What's the Best Way to Resolve a Plausibility Paradox?

Mathew Coakley

doi: 10.2478/disp-2023-0019 BIBLID: [0873-626X (2023) 71; pp. 491–511]

Abstract

A "plausibility paradox" has very plausible premises that seem to require acceptance of an implausible conclusion. Resolutions can reject a premise, accept the conclusion, or deny that either is necessary. Some general strategies, however, may in expectation be superior. This paper argues that if acceptance of the premises requires acceptance of the conclusion, then the best resolution would be to show that the paradox relies on a psychological framing effect. Specifically, the premises are accepted yet replaced with functionally equivalent ones where the conclusion is now no longer implausible. To illustrate, the paper provides such a resolution to Hempel's ravens' paradox and contrasts it with existing rivals.

Keywords

confirmation theory; paradoxes; philosophical method

1 Scope and strategy

Let a "plausibility paradox" be one with:

(1) Some premises that seem very plausible and that, therefore, should be accepted (at least under certain specified conditions).[1]

[1] 'Very plausible' and 'accepted' are meant permissively as placeholder terms capturing a certain level of epistemic commitment, such as alternatively 'highly probable' or 'believed'.

(2) A conclusion that seems implausible and that, therefore, should not be accepted.

(3) A claim of "transmission" whereby acceptance of the premises seems to require acceptance of the conclusion.

Diverse examples include the mere addition paradox and the repugnant conclusion in population ethics (Parfit [1984: part IV]), Smilansky's misfortune paradoxes [2007], Hempel's [1945] confirmatory ravens' paradox, Nozick's experience machine in value theory [1974: 42–5], and Newcombe's decision ranking paradox (see Sainsbury [2009: 69–82]).

For a small subset of plausibility paradoxes, the core dispute concerns whether transmission holds (*e.g.*, Zeno's paradox, Huggett [2024]). For other instances where it does hold, the options are to reject a premise or accept the conclusion. Neither, however, will necessarily provide a satisfactory resolution. Prominent paradoxes are prominent for a reason: most people's judgement of the premises' plausibility and the conclusion's implausibility seems well justified. A resolution, if it does not deny transmission, should also explain why one of these widely held judgments is mistaken. The better that explanation, the better the resolution.

This paper proposes a tightly defined resolution strategy, "replace-yet-accept", that responds directly to this explanatory demand. It has two necessary components. First, the original premises are replaced with alternative premises where both sets of premises—via transmission—lead to the same conclusions, these and these alone. Second, under the new premises, the previously problematic conclusion is no longer implausible.

Under replace-yet-accept, the cognitive implausibility of the paradox's conclusion is shown to be driven by the manner in which the premises are presented, rather than their joint substantive content[2]. That is, people's judgements of plausibility and implausibility may often depend not only on the formal logical content but also on how something is framed.

For example, here are the premises and conclusion of what could be called the "Sheepdog Paradox":

P1. All the animals on the farm are grey.

---

[2] There is a large body of evidence that framing can make a huge difference to conclusion plausibility / implausibility via at least two mechanisms: the directing of attention to certain features and the selective activation of various heuristics. See most influentially Kahneman & Taversky [1979], or on mechanism variation Levin, Schneider & Geath [1998].

P2. All the mammals on the farm are not-grey.

Conclusion: All the dogs on the farm are sheepdogs.

This seems mildly puzzling. However, assuming classical logic (which this paper does both here and in the subsequent ravens' discussion), transmission does obtain. Yet, the conclusion appears undersupported. Consequently, one might worry that the definitions in the premises are somewhat suspect.

Imagine you were to ask students: 'if you accept the premises, should you accept the conclusion?'. My guess is their responses would vary, with most uncertain or perhaps even leaning against acceptance.

However, the syllogism is perfectly valid. If the premises are true, so is the conclusion. It is the wording that is misleading; the emphasis on sheepdogs and the colours of mammals detracts from the real issue—namely, the existence of farm dogs. But it can be reframed to remove the paradoxical feel.

P1*. All the animals on the farm are grey.

P2*. There are no mammals on the farm.[3]

Conclusion*: There does not exist a (non-sheepdog) dog on the farm.

There's not much paradoxical about this. Indeed, if you were to ask the same students the same question—'if you accept the premises, should you accept the conclusion?'—presumably, many more would answer yes, confidently. Yet, the premises of both syllogisms are jointly functionally equivalent, in that (P1 & P2) entails (P1* & P2*), and vice versa. The two conclusions are logical equivalents. With a non-puzzling frame, it is intuitively easy to understand why the conclusion holds given the initial premises, and therefore to resolve the paradox. In this case, the statement 'All the mammals on the farm are not grey' can be misinterpreted as necessarily implying the existence of a non-grey mammal, whereas

[3] A different P2*, and one that does not entail the conclusion on its own, would be 'If there are any mammals on the farm, they are not grey'. A reframe with a premise that entails the conclusion is a valid example of the strategy, but sometimes one that relies on all the new premises to generate the previously problematic conclusion might make the explanation cognitively more easily accessible. I'm grateful to Elia Zardini for the suggested alternative.

it would be true, for instance, of a farm with only fish, chickens and wheat. The initial framing makes one aspect—the possibility of the farm having both mammals and non-mammals—overly cognitively prominent.

In addition to misleading-feature salience, apparent puzzles can be created via frames that reinforce a false yet common external assumption. Here is an informal description of what might be termed the 'Waitress Paradox'.

Some bankers are having after-work beers in a pub. A waitress carrying a plate of drinks accidentally drops them with a loud smash and rushes into the kitchen, visibly upset. 'I'm sorry', she confides to a waiter, 'my dad is in the hospital undergoing surgery, and I'm really distracted'. Meanwhile, one of the bankers, looking somewhat embarrassed, admits to a colleague that, 'Erm, the clumsy waitress is actually my daughter'. Neither the waitress nor the banker is lying. How come?

This could be formally expressed as a series of premises, with the conclusion that both are telling the truth appearing mildly paradoxical to some. However, this type of scenario is now such a staple of undergraduate ethics lectures that many will immediately recognise that the banker is the waitress's mother.

The case can also be trivially reframed so that anyone newly asked it is unpuzzled—simply stating that the banker turned to a colleague and that "she or he" confided would probably suffice. Of course, such an addition hasn't added logical content. Instead, a background assumption has been weakened. Indeed, some years ago, I used to tell students a version using a surgeon as the illustrative example. However, as medical gender norms shifted, too many began to immediately intuit the answer, so to illustrate the nature of role stereotypes, it had to become about bankers.

After a few moments of reflection, neither paradox is particularly puzzling. The key conceptual point, however, is that whether a conclusion *appears* paradoxical can depend on how the premises are framed, even if the different frames leave the joint entailments the same. The property of "seeming paradoxical" relies on which presuppositions or features of the scenario are psychologically prominent, not simply the overall propositional content. It may, therefore, be possible to reframe premises to remove a conclusion's implausibility, even while retaining the joint content.

This paper's weaker thesis is that replace-yet-accept is a clear explanatory strategy with advantages over its rivals. The stronger thesis is that absent the justified denial of transmission, it should normally be regarded as the best possible resolution when compared to alternatives.

Finally, there is an important group of paradoxes that may not be covered here—

"impossibility paradoxes"—in which something claimed to be possible is allegedly shown to be impossible, or *vice versa*. These fit at least one of the following two structures:

(1) A (sometimes implicit) premise that X should always be impossible.

(2) A case or set of conditions that are apparently possible.

(3) The claim that if (2), then X.

Or:

(1^*) A (sometimes implicit) premise that Y should always be true (or hold).

(2^*) A case or set of conditions that are apparently possible.

(3^*) The claim that if (2^*), then Y is not true (or does not hold).

An example of X might be that a set is a member of itself (Russell [1919]), or of Y, that a statement that is internally logically coherent should be capable of being either true or false but never simultaneously both (the liar paradox)[4]. Whether these are slightly disguised plausibility paradoxes depends on the grounds for accepting the first premise—in principle, some things might be ruled impossible or always true due to purely theoretical considerations, such as the need to fulfil a certain desirable function. Whether those sorts of goals ultimately depend upon plausibility judgments will determine if there is a separate type of paradox, a question this paper leaves open.

It proceeds as follows. Part two outlines the replace-yet-accept approach. Part three uses the ravens' paradox to illustrate the strategy. Part four examines the strategy's comparative advantages over rivals. Part five considers the broader advantages.

2 Replace-yet-accept and alternatives

The resolution strategy explored here is the following:

> Replace-yet-accept: replace the original premises with new plausible premises such that—via transmission—every conclusion of the original premises is a conclusion of

[4] This example's formulation is merely illustrative—a key issue for impossibility paradoxes is exactly what is supposed to be impossible or not.

> the new premises, *and vice versa*, and where the previous conclusions are not now implausible.

In this paper, sets of premises will be described as "reframes" of each other if and only if both have—via transmission—the same set of implications, and only these. As the content of transmission is field-specific, so too will be the conditions for a valid reframe. Thus, in confirmation theory, premise reframes presumably have the same possible objects confirming (or not) the same possible hypotheses. With Parfit's population ethics paradoxes, reframes would normatively rank all possible populations identically and have rankings over all the same possibilities.

Replace-yet-accept is intended for when "transmission" is valid. In such cases, there are well-explored alternatives:

> Simple-reject: reject a premise without committing to a specific replacement.
>
> Accept-and-hypothesise: accept the premises, and offer a hypothesis for why the implausible conclusion should be nonetheless accepted.
>
> Reject-and-replace: reject a premise, replace it with an alternative, and then—possibly—offer some explanation for the erroneous previous judgement.

Replace-yet-accept and accept-and-hypothesise may superficially seem similar, as both aim to explain why the paradox should be accepted. However, they differ in what is explained and how. Under replace-yet-accept, the conclusion's implausibility is shown to be a contingent framing effect, and something we can, therefore, remove. Under accept-and-hypothesise, the implausibility is potentially retained, but there's a suggestion of why we should accept the conclusion nonetheless.

With Parfit's [1984] core population ethics paradoxes, for instance, an accept-and-hypothesise resolution would provide reasons for accepting the repugnant conclusion. A replace-yet-accept resolution would reframe the premises so that the conclusion is no longer repugnant.

3 Illustrating the resolution method

The previous strategies are required if transmission holds. Here—via Hempel [1945]—is one such case, the ravens' paradox ('RP'):

Premises:

Nicod's Condition (NC). *'All F are G' is confirmed by an F that is G.*[5]

The Equivalence Condition (EC). *Whatever confirms (disconfirms) one of two logically equivalent sentences also confirms (disconfirms) the other.*

Conclusion:
A non-black non-raven, such as a white shoe, confirms 'All ravens are black'.

Transmission is valid. 'All ravens are black' is logically equivalent to 'All non-black things are non-ravens' and thus would be confirmed under NC by a non-black non-raven.

This is also potentially a *plausibility* paradox: a white shoe appears irrelevant to a theory about raven colours, yet, absent confounding background information, it does seem that an F that is G should confirm 'All F are G'.

The RP literature is truly vast. However, there are six types of extant resolutions. Here is an example of what I take to currently be either the best, most prominent, or most discussed version of each. (1) Reject Nicod's Condition via apparent violator examples (Good [1967]). (2) Reject the Equivalence Condition by arguing that it may not need to be accepted in certain contexts (Clarke [2010]). (3) Accept the premises but provide an explanation for the problem case (Hempel [1945], Goodman [1955], assorted authors). (4) Reject the premises by replacing them with a probabilistic framework that disputes the cases (Good [1960]). (5) Reject the premises by replacing them with a probabilistic framework that explains a problem case—such as the white shoe—but disputes others (Howson & Urbach [2006]). (6) Reject the premises by replacing them with a probabilistic framework that agrees on the core cases (Maher [1999]).

These adopt one of the three earlier outlined resolution strategies: simple-reject (1 & 2), accept-and-hypothesise (3), or reject-and-replace (4, 5 & 6).

Finding a rival replace-yet-accept resolution requires three potentially iterated steps. First: identify all the conclusions of the paradox's premises. Second: find different plausible

[5] Nicod's Condition, or sometimes 'Nicod's Sufficient Condition', is conventionally phrased with a hypothesis statement 'All F are G' being confirmed by an object—an F that is G. An alternative is to hold that the confirmer is a statement—such as 'There exists an F that is G'. For internal literature consistency, this paper adopts the former, as it also does with all similar instance-principles.

premises that produce these and only these. Third: see if the conclusions are now plausible in light of the new premises. If not, repeat step two…

With the RP, step one involves identifying all the confirmatory instances given Nicod's Condition and the Equivalence Condition. This is that 'All F are G' is confirmed by (a) an F that is G, (b) a not-F that is not-G, (c) a not-F that is G, (d) a not-F (of unknown G-predicate status), and (e) a G (of unknown F-predicate status).

Thus, 'All ravens are black' is confirmed by (a) a black raven, (b) a non-black non-raven (*e.g.*, a white shoe), (c) a black non-raven (*e.g.*, a black shoe), (d) a non-raven of unknown colour (*e.g.*, a shoe simpliciter), and (e) a black object of unknown raven/non-raven status (a black thing).

Cases (c), (d), and (e) are oddly significantly less literature-prominent than the white shoe confirmer, despite being clearly noted by Hempel [1965: 14–5]. When discussed, these cases are usually framed as an additional difficulty. For Whitely [1945], for instance, this is a "second paradox"; for Earman and Salmon confirmation is rendered for many "too easy to be true" ([1992: 54]).

Whether well discussed or not, the confirmatory cases form the set of transmitted conclusions. To phrase them generally, 'All F are G' is logically equivalent to 'All (F or not-F) are (not-F or G)', and thus anything [(F or not-F) that is (not-F or G)] is a confirmer—that is, anything that is (F and G) or (not-F) or (G).

Step two is to identify plausible premises that produce these, but only these, and that remove the conclusions' implausibility. Here is one such set of premises (and, I believe, the best—a potential rival drawing on Maher [1999] is in 4.3):

"*Non-existence ravens' paradox*" *premises*:

> **Non-Existence Condition (NEC).** *'There does not exist an H' is confirmed by anything not-H.*
>
> **The Equivalence Condition (EC).** *Whatever confirms (disconfirms) one of two logically equivalent sentences also confirms (disconfirms) the other.*

This meets the replace-yet-accept requirements. First, for any hypothesis, everything that is confirmatory under the original RP is also confirmatory under the non-existence RP, and only these things. Second, for any hypothesis, everything that is confirmatory under the non-existence RP is likewise confirmatory under the original RP, and only these things. Third, upon examination, the conclusions of the non-existence RP do not seem particularly implausible.

Under the non-existence RP, 'All ravens are black' is confirmed by a white shoe, not because the shoe is white but because it is not a non-black raven. 'There does not exist a (non-black raven)' is confirmed by anything that is not a (non-black raven). That's not implausible: 'There does not exist a pink unicorn' should normally be confirmed by anything not a pink unicorn.[6]

Imagine—counterfactually—that Hempel had set out and accepted the Equivalence Condition and then proposed the Non-Existence Condition (NEC), which states that 'There does not exist an H' is confirmed by anything not-H.

He could then perhaps have noted that 'There does not exist a (non-black raven)' is confirmed by anything not a non-black raven. Would anyone seriously have elevated this to a paradox? Count me deeply sceptical. It seems the opposite. Absent confounding information, the weird result would be 'There does not exist an H' *not* being confirmed by a not-H.

The key lacuna in Hempel's otherwise impressively exhaustive treatment arises when discussing whether generalisations should have existential import (see section 5.11, [1965: 16–7]). He offers three specific reasons why they should not. First, doing so would block 'All F are G' and 'All not-G are not-F' from being logical equivalences, but sometimes this is useful, as with his sodium salts example. Second, in some scenarios, the object of the existential clause needs to be specified, which might be done arbitrarily, as with his diphtheria case. Third, some hypotheses clearly make claims about what might exist, not what does, and thus don't appear to imply any existential clause. Having highlighted these three cases, he concludes that generalisations should not have existential import, and the argument moves on.

His discussion, however, omits three important possibilities. First, there is no good reason not to allow generalisations both with and without existential import, so long as it is clear in any given case which applies, which modern logic trivially permits. His three substantive counter-examples only target the straw-person position that all generalised scientific hypotheses must have existential import, a dogmatic stance that few would wish

---

[6] As the Hempel-Good debates made clear, such judgements do need to exclude some possible background knowledge and have to be domain-specific. Thus, the claim here is in the context of thinking about objects that might exist. It might still seem paradoxical that non-(pink-unicorns) on Earth support scepticism about pink unicorns on the moon, but the positive case would be that if they are known to exist somewhere, there's an increased probability they do elsewhere, and the reverse. Becoming more sceptical of them on Earth thus warrants at least some more scepticism about them elsewhere.

to assert.

Second, the set of confirmatory instances for any specific generalisation may differ, and at times should differ, based on whether it has existential import. 'All ravens and black' and 'All ravens are black and there exists a raven' should not necessarily be confirmed by exactly the same set of things. Eliminating one of these hypotheses by near-fiat leaves this fecund issue unexplored.

Third, as with the "Sheepdog Paradox", if people tend to assume that generalisations have existential import, even when formally they do not, then this might produce results that seem puzzling. If the paradoxical feel of the conclusions relies on a cognitive misinterpretation of 'All F are G', then we may be able to make progress simply by phrasing the premises in a way that prevents any confusion over existential import, such as by using non-existence claims.

The reframed RP premises, where 'There does not exist an H' is confirmed by a non-H, therefore straightforwardly explain all the confirmers: (a) a black raven, (b) a non-black non-raven (*e.g.*, a white shoe), (c) a black non-raven (*e.g.*, a black shoe), (d) a non-raven of unknown colour (*e.g.*, a shoe *simpliciter*), and (e) a black object of unknown raven/non-raven status (a black thing). If we know an object is not a (non-black raven), then the hypothesis is confirmed.

The premises also explain the non-obvious *non-confirmers*, such as a non-black object of unknown raven/non-raven status, or a raven of unknown colour status. If we do not know that an object is not a (non-black raven), then the hypothesis is not confirmed. The logic here is clean and clear, and it straightforwardly covers—without exception—every confirmatory instance as well as every non-confirmatory one.

If, after exposure to the non-existence RP, someone maintains that the conclusions of the original are still truly paradoxical, they then face a dilemma. They can argue that the conclusions of the reframed RP are themselves paradoxical, even though no one new to the reframe would likely think that. Alternatively, they can argue that if two equivalent sets of premises entail the same conclusions, and one set renders the conclusions non-paradoxical, the conclusions should still be regarded as inexplicably paradoxical. This would be akin to simultaneously holding all three of the following: (i) the Waitress Paradox is puzzling if it is assumed the banker is male; (ii) the Waitress Paradox is not puzzling once it is recognised that the banker might be female; (iii) and yet the Waitress Paradox is still puzzling. Aside from a fondness for bafflement, it is difficult to see the coherent motivation for holding all three views simultaneously.

It is interesting that a replace-yet-accept solution to the ravens' paradox exists. However,

the broader issue relates to which paradox strategies are likely to be better and why. Below, I offer arguments for why replace-yet-accept enjoys expected advantages over alternative strategies (section 4) and offers a platform for future progress (section 5).

4 The comparative advantages of replace-yet-accept

4.1 Some downsides to simple-reject

One option is to reject one or more of a plausibility paradox's premises. In the ravens' case, for example, Clarke [2010] has argued that the Equivalence Condition should, in fact, be contingent on the context of enquiry. Good [1967] has proposed situations where it seems Nicod's Condition should not obtain.

This strategy has at least three downsides: a likely methodological narrowing, a failure to explain diverse puzzlement, and a vulnerability to scope restrictions.

First, a paradox's premises are typically methodologically ecumenical in that they are compatible with a diverse set of approaches, ones that have independent reasons for being taken seriously. If a paradox were to lack this feature, the incompatible approach would offer a ready resolution. As a result, the rejection of one of a paradox's assumptions is generally expected to entail a controversial methodological position.

For instance, rejecting the Equivalence Condition is extremely restrictive of the set of potential methods in confirmation theory. Similarly, claiming that justified confirmation without probabilistic background assumptions is impossible is highly reductive of approaches.

If this methodological narrowing is indeed the only viable option, it may be an unavoidable cost. But, in principle, resolution strategies that retain a paradox's permissive theoretical approach are preferable.

Second, there is a selection effect: to become a well-known paradox, the premises must appear very plausible to a wide range of individuals who might otherwise strongly disagree. Therefore, showing that it is possible to reject the premises seems too weak. We need an explanation of why so many thoughtful individuals were mistaken.

Third, even if there are cases where the premises seem not to obtain, it is usually possible to find many cases where they do, leaving simple-reject vulnerable to scope restrictions. Hempel's [1967] response to Good is exactly of this type. Good invented odd background evidence where Nicod's Condition appears not to apply. Hempel aptly responds: sure, but what about when there is either no background evidence or typical background evidence?

The underlying weakness of simple reject is that there are prima facie good reasons to accept the premises of well-established paradoxes: their compatibility with most methodological approaches, their considered endorsement by a diverse set of philosophers, and their seemingly justified application to most cases, even if not all.

This does not guarantee they are correct. But, it does create a presumptive burden to demonstrate that rejection is preferable to cautious or qualified acceptance and, furthermore, to *explain* why most people are mistaken. Simple rejection struggles with both.

4.2 Some downsides to accept-and-hypothesise

Under accept-and-hypothesise, the premises are retained, and a hypothesis is proposed to explain why we should still accept the paradoxical conclusion. In contrast to rejection strategies, this method has the advantage of preserving a paradox's methodological and philosophical presuppositions.

Its central weakness is that it lacks any necessary framework to differentiate a good hypothesis from a merely possible one, and the fact that a hypothesis is possible may give it undue prominence, even if it is a poor explanation. The bar on accept-and-hypothesise is too low.

Take, for instance, perhaps the most widely cited explanation of the ravens' paradox: the "observation order" hypothesis (this is initially briefly explored by Hempel [1945: 5.2 (b)], and then becomes a common explanation). Here is Sainsbury's example of the approach [2009: 97]:

> Suppose that we are on an ornithological field trip. We have seen several black ravens in the trees and formulate the hypothesis that all ravens are black. We then catch sight of something white in the topmost branch. For a moment we tremble for the hypothesis, fearing a counterinstance—fearing, that is, that we have found a white raven. A closer look reveals that it is a shoe. In this situation, we are more likely to agree that a white shoe confirms the hypothesis.

Similarly, here is Godfrey-Smith's [2003: 49] crisp illustrative description:

> You believe that all ravens are black, and someone comes up and says, 'I have a white object behind my back; want to see what it is?'. You should say yes, because if he has a raven behind his back your hypothesis is refuted. He pulls out a shoe, however, so your hypothesis is OK.

The common interesting underlying idea is that separate sequentially observed predicates can be used to test theory predictions. What, though, does this have to do with the ravens' paradox? Yes, it offers us an explanation of the white shoe confirmer, true. But under the paradox, a black shoe confirms 'All ravens are black', as does any shoe, including one of unknown colour. Furthermore, a white shoe confirms 'All ravens are black' and confirms 'All ravens are white'.

To wit: we're on an ornithological field trip and catch sight of something white in the topmost branch. Should we tremble for the hypothesis 'All ravens are white', fearing a counterinstance? No. But a white non-raven would confirm it (as would the white object *simpliciter*).

Similarly, someone says, 'I have a white object behind my back, want to see what it is?' Your hypothesis is, 'All ravens are white'. If you're worried about your hypothesis being refuted, why say yes (or no)? Under the RP, the white shoe confirms it.

The observation-order hypothesis explains an undeservedly prominent case in the literature—the white shoe confirmer—but it does not explain the paradox.

To be fair, both authors qualify their implied semi-endorsement of observation order as a potentially good answer. Sainsbury hedges that for such background evidence with contingent observational stories, "Perhaps these can be developed to justify the first response" [that the paradoxical conclusion is, after all, acceptable]. "Clearly", writes Godfrey-Smith, "this discussion of order-of-observation does not entirely solve the ravens problem. *Why* does order matter, for example, and what if both properties are observed at once?".

But order doesn't matter to the paradox except in the trivial derivative sense that 'All ravens are black' is confirmed the moment you know that an object is not a non-black raven, and never, ever, before then.[7]

A resolution strategy as open-ended as accept-and-hypothesise does not necessarily explain all the implications of a paradox's premises, even if it might explain a historically salient one. Someone can offer an explanation of the problem case, it seems credible, and yet there can be implications of that explanation that misfit other premise-conclusions.

---

[7] One could try to claim the first observation is that it is an object, the second is that it is not a non-black raven. This, however, then has nothing to do with observation order, only the status of not being a non-black raven matters, and we require an explanation of why (which both this paper's non-existence RP and Maher's rival counterexample explanation potentially supply).

4.3 Some downsides to reject-and-replace

An alternative is to reject a premise or methodological presupposition and replace it with a different one that might provide insights into the original paradox. For instance, Nicod's Condition doesn't incorporate contingent background information, but it can be replaced with an approach that does, and the cases then explored.

This strategy underpins all probabilistic RP resolutions and was explicitly applied by Good [1960], although the earliest discussion appears to be in Hosiasson-Lindenbaum [1940: section 4]; see also Hempel [1945: 21, fn 2]. The two most prominent—and also notably contradictory—modern accounts are those of the Standard Bayesian Solution (see Howson & Urbach [2006]) and Maher's [1999] Carnapian solution.

The drawbacks of reject-and-replace as a general strategy depend critically on precisely what is rejected: a conclusion or a premise.

If a reject-and-replace resolution rejects one or more of a paradox's seemingly implausible conclusions, it will not explain the original paradox and will inherit the drawbacks of simple rejection.

The Standard Bayesian Solution ('SBS'), for example, superficially seems to explain the paradox. The white shoe confirms that 'All ravens are black', but only by a minuscule amount. The concept of small-degree confirmation appears to explain why we misinterpret the paradox's apparently puzzling cases.

This, however, cannot be generally true. Under the original paradox, 'All ravens are black' is confirmed both by a black object and by a black non-raven. Yet, under the SBS, a black object neither confirms nor disconfirms 'All ravens are black', and a black non-raven actually disconfirms it.

There is a careful dispute regarding whether the SBS's conditional probability assumptions can be justified: see Vranas [2004] and then Howson & Urbach [2006: 101] for a response. My point here is different. The SBS is not an explanation of the paradox. It does not explain why the premises should be rejected. It does not explain why the conclusions should be accepted, except for one conclusion, that of the white shoe confirmer, but not the others. It demonstrates that there exist possible background assumptions under which Nicod's Condition might not hold. But both sides of the Hempel-Good debate from decades earlier accepted that. Their real dispute was: should Nicod's Condition hold in "typical" cases, or absent background information about the numbers of types of objects in the universe?

A reject-and-replace strategy that rejects a conclusion adds nothing to our understand-

ing unless it also explains why it—and the premises—should be rejected. One or more conclusions are already deemed implausible; that's why there's a paradox in the first place.

Imagine, instead, that a reject-and-replace proposal accepts the core conclusions. Maher [1999], for example, provides a probabilistic resolution that agrees with the classic paradox concerning both the white shoe confirmer and the black shoe confirmer. His explanation [1999: 57] is that:

> A non-raven (of whatever colour) confirms that all ravens are black because (i) the information that this object is not a raven removes the possibility that this object is a counterexample to the generalisation and (ii) it reduces the probability that un-observed objects are ravens, thereby reducing the probability that they are counterexamples to the generalization.

Call this 'the counterexample explanation'.

The specific downside of a reject-and-replace resolution depends on whether the rejection is merely used to identify an explanation for the original paradox or is necessary for the newly proposed resolution.

If a proposed explanation is meant to cover the original paradox, it should be convertible to a replace-yet-accept resolution. It can then be directly scrutinised to ascertain whether it provides the best existing explanation.

If it does, why reject? Replace-yet-accept maintains a paradox's ecumenical presuppositions and broad theorist appeal. Reject-and-replace is likely to introduce either a methodological narrowing or obfuscating complexity: most paradoxes are admirably clear and accessible.

If, by contrast, its explanation *doesn't* cover the original paradox, then it appears inferior to one that does, namely, an accept-yet-replace resolution.

Maher's counterexample explanation seems intended to explain the original paradox—the formal probabilistic framework is being used as a source of insight. Therefore, it ought to have a replace-yet-accept variant, and indeed, it does. Here is what seems to be the obvious candidate:

Replace-yet-accept "Counterexample" RP premises:

> **Counterexample Condition.** *'All F are G' is confirmed by anything known not to be a counterexample to it.*
>
> **The Equivalence Condition (EC).** *Whatever confirms (disconfirms) one of two logically*

equivalent sentences also confirms (disconfirms) the other.

Why, however, reject the original paradox's simple instance-based framework if this is a good explanation? If the explanation generalises convincingly, we don't need to adopt a Carnapian sampling framework with all its methodological commitments and controversial assumptions.

To be clear, I believe that the counterexample resolution—though interesting and promising—offers a comparatively worse explanation compared to that of this paper for the reasons outlined in 5.3. The important point here, however, concerns the potential redundancy of a reject-and-replace resolution that accepts all the conclusions while claiming that its explanation applies to the original paradox, yet implicitly rejects the original paradox's simple frame with its broad theorist appeal. If it provides a good resolution, then the methodological narrowing should be rejected, and replace-yet-accept adopted.

4.4 Why these comparative advantages are to be expected

The comparative advantages of replace-yet-accept stem from two underlying structural features, a combination that other resolutions inevitably lack. First, a paradox's premises will only be plausible if both their within-paradigm content and their broader methodological assumptions are credible and allow for a wide range of approaches. Reject-the-premises strategies will therefore typically be theoretically costly, requiring us to shrink our set of viable methods or resolve a controversial question that the paradox leaves open. A resolution that does not require this narrowing is prima facie better.

Second, ongoing plausibility paradoxes are not merely internal logic puzzles. There are widely shared human judgements underpinning them, judgements that have proven stable under prolonged scrutiny. Since some of these plausibility judgements will ultimately require revision—the conflict creates the apparent paradox—a resolution that explains the source and nature of the erroneous judgement will be prima facie better than one that does not. (As such, a resolution strategy that instead showed why the premises were *cognitively* appealing, albeit mistakenly so, would also hold this advantage and be a rival to replace-yet-accept.)

Replace-yet-accept, by its very structure, both preserves widely held judgements over the premises and explains—if the implausibility of the conclusions is truly removed—why a paradox initially appeared to exist. A resolution that fails to accomplish both is inferior to one that does. A resolution that achieves both, through stronger assumptions, has built-in redundancy: it can be transformed into a replace-yet-accept resolution, which can then be

straightforwardly compared to any others.

5 Some broader advantages of replace-yet-accept

5.1 Replace-yet-accept provides a coherent frame for analysing why a paradox might cognitively mislead

If a conclusion is implausible given one set of premises and plausible given another, with each set requiring acceptance of the other, then a framing effect is present. This offers a psychological explanation for the earlier clash of plausibility judgements. An offshoot is that there is now a coherent cognitive frame for analysing the different parts of the initial paradox and what we previously misunderstood.

In the ravens' case, for instance, it appears that the classic frame lends itself to two related types of errors. First, focusing on which objects surprisingly confirm 'All ravens are black' obscures the question of which objects confirm one 'All ravens are X' hypothesis rather than another. If an object confirms 'All ravens are X', our explanation for why should also explain, and thus be informed by, why it confirms 'All ravens are not-X' or not.

For instance, imagine that you ask students familiar with the paradox: What confirms 'All ravens are black' but not 'All ravens are non-black'?

My guess is that a few may take some time to respond. Of those who do, some might venture the answer 'A black raven and a non-black non-raven', which, considering the focus of the existing literature, would not be unreasonable to conclude.

But of course, it is false. Under NC and the EC, the only objects that confirm 'All ravens are black' and also do not confirm 'All ravens are non-black' are a black raven and a black object of unknown raven status. If the only cognitively prominent hypothesis is 'All ravens are black', this can be non-obvious.

Second, the classic frame leads people to mistakenly believe that the white shoe confirming 'All ravens are black' necessarily reveals something about the colour of unobserved ravens. Thus, we risk engaging in what Goodman [1955: 71] mischievously dubbed "indoor ornithology".

This, however, is false. It would only be true if a white shoe confirmed a hypothesis of form: 'All ravens are black, and there exist x ravens', where x is a positive integer larger than the number of observed ravens. But that isn't what the original paradox entails. There is no existential import; hence, it can be reframed via non-existence hypotheses.

5.2 Increased cognitive ease of use

If a new frame renders a paradox's conclusions plausible, this is likely to also make them more cognitively accessible, perhaps providing a small but non-trivial benefit.

Consider the hypothesis 'All two-legged swans are (male or white)'. Under Nicod's Condition and the Equivalence Condition, which of the following objects would confirm the hypothesis: (i) a white female swan, (ii) a black, no-legged swan, (iii) a two-legged raven, (iv) a non-white swan, (v) a black shoe?

My guess is that answering may require a few moments of reflection, at least for many. However, with the reframe, this becomes much more straightforward.

What should confirm 'There does not exist a (non-white non-male two-legged swan)'? Anything known not to be a (non-white non-male two-legged swan). So (i), (ii), (iii) and (v).

Answering becomes easier when using the reframed premises, and that's the point. A reframe that aligns with our plausibility judgements makes it much easier to cognitively map premises to conclusions.

5.3 Multiple resolutions

Multiple distinct premises may meet the replace-yet-accept criteria. Within the paradox's scope, these will always agree upon the transmitted conclusions. However, the premises are not necessarily identical, as they may offer potentially competing options when moving beyond the paradox via a wider scope.

The non-existence resolution and the counterexample resolution, for instance, both cover only one type of hypothesis: those logically equivalent to 'There does not exist an F'. They agree here on the confirmatory instances. But, in expanding the scope, this need not be the case.

Should, for example, the hypothesis 'All F are G, and there exists an F' be confirmed by a not-G that is not-F? The underlying logic of the counterexample explanation appears to support yes, as it clearly could have been a counterexample.

The same is not true of the resolution of this paper (see 5.4). On the surface, this confirmation seems paradoxical—should a white shoe confirm 'All Ravens are black and there exists a raven'? Thus, this appears to provide a comparative reason to reject the counterexample explanation, although extensions or additions might change this assessment. At a minimum, the counterexample explanation suggests questionable confirmations that the non-existence resolution does not.

When thinking outside the scope of a paradox, various accept-yet-replace resolutions suggest different approaches or research paradigms. Some fail to be compellingly generalisable, while others lend themselves to ready and simple extension. This provides a way of comparing the resolutions and their underlying rationales even if, within the paradox's scope, they agree on the conclusions.

5.4 Replace-yet-accept is methodologically liberating and provides a platform for additional premises

If a replace-yet-accept resolution exists, this has implications for the set of broader theoretical approaches within a given domain. For example, the existence of replace-yet-accept resolutions to the ravens' paradox indicates that an instance-based (background-assumption-free) approach need not be rejected in favour of a probabilistic account. Both can be explored with regard to their differing merits. This is methodologically liberating, unlike the situation where the only options are to reject or to reject-and-replace.

One consequence is that we can build upon a paradox in a way that preserves its underlying approach and explains why it is interesting. Take, for instance, this paper's reframed ravens' paradox premises:

> **Non-Existence Condition (NEC).** *'There does not exist an H' is confirmed by anything not-H.*

> **The Equivalence Condition (EC).** *Whatever confirms (disconfirms) one of two logically equivalent sentences also confirms (disconfirms) the other.*

We could expand the scope of these also to cover hypotheses about generalisations of objects that are claimed to exist, such as that:

> **Nicod's With-Existence Condition (NWEC).** *'All F are G and there exists an F' is confirmed by an F that is G.*

Adding this produces no additional raven-style paradoxical-seeming cases. It helps make clear that 'There does not exist a non-black raven' and 'All ravens are black and there exists a raven' should have different confirmatory instances. It seems revealing that only a black raven confirms both, whereas a non-raven—black or non-black—confirms only the first.

The broader point is that reframed premises that remove a conclusion's implausibility are easy to build upon. Crucially, this can be done while retaining the overall approach; in

this case, that we have an instance account absent background information. The replace-yet-accept criteria can also identify which challenges can be met by psychological explanation (such as the RP) and which cannot, and are therefore more fundamental (such as blite or grue cases).

6 Conclusions

Plausibility paradoxes are important philosophical puzzles stemming from an apparent conflict between plausible premises and at least one implausible conclusion. In principle, there are multiple ways to resolve them. However, we can do more. We can also explain why the original paradox arises.

Replace-yet-accept is a tightly defined resolution strategy that does exactly that, drawing on the possibility that framing effects shape our plausibility judgements. Furthermore, it does so while being able to retain why the paradox was interesting in the first place and why the premises are widely judged convincing.

A method that resolves a paradox, corrects previous errors in judgement and offers a foundation for building on the premises seems at a minimum to offer a good resolution. Absent the justified denial of transmission, it might provide the best one.

There are many plausibility paradoxes. If we can find replace-yet-accept resolutions for them, it appears we will be making progress[8].

Mathew Coakley
m.p.coakley.97@cantab.net

References

Clarke, Roger [2010]. ""The Ravens Paradox" is a misnomer". *Synthese* **175**: 427–40.

Earman, John, Salmon, Wesley C. & Salmon, Merrilee H. *et al.* [1999]. *Introduction to the Philosophy of Science.* Indianapolis (IA): Hackett.

Godfrey-Smith, Peter [2003]. *Theory and Reality*. Chicago: Chicago University Press.

Good, Irving John [1967]. "The White Shoe is a Red Herring". *British Journal for the Philosophy of Science.* **17**: 322.

[8] I'm extremely grateful for the anonymous reviewers' comments.

Goodman, Nelson [1955]. *Fact, Fiction and Forecast.* Cambridge (MA): Harvard University Press.

Hempel, Carl Gustav [1945]. "Studies in the Logic of Confirmation I & II". *Mind* **54**: 1–26 & 97–121.

Hempel, Carl Gustav [1965]. *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*. New York (NY): The Free Press.

Hempel, Carl Gustav [1967]. "The White Shoe - No Red Herring". *The British Journal for the Philosophy of Science* **18**: 239.

Hosiasson-Lindenbaum, Janina [1940]. "On Confirmation". *The Journal of Symbolic Logic* **5**: 133-48.

Howson, Colin & Urbach, Peter [2006]. *Scientific Reasoning. The Bayesian Approach.* Illinois: Open Court.

Huggett, Nick [2024]. "Zeno's paradoxes". *The Stanford Encyclopaedia of Philosophy*, [Fall 2024 Edition], edited by E. N. Zalta & U. Nodelman. https://plato.stanford.edu/archives/fall2024/entries/paradox-zeno/

Kahneman, Daniel & Tversky, Amos [1979]. "Prospect Theory: An Analysis of Decision under Risk". *Econometrica* **47**: 263–91.

Levin, Irwin., Schneider, Sandra, & Gaeth, Gary [1998]. "All frames are not created equal: A typology and critical analysis of framing effects". *Organizational Behaviour and Human Decision Processes* **76**: 149–88.

Maher, Patrick [1999]. "Inductive logic and the ravens' paradox". *Philosophy of Science* **66**: 50–70.

Nozick, Robert [1974]. *Anarchy, State and Utopia*. New York: Basic Books.

Parfit, Derek [1985]. *Reasons and Persons*. Oxford: Oxford University Press.

Russell, Bertrand [1919]. *Introduction to Mathematical Philosophy*. London: George Allen and Unwin Ltd.

Sainsbury, R. Mark [2009]. *Paradoxes*. Cambridge: Cambridge University Press.

Smilansky, Saul [2007]. *10 Moral Paradoxes*. Oxford: Blackwell.

Vranas, Peter. B. M [2004]. "Hempel's ravens paradox: A lacuna in the standard Bayesian solution". *British Journal for the Philosophy of Science* **55**: 545–60.

Whiteley, Charles Henry [1945]. "Hempel's paradoxes of confirmation". *Mind* **54**: 156–8.