DE GRUYTER OPEN

STUDIES IN LOGIC, GRAMMAR AND RHETORIC 46 (59) 2016
DOI: 10.1515/slgr-2016-0036

David Botting
Universidade Nova de Lisboa

HIGH CONFIRMATION AND INDUCTIVE VALIDITY

Abstract. Does a high degree of confirmation make an inductive argument valid? I will argue that it depends on the kind of question to which the argument is meant to be providing an answer. We should distinguish inductive generalization from inductive extrapolation even in cases where they might appear to have the same answer, and also from confirmation of a hypothesis.

Keywords: probability, inductive generalization, inductive validity, confirmation.

1. Arguing with probabilities

Consider the following four arguments:

1. Every raven is black; this is a raven; therefore, this raven is black.
2. Every raven is black; this is a raven; therefore, likely this raven is black.
3. 9 out of 10 ravens are black; this is a raven; therefore, likely (with a probability of 0.9) this raven is black.
4. Every observed raven is black; therefore, likely every raven is black.

Are these arguments valid? I will argue that this depends on whether we read "every raven is black" as a universal material conditional or as a probability statement. However, only one reading seems correct in each instance. On the correct reading then, I will argue that the first three are deductively valid and the fourth is deductively invalid.

The main aim of this paper, then, is to consider whether the fourth argument is valid in some other sense (perhaps "inductively valid") or simply not a valid or good argument as it stands. It will turn out that this depends on which of two possible readings (4) is given, namely as an inductive generalization or as the inductive confirmation of a hypothesis. As a generalization I will argue that the argument is inductively valid, but as a confirmation it is not a good argument as it stands.

ISBN 978-83-7431-501-2 ISSN 0860-150X

Ostensibly, (1) should be uncontroversial. Surely, this is deductively valid if anything is. However, although I agree with this assessment I wish to point out that this requires reading "every raven is black" as a universal material conditional and not as a statement that the probability of ravens being black is 1.0. One could be forgiven for thinking that when the probability is 1.0 this amounts to same thing as saying "every" and when it is 0.0 this amounts to same thing as saying "none." But this is not so: the probability statement "every raven is black" can be true while the universal material conditional "every raven is black" is false, and conversely the probability statement "every raven is black" can be false while the universal material conditional "every raven is black" is true.

The probability statement "every raven is black" can be true while the universal material conditional "every raven is black" is false because negative instances will make the universal material conditional false but do not prevent a frequency series converging on 1.0 when the number of negative instances is finite or their relative frequency tends to zero with increasing sample size. So, on a frequency interpretation of probability it is logically possible for the premises of this argument to be true but the conclusion false – the relative frequency of black ravens may converge on 1.0 whether *this* raven (the raven referred to in the conclusion) is black or not. Reading "every raven is black" as a statement that the probability of ravens being black is 1.0, then, (1) is deductively invalid. We will see later that it is, perhaps surprisingly, still invalid even if there is a complete enumeration of ravens. This is because the probability statement, on the frequency interpretation, is not about the set of ravens as such.

The probability statement "every raven is black" can be false while the universal material conditional "every raven is black" is true because the latter is true when there are no ravens, i.e., when the class is empty. No frequencies can be associated with such classes, so a probability statement (which by definition is a statement about infinite classes on standard frequency views of probability) makes no sense when the class is empty. Perhaps we should not say that it is false; rather, it is meaningless. Similarly, before the frequency series has started – that is to say, before the first trial – it is impossible to assign any value at all to the probability, although here the probability statement would not be strictly meaningless.[1] This goes also for the set of ravens: the reason that the probability statement cannot be about the set of ravens as such is that the set of ravens existing at any time is finite and as such has no probability.

Let us make this clearer. A probability (relative frequency) statement is not a universal or general statement at all but a singular statement about two infinite sets, specifically a statement of the frequency ratio of instances of one to instances of the other of these sets. Now, the set of ravens is not an infinite but a finite set, and (in most relative frequency views of probability, though there have been exceptions) probability is undefined for finite (including empty) sets. Additionally, ravens are not events, so it makes little sense to refer to the "frequency" of ravens. Perhaps it might be objected that this shows that the frequency interpretation is just wrong, or is not applicable to this situation. However, what the probability statement "every raven is black" means in this situation is something like the following: imagine making a random choice from the set of ravens, determining that it is black, and replacing it. The two infinite sets here are the set of choosing ravens and the set of choosing black ravens (both of which are events), and the probability statement states that the frequency ratio of black ravens chosen to the number of ravens chosen will converge on 1.0. This probability statement is the ground of the statement of the universal material conditional, but the statements are by no means the same. The ground of the probability statement, in turn, is a statement about the frequency ratio in the current sample or just an enumeration of the current sample, which is not in itself a probability statement; we predict the ratio in the infinite set on the basis of the ratio in the current sample. The probability statement always describes a counterfactual situation for which we may have evidence but is never deductively entailed by that evidence, even in complete enumerations, because completely enumerating the set of ravens is not the same as completely enumerating the set of raven-choosings, and the latter set, being infinite, cannot be completely enumerated.

Coming now to (2), we should interpret (2) as a statistical syllogism on a par with (3), which is to say that "every raven is black" should be considered as a probability statement just as "9 out of 10 ravens are black" is considered in (3). Interpreted this way I will argue in a moment that (2) and (3) are deductively valid, but now I wish to consider the situation reading "every raven is black" as a universal material conditional. Is this deductively valid?

We have already seen that, because the probability statement can be true (i.e., the probability can be 1.0) despite the existence of negative instances, (1) was deductively invalid when "every raven is black" is read as a probability statement. Nevertheless, one might reason as follows: granted it is logically possible for the premises to be true but for the raven you choose not to be black. But the conclusion of (2) says only that it is *likely* to be black. Is it logically possible for the premises to be true but for it *not* to be *likely* that the raven you choose is black? It does not seem so (though I admit I am unsure of this point). I think that (2), then, is deductively valid however you read "every raven is black," although I would still maintain that the probability reading (whose deductive validity is yet to be shown) is the correct one in this case.

One might think that what can be said for (2) follows also for (3). Granted, in one out of ten cases the raven you choose will not be black, consequently it is logically possible for the premises to be true but for the raven you choose not to be black. But the conclusion of (3) says only that it is *likely* to be black. Is it logically possible for the premises to be true but for it not to be likely that the raven you choose is black? Using a weather forecast as an example, Weddle (1979, p. 3) does not think so; the argument

> stated a probabilistic connection between its premises and rain [the conclusion]. But the arguer only said that it was likely to rain. The connection between those premises and the likelihood of rain is not similarly probabilistic. We could not reasonably grant those premises ... and yet deny that it is likely to rain.

But this is not unreasonable, and it is logically possible for the premises to be true but for it to be unlikely that the raven you choose is black. This is because it is entirely possible for something to be highly probable relative to one reference class yet highly improbable relative to another. One cannot by "hedging" (Weddle's term) or qualifying the conclusion make it detachable unless one has all the evidence. Granted, Weddle does not say that all arguments like this will be deductively valid but only those made carefully and on the basis of sufficient evidence. The issue is really that this reading depends on additional information that we may not have. We may be able to fill in all the unexpressed premises so that the requirement of total evidence is satisfied, but there seems no reason to assume this and judging whether this is true seems as difficult as our original problem, and judging whether the arguer's evidence is really sufficient seems to lead us back to where we started.

Does this mean that (3) is deductively invalid? No: it means rather that we should not read the modal qualifier as qualifying the conclusion. Instead, it qualifies the illative: (3) says that the premises (or body of evidence) make likely or 'probabilify' the conclusion. There is no detachment here, only a certain relation to the premises. On this reading I have been asking the wrong question. The right question is: "Is it possible for the probability relation between these premises and this conclusion to be anything other than 0.9?" The answer to this seems to be "No." We can change the probability by adding premises, of course, but this is just to change the relation. Reading the "likely" in this way, then, does seem to make (3) deductively valid.[2]

I suggest that we read (2) and (3) in the same way, that is to say, as containing a probability statement where the probability given in the content of this statement becomes attached to the illative. Interpreted this way, all statistical syllogisms are deductively valid. On the interpretation where the conclusion is modally qualified the statistical syllogism is not valid as it stands, either deductively or inductively. This is not to say that we do not make modal statements about whether this raven is black or whether it will rain tomorrow. In fact, we bet on such occurrences. This illustrates one important difference between inductive generalization and inductive extrapolation (i.e., inference to the next instance): generalization depends, so I will argue, on the evidence and nothing else, but extrapolation and confirmation of hypotheses does not.

2. History of the dispute

Consider (4) again: "Every observed raven is black, therefore likely every raven is black." Is this argument valid?

Is it deductively valid? It could be argued that this is in fact deductively valid, on grounds similar to those we have seen. The reasoning is more or less as before: granted, it is not logically impossible for the premises to be true but for it to be false that every raven is black, but what the conclusion actually says is only that it is *likely* that every raven is black. So the relevant question is whether it is possible for the premises to be true but for it not to be likely that every raven is black, and if the answer to this is "No" then the argument is deductively valid.

However, the answer to this question too is actually "Yes." The important point here is that the "likely" is being read as qualifying the conclusion, and a likely conclusion can only be validly detached when the premises contain all the relevant information, as we have just said. Suppose that it is only one particular kind of raven that you have observed, or only in one particular place. Under these circumstances it would be in no way inconsistent to deny that it was likely that every raven is black. One may have little evidence for saying this, but certainly there is no logical inconsistency in saying it.

Let me put this another way that may seem slightly surprising and is more suited to the problem we started from: given that all observed ravens have been black, it is rational to generalize that every raven is black but it is not necessarily rational to suppose that the next raven we observe will be black. That is to say that I do think the argument "Every observed raven is black, therefore likely every raven is black" to be valid. It is not deductively valid for we are not here dealing with a statistical syllogism, but if someone wants to describe generalizations with high frequency ratios as "inductively valid" then I have no objection. The main thing is that it is a good argument. "Every observed raven is black, therefore likely this (unobserved) raven is black," on the other hand, is not valid unless more is known about the representativeness of the sample. I will argue that there is a way of reading (4) – as the confirmation of a hypothesis – that is also invalid for the same reason as the extrapolation. In fact, it will turn out to be questionable whether this can be made valid even with the addition of information about the sample; I will try to argue that there is one kind of scenario in which it is valid and another in which it is not.

The dispute whether or not this argument is good has a considerable history. Most recently it has appeared in two articles in *Studies in Logic, Grammar and Rhetoric*. Bermejo-Luque (2009, p. 299) says that this is an inductively valid argument, Szymanek (2014, pp. 237–8) denies it. Szymanek cites Hitchcock (1999) who in turn is commenting on a dispute between Thomas and Nolt (Nolt 1984; Nolt 1985; Thomas 1985). An argument similar (though not quite the same) to the one that Hitchcock provides, and that Szymanek more or less borrows, was actually also made by Popper (1962) against Carnap's (1952) inductive methods and even against instance-confirmation as such. Bermejo-Luque and Thomas think (4) is a good argument,[3] while Szymanek, Hitchcock, Nolt and Popper think otherwise.

3. The case against

a) Szymanek's argument

Szymanek (2014, pp. 237–38) says that the argument's goodness depends on the receiver's set of beliefs. He illustrates this with an example. Imagine an urn containing 100 balls, each of which may be either black or white. We draw 99 balls at random, leaving one in the urn. All the balls we draw have been black. "Can we say that it is likely that the last ball remaining in the urn is black?" Szymanek (2014, p. 237) asks, before answering "... the probability of the last ball being black cannot be calculated on the basis of the presented data. ... [I]t is necessary to know the *a priori* probability of the last ball in the urn being black (or the *a priori* probability of various sets of balls in the urn)." "Each of the balls drawn were black, therefore the last ball drawn will be black" is not inductively valid, and by parity of reasoning "Every observed raven is black, therefore likely every raven is black" is likewise invalid (Szymanek 2014, p. 238).[4] The point is that if the prior probability of all the balls in the bag being black is low enough then the posterior probability may not be raised high enough by the empirical evidence for this to be an inductively valid argument.

b) Hitchcock's argument

Hitchcock's urn contains only 50 balls and each ball is either blue or not, but otherwise the example is much the same. Is the argument "There were fifty balls in the urn; the first 49, drawn at random, were all blue; therefore, probably, the remaining ball is blue" valid? He says not.

Hitchcock starts by saying that no relative frequency interpretation of probability is applicable. Instead the probability is to be construed as epistemic probability. Because there is only one ball left, and if this is blue it means that all the balls in the urn were blue, he then says that the epistemic probability of the next ball being blue will be the same as that of every ball's being blue (Hitchcock 1999, pp. 202–203); when there is only one ball left, extrapolation and generalization collapse into each other. This step was missing in Szymanek's argument, who assumed too quickly that an extrapolation ("Each of the balls drawn were black, therefore the last ball drawn will be black") was on a par with a generalization ("Every observed ball is black, therefore likely every ball is black").

Hitchcock goes on to prove it by appeal to conditional probabilities. Hitchcock's claim amounts to the claim that p(remaining ball is blue|E) = p(all 50 balls are blue|E), which amounts (with the conditional cashed out) as the claim that p(remaining ball is blue & E)/p(E) = p(all 50 balls are blue & E)/p(E). Presumably E here is the set of premises "There were fifty balls in the urn; the first 49, drawn at random, were all blue." Conjoined with "The remaining ball is blue" this entails that "All 50 balls are blue," and hence that "The remaining ball is blue & E" is logically equivalent to "All 50 balls are blue and E." Since logically equivalent propositions have the same probability, this means that p(remaining ball is blue & E) = p(all 50 balls are blue & E), and since both are divided by p(E) to get the conditional probability, it follows that the conditional probabilities are equal also, and so the probability of the generalization being true given the premises is the same as the probability of the extrapolation being true given the premises (Hitchcock 1999, p. 203). He then uses Bayes Theorem to calculate what this probability is, ending up with

$$p(50 \text{ blue balls}|\text{first } 49 \text{ blue balls } \& \text{ K}) = 50/(50 + r^*)$$

where r^* is the inverse ratio of the prior probability that 49 balls are blue to the prior probability that 50 balls are blue (Hitchcock 1999, p. 206). When these priors are not among the information we are given, as in Thomas (1994), then we should not say that generalization is a good argument or likely to be true, no matter how well confirmed it is, that is to say, no matter how often we draw blue balls. As Szymanek (2014) objects to Bermejo-Luque, it depends on the beliefs of the receiver, specifically of these prior probabilities, without which the probability is simply impossible to calculate. When the priors are added, the inductive validity of the argument depends on the calculated value of the probability: the argument is valid if this probability is high.[5]

Hitchcock (1999, p. 208) notes some of the surprising results of this: "Since the result we observed was inevitable on the first hypothesis, but highly unlikely on the only alternative hypothesis consistent with our evidence, does this result not make it highly probable that the first hypothesis was true ...?" This reasoning is invalid, Hitchcock says, again because it ignores the priors. But this seems to rule out at the same time a kind of reasoning that is often used in certain cases of inference to the best explanation. Where we have several hypotheses that are consistent with the evidence, we normally think ourselves entitled to select the hypothesis for which the evidence, especially surprising evidence, becomes highly likely or even inevitable – the "empirical adequacy" is at least one factor in deciding which hypothesis provides the best explanation, and this in turn is taken as indicative of truth. But if Hitchcock is right it seems that it never is, unless you have the relevant priors. Hitchcock (1999, p. 209) seems ready to accept this.

c) Popper's argument

Instead of an urn Popper has two bags: one with twenty white balls and one with nineteen white balls and one black ball, but you do not know which bag is which. Consider now the situation where we draw balls from only one of the bags but we do not know whether this bag is the one with the black ball. We draw nineteen white balls from the bag. It is clear that the probability of the twentieth being black is still 0.5, quite irrespective of the evidence of previous drawings. Popper (1962) argues as follows (I have adapted his example slightly): suppose that all the balls except one have been drawn and have all been found to be white. Take a as the name of the statement "All the balls in our selected bag are white" and b as the name of the statement "Precisely one ball in our selected bag is black." What is the probability of the statement a, given the information a or b, or equivalently, what is the probability of all balls being white, given that either all are white or exactly one is black? The formula for this is

$$\mathrm{p(a,\ a\ or\ b) = p(a)/(p(a) + p(b) - p(ab))}$$

Since a and b are incompatible their intersection must be null, i.e., p(ab) = 0. This gives p(a, a or b) = 0.5/(0.5 + 0.5) = 0.5/1.0 = 0.5. In other words, however many white balls you draw from the bag you have no better reason to think that all the balls in the bag are white than a mere guess; it depends only on what bag you chose at the start.

The conclusions Popper (1962, p. 71) draws are dramatic:

> Our formula does not refer to [the size of the sample]. It is, therefore, valid for [samples] of all sizes (and it may even be extended to infinite samples). It shows that, even on the assumption that we have checked all [members] of a sequence (even an infinite sequence) except, say, the first, and found that they all have the property A, this information does not raise the probability that the as yet unknown first [member] has the property A rather than B.
>
> Thus the probability of a universal law-such as the statement a-remains, on our assumptions ... equal to f = 1/2 even if the number of supporting observations becomes infinite, provided one case of probability r = 1/2 remains unobserved.

Here we do know what Szymanek referred to as "the *a priori* probability of various sets of balls in the urn," figuratively speaking: there are two possible sets, and each is equally probable. Popper's argument goes further than Szymamek's and Hitchcock's, for whereas they at least allowed for Bayesian updating of the probability of all the balls drawn being of the same colour, Popper says that the probability is the *a priori* probability set up by the initial conditions and never changes from that; it is invalid reasoning to suppose that because you have drawn only white balls so far this makes it more likely that all the balls in our selected bag are white.

This is similar to Hitchcock's contention that because what you have observed is highly likely on one hypothesis and less likely or unlikely (though consistent with) another, this is any reason in itself to suppose that the first is more likely to be true. Popper's contention seems to be more extreme in its ramifications. Hitchcock does at least allow for highly confirmed hypotheses to follow validly under circumstances where the priors are known, but on Popper's view the priors are all there is and confirmation by empirical data is beside the point; which of the hypotheses is true is analogous to which bag we are picking from, and just because the data is what we would expect if the hypothesis were true there is no greater likelihood of its being true. This seems to suggest that this kind of inference to the best explanation is just an invalid form of inference.

There is a similar argument in Weisberg (2011, pp. 506–507). He considers a sequence of ten coin-tosses. Applying a Principle of Indifference seems to tell us that we should consider each sequence equally likely, that is to say, a sequence in which heads comes up ten times in a row is as likely as a sequence such as HTTTHHTTTH. Consider now the position after nine coin tosses which have all come up heads. There are now two possible sequences that are consistent with the coin tosses so far, namely the sequence where the next toss is a heads and the one where it is a tails. Indifference seems to tell us that these are equally probable, but Weisberg notes the anti-inductivism of this result, for it means, as we will see Carnap claiming, that one does not learn by experience. Weisberg takes this as reason to reject the Principle of Indifference, but it is not clear how much the Principle of Indifference itself is essential to this argument, for we may arrange our priors in some completely arbitrary (though probabilistically consistent) way and the result will be that only the prior probability really counts, as Popper argues; the probability that the last ball in the bag is white depends on our initial choice of bag, for we only ever pick from the same bag, and the bag was chosen before we had any evidence. Similarly with scientific hypotheses, which hypothesis is true being analogous to which bag we are choosing from (although in this case there is no actual choice as such): suppose that we know that one and only one of two hypotheses is universally true. Is it the case that evidence that is logically consistent with both hypotheses can nonetheless favour one hypothesis and make it (by conditionalizing on the evidence) more likely than the other? This argument suggests not.

4. The case in favour

I would like to make a couple of preliminary remarks first. Firstly, I disagree with Hitchcock that the probability referred to has to be construed as an epistemic probability and cannot be construed as a relative frequency, or other kind of probability. In Section 1 I explained the way it can be seen as a relative frequency as well (as sampling under replacement), and relative frequencies are typically derived by applying the Straight Rule.[6]

I have no objection to them being treated as epistemic probabilities as well, however. This epistemic probability should not be identified with our confidence that the frequency statement is true; although this confidence also is an epistemic probability, its value is not the same as the observed frequency. The epistemic probability in question belongs instead to either the next instance or to a conditional like "If this is a raven, then it is black," and if the relative frequency is 9 out of 10 as in (3), then this means that 9 out of 10 substitution-instances of this conditional (with true antecedents) will be true. It is the probabilistic influence of being a raven on being black.

I would distinguish two versions of (4), hinted at earlier. In the first (4) expresses generalization itself, and as such it is clear that the conclusion follows from the premise by the Straight Rule. In the second we are making a comment about the probability statement; we have made such a statement, and the premises are now confirming that this statement is true. From what was said above it should be clear that every raven is black can be the correct probability statement and, because of our doubts about the sample, for our confidence that this probability statement is true to be low, and if what we have just said about the validity of inference to the best explanation is true, then we can never be confident unless blessed with favourable priors, and perhaps not even then. That this confidence is not the epistemic probability that was just mentioned is clear from the fact that that epistemic probability was determined solely by the evidence, whereas our confidence is determined by considerations of the quality of the evidence, such as the variety of instances. It is the weight of evidence; it has its psychological indicator in confidence, but it is not itself subjective but has an objective and rational basis.

Unlike the other arguments that place the problem in the context of Bayes Rule, in Popper's argument it seems to be the Straight Rule that is informing the discussion. As we will soon see, Carnap also rejects the Straight Rule, though for different reasons. I will give a partial defence of the Straight Rule, and as we have seen once this is done then (4) – considered as a generalization – is a good argument. In contrast, considering (4) as a confirmation, its goodness is (as we have just seen) extremely problematic.

Let us look at Popper's argument in more detail. Popper's method of working out the probabilities amounts, in Carnap's λ-system, to $\lambda = \infty$ or in other words giving weight to the 'logical factor' (where a Principle of Indifference determines that all logical possibilities – or in Carnap's modified version, all structure-descriptions – are given equal probabilities) and zero weight to the 'empirical factor'. Carnap (1952, p. 38) rejects this method because it has as a consequence (here apparently embraced by Popper) that we do not learn anything from experience, and he uses instead the rule $P = (s_A + 1)/(s + 2)$ where s is the number of balls in a sample and s_A is the number of these that are white. If we have examined every ball in the population apart from one (i.e., $s_A = n - 1 = 20 - 1 = 19$) then this reduces to $P = n/(n + 1) = 20/21$ (Popper 1962, p. 72). Carnap does not use the Straight Rule, then, and in defending the Straight Rule I will also have Carnap to contend with as well as Popper.

One curious thing about Popper's paper is that although he takes himself to be talking about singular predictive inferences (i.e., extrapolations) (Popper 1962, p. 69 & p. 71) a is not a singular statement at all but a general one, and in his conclusion is said to be a universal law. He seems to take this to be equivalent to a singular statement in the case where every member of the population except one has been observed. Since observing the only as yet unobserved member will complete the enumeration he takes the probability of its being A as the probability that they are all A (Popper 1962, p. 70). Szymanek also takes this for granted. Only Hitchcock provides an argument. But I deny this: it is an error born of reading the generalization "All fifty balls are blue" not as a probability statement but as a universal material conditional (the "fifty" is actually redundant here). Nor does it follow without further assumptions that because the universal material conditional is non-vacuously true the probability must be 1.0, for the probability statement concerns ball-drawings and not balls, and completely enumerating the balls (which is what the universal material conditional amounts to) does not amount to completely enumerating the ball-drawings – the ball-drawings referred to in the probability statement can never be completely enumerated, for there is by definition an infinite number of them. Only if we assume, for instance, that the balls can never change colour, and things like this, can we infer with complete confidence that the probability is 1.0. The slightly peculiar, though nonetheless valid, upshot of this is that complete enumeration of the balls entails the universal material conditional but does not entail the probability statement, though it does provide grounds for it that are as good as you are going to get. In other words, the argument

Ball 1 is A
⋮
Ball n is A
<u>There are n balls</u>
All the balls are A

is deductively valid[7] when the conclusion is read as the universal material conditional (there being no way in which the premises can be true and the conclusion false) but deductively invalid when the conclusion is read as the probability statement giving the frequency ratio of 1.0 (for the premises can be true but the conclusion false, for the conclusion is making a statement about ball-drawings and the premises do not say anything about ball-drawings as such, though they provide evidence for statements about what ball-drawings would be made under the counterfactual conditions described earlier).

The question "What is the probability of all balls being white, given that either all are white or exactly one is black?" cashed out probabilistically means "What is the frequency with which I would draw white balls if I were to draw a ball, replace it, and repeat this ad infinitum, given that all the balls drawn so far have been white but one may be black?" We are being asked to provide a generalization from the data, and the answer to this is given, I would maintain, by the Straight Rule: if I have drawn 19 white balls from 19 attempts the probability in the generalization should be 1.0. The mere possibility of drawing a black ball does not, I maintain, alter the value of this probability. This probability, then, depends only on the evidence, that is to say, on the frequency series we have found by sampling.

This has to be distinguished from two other questions that it might be confused with, especially when the probability is 1.0. We have already seen these, but they are worth emphasizing. Here is the first: how well is "every ball is white" confirmed? When we have a confirming instance we can see that confirmation in two ways, perhaps even within the same inductive process. Suppose we start off not knowing what colour ravens are and start observing ravens. The relative frequency of (observed) black ravens to (observed) ravens is 1.0. Then there might come a point where we decide to make "every raven is black" a hypothesis, and then we tend to see the observations as confirming this hypothesis. In a sense the observation does double duty. Here is the second question: what is the probability that the next ball I draw will be white or the next raven I observe will be black? Generalization should not be confused with extrapolation, even if there is only one ball left.

Let us recapitulate. We are concerned with the question of whether (4) is a good argument. It is not deductively valid, so the question is whether it is inductively valid. I proposed that we read (4) "Every observed raven is black, therefore likely every raven is black" as an inductive generalization that generalizes from observed ravens that are black to all ravens being black, and further suggest that this should be cashed out as "Every observed raven is black, therefore the frequency with which I would draw black ravens if I were to draw a raven from the population of ravens, replace it, and repeat this ad infinitum, given that all the ravens drawn so far have been black, is high." The generalization, and equivalently the argument in (4), is as inductively strong as this frequency is high (and in this case it is 1.0, as indicated by "every"). That is to say that the generalization itself, or what I will be calling the probability of the hypothesis, depends entirely on the evidence, and prior to any evidence no probability can be attributed. If I have only drawn black ravens then the probability of the hypothesis (assuming the Straight Rule) is 1.0. It depends only on the empirical factor and not the logical factor. The logical possibility of a non-black raven means only that if we are asked to bet on "every raven is black" being true then we might not do so. We may not be at all certain that the probability is 1.0, but this does not mean that we believe that the probability is other than 1.0 or that this is not the answer we would give for the probability of the hypothesis if we were forced to give one. When the sample size is still small we might not be very confident at all that "every raven is black" but this makes no difference to the generalization itself. Conversely, the generalization does not change in this case whether you have made one observation or a thousand – the frequency ratio never shifts from 1.0. Obviously, it would be wrong to suppose that having a larger sample size makes no difference in this case; in the next section it will be seen that because the Straight Rule implies that one cannot learn from experience, Carnap adduces this as a reason that it must be false. What these confirmations give you is confidence that the frequency ratio has reached its final value, that is to say, that the frequency series has converged on its limit.

Suppose now that we are asked to bet that a hypothesis is true. Having drawn the white ball, we do not bet that the proportion of white balls to black balls in the bag is 0.975. Indeed, we know that this hypothesis is false because it would amount to something like one ball being half-black and half-white! The difference made by the presence of the additional bag is that we will require greater inducement to be forced to bet at all; we do not give the evidence great weight, in the sense of being confident that the hypothesis is true, i.e., that the stated probability is the real probability.

So, we should beware of confusing the probability value of the generalization (i.e., the probability of the hypothesis) and the weight of evidence. The probability value of the generalization is always given by the Straight Rule. Carnap's preferred rule referred to in Popper's example, being different from the Straight Rule, is, I think, wrong. However, Carnap is right in so far as a satisfactory answer to the question of the probability of the next instance requires some rule involving both a logical and an empirical factor. This was the second question that I said above might be confused with the question of what was the probability of the hypothesis. First of all, we need to look at the first question that I said above it might be confused with, viz., as inductive confirmation.

This is where we have got to: (4) is an inductively valid argument on the suggested reading. Is it inductively valid on the alternative reading of (4) as a confirmation of a hypothesis?

I think that there are two kinds of situations that we need to consider here that the previous discussion blurred. That discussion tended to treat the hypotheses being confirmed as mutually exclusive, i.e., only one scientific explanation of the same empirical data can be true, and only one bag can be chosen from, which is to say that the different possible distributions of balls in the bag being considered were only epistemic alternatives rather than real alternatives. In this situation Popper's argument seems valid – confirmation of a hypothesis (that is to say, making observations that are consistent with its being true) does not make that hypothesis more likely to be true. If we add to (4) unexpressed premises in the form of prior probabilities then we may get an inductively valid argument, but only because the prior probabilities themselves, without any contribution from the empirical data, are high;[8] "All observed ravens are black" would be a redundant premise in this situation. Since this is the situation that Hitchcock and Szymanek discuss as well, I think that they are wrong to imply that "All observed ravens are black" and statements of the prior probabilities together and without redundancy constitute an inductively valid argument.

However, there is another kind of situation where both hypotheses are true, and further special cases of this where both may have a probabilistic influence on the sample. In such cases I think it does make sense to treat the inverse probability as a confirmation measure. This, I think, is because there is an objective chance of the other alternative being responsible for the particular data. True, in a sense there was an objective chance of the alternative in the two urn case because there really are two urns, and not just one urn with two epistemically possible distributions of balls, but once the first ball was drawn there was no objective chance that any further ball was drawn from the other urn for we have made the determination to draw only from that urn. It is this, rather than inductive confirmation as such, that runs Popper's argument, and although it is a plausible model of the kind of situations Popper was most concerned with, it is not a plausible model of situations of the following kind: suppose that we do not know what urn we are drawing from in any particular drawing, and we draw 19 white balls as before. Objectively, we could have been drawing from either urn, or even drawing some balls from one of the urns and some from another. It seems now that drawing 19 white balls is more likely on the hypothesis that we are drawing from the urn containing 20 white balls than on either of the alternative hypotheses. Now, I need to avoid a possible confusion between two sets of hypotheses. The probability statements, each representing the distribution of balls in a specific urn, express true hypotheses about the world, or so we suppose, for otherwise there would be no objective chance of drawing from that urn, making that urn an epistemic alternative only. It is a different set of hypotheses that we are talking about here, namely the hypotheses that a particular sample is drawn from a particular urn, and in saying that one hypothesis is more likely to be true than the other hypotheses we are saying only that the other urns are less likely to be responsible for the empirical data and not that the probability statements are false, since (by supposition) they are true. We can tentatively draw conclusions about what population the sample is a sample of and what the probability distribution within that population is.

In summation, reading (4) now as a confirmation and as having the prior probabilities[9] as unexpressed premises, I think that it is inductively valid in cases where there are objective chances of the alternatives being responsible for any particular datum or data-set, but invalid in cases where there is no such objective chance and only one out of a set of candidate hypotheses may be true and hence responsible for the data. This is unfortunate for inference to the best explanation.

Now we need to consider cases of inference to the next instance. Consider the case where you know that there are both white balls and black balls but have not yet drawn one, and consider the question "What is the probability that the next ball that you draw will be white?" As Carnap intimates, here the logical and empirical factors do seem to work together. The logical factor does not favour the choice of one bag over another, and since you do not know the distributions in the bags, it does not favour a white ball over a black ball either. Also, since no balls have been drawn, there is no frequency series and thus no probability to which the empirical factor could be applied. If the probability of the next instance and the probability of the hypothesis were identical, it would not be possible to assign any rational quotient to the bet in question, for no probability of the hypothesis can be assigned prior to a trial. However, because the probability of the next instance depends partly on the logical factor (as the probability of the hypothesis does not) this logically determined probability is the rational quotient in this situation, which is to say that it would be rational to spread one's bets evenly between drawing a white ball and drawing a black ball.

You begin to amass evidence by drawing balls. The logical factor decreases as the sample size increases since the weight of evidence that the bag being drawn from is the one with the black ball decreases as one draws white balls, since if one had been drawing from the bag that had the black ball then the probability that one would have drawn it by now increases as the number of drawings increases; therefore, the fact that one has not drawn the black ball is evidence against it being the bag with the black ball and I become less willing to bet on this hypothesis. But let us suppose that you believe the sample you have taken so far not to be representative. Then, although as the sample size increases (and the probability that the sample is biased decreases) more and more weight will be put onto the empirical factor and less and less weight will be put on the logical factor, you will not necessarily eliminate the logical factor completely until it is believed that the frequency series has converged, and possibly not even then if one believes that every logical possibility has some non-zero probability. For extrapolations, then, the empirical and logical factors seem to combine directly.

When two hypotheses may be responsible for the next datum, we have to multiply the weight of evidence by the probability of the hypothesis for each hypothesis and add them together, and if there is no such probability as yet, we can apply a logical factor to a priori probabilities. So, the probability for the next instance in the two bag example could be 0.975,[10] though this is not a possible value for the probability referred to in "Likely all the balls in the bag are black." However, because we have a logical factor here as well we do not necessarily need to treat the hypotheses as equally likely and might be able to find a priori reasons (e.g., simplicity, where this might mean only the number of qualitatively distinct objects posited by the hypothesis) to weigh the hypotheses so that one (e.g., the simpler) is favoured over the other.[11]

However, when only one hypothesis may be responsible for the next datum (as was the case in our examples, when only one distribution of balls was actually responsible for the data though we did not know what that distribution was) then the extrapolation can be based only on that hypothesis. In Popper's example, since we may have chosen from either bag at the start, the probability for the next instance is either 1.0 or 0.95 and cannot be anything else. The problem is that we do not know which, but it would be wrong to conclude from that that it is, as again we might be tempted to think, 0.975. If we have the option of spreading our bets over both hypotheses, though, it is true that we should take 0.975 as the rational betting quotient. But it is not the probability of the next instance.

Let us go back to (4) now. The conclusion is "every raven is black." How should this statement be interpreted? We could take it as the universal material conditional and the evidence contained in the premises as confirmations that this statement is true. I think that Hitchcock and Szymanek take it this way, and taken that way their criticisms are probably sound and may not even go far enough. Taken instead as a generalization the argument merely says that the probability value is the limit in the observed instances, where these are reported in the premises. As long as there is one instance a limit and thus a value for this generalization may be posited. (4), then, is basically a corollary of the Straight Rule, and if the Straight Rule is sound then I believe we can say that the argument is inductively valid.

It might be argued that this is much too permissive, and in a sense it is, for the argument is as much valid for small samples as for large ones, and when the probability is 1.0 the value will (generally) always be the same whatever the sample size. But this, I think, is because there are assertibility conditions on the argument. Asserting a generalization amounts to a certain extent on betting on it; we cannot sincerely assert that "every raven is black" unless we feel confident about our sample, that the frequency series has reached its point of convergence, etc. This is especially so when we are asserting this as a universal material conditional, but also as a probability statement. All the considerations raised above concerning the weight of evidence come home to roost. The situation is not entirely dissimilar to being asked a question when we do not know whether the presuppositions of the question are satisfied. We do not know whether the series will converge, whether the predicate is projectible, etc. An assertion cannot be expected in reply to such a question or request for a probability, and in such a case I do not think the question can be said to transfer the burden of proof. But if a reply is forced from us the best we can do is to give the limit so far, though this would not qualify in this case as an assertion.

5. Defending the Straight Rule

What I have said so far regarding the inductive validity of (4) depends, then, on whether the Straight Rule is itself a good rule. We have seen how to avoid Popper's conclusion that you cannot learn from experience; Popper confuses the generalization with extrapolation and with confirming the hypothesis. However, interestingly Carnap criticizes the Straight Rule on the same grounds but for different reasons. The Straight Rule says "Always take the practical limit" as the limit of the infinite series irrespective of the sample size. The practical limit after drawing two white balls from the bag is 1.0, and the practical limit after drawing twenty white balls from the bag is 1.0. This also suggests that we would be prepared to accept exactly the same betting odds after two drawings as after twenty (Carnap 1952, pp. 42–43). We do not seem to have learnt anything after the twentieth drawing that we didn't already know after the second or the first.

This is basically the paradox of ideal evidence. Positive evidence, we would normally say, should raise the probability of the hypothesis being true. But here the probability of the hypothesis is set at 1.0 from the start and never moves. The usual response in the case of ideal evidence is basically the same as what I have called weight of evidence – what increases with sample size is willingness to bet or (what is really the same) our confidence that the sample is representative. This is indicated in Carnap's "prepared to accept," which Carnap does not seem to notice may come in degrees, where these degrees are a separate thing from the betting odds themselves, that is to say, I may be more or less prepared to accept 1:1 as betting odds, but if the latter this does not mean that other betting odds would be more acceptable, for this again would be confusing weight of evidence with the probability of the hypothesis.

Similarly Carnap (1952, p. 42) gives the following scenario as leading the Straight Rule into counter-intuitive results. Suppose that we have observed instances of P1 and instances of P2. These are such that their conjunction is logically possible but have never yet been observed. The Straight Rule says that the probability of something's being both P1 and P2 is 0.0 because it has never been observed, but Carnap objects that something that is logically compatible with the evidence cannot have a probability of 0.0. My response is that we should not infer that the probability that this will be observed in the next instance is 0.0 (because of the logical factor that is operative there), but there is nothing counter-intuitive in inferring that the probability of the hypothesis is 0.0.

I will put Carnap's (1952, p.43) next scenario in his own words:

> If all we know about the universe is that the only observed thing is M, then the method of the straight rule leads us to assume with inductive certainty all things are M and to take the estimate of the rf [relative frequency] of M in the universe to be 1. ... Thus, on the basis of the observation of just one thing, which was found to be a black dog, the method of the straight rule declares a betting quotient of 1 to be fair for a bet on the prediction that the next thing will again be a black dog and likewise for a bet on the prediction that all things without a single exception will be a black dog ... Thus, this method tells you that that if you bet ... on either of the two predictions mentioned, while the stake of your partner is 0, then this is a fair bet.

Carnap concludes that this is obviously not a fair bet, and any method that tells you it is must be wrong as a consequence.

It is Carnap this time who confuses the question concerning generalizing itself and that of the singular predictive inference, i.e., extrapolation. Carnap is right to say that the Straight Rule is inadequate for the latter purposes which I agree require some kind of logical factor, but it does not follow, as he seems to say that it does, that the Straight Rule is inadequate for determining the empirical factor, and the empirical factor is what we have to give as what the probability of Z's being Y is given that they are Z – e.g., the probability of ravens being black given that they are ravens – when we are forced to give one.[12] When the weight of evidence is low, we would prefer not to give one and prefer not to bet; even though on this reading "Every observed raven is black, therefore likely every raven is black" is inductively valid, I have suggested that it may nevertheless not be assertible unless the weight of evidence is high.

6. Conclusion

Is "Every observed raven is black, therefore likely every raven is black" valid?

It is not deductively valid: it is possible for the premise to be true and conclusion to be false, even when we "hedge" and give a modal reading to the "likely." Nor is it deductively valid in the way that I have argued statistical syllogisms are deductively valid: there we are reasoning deductively from a probability statement, whereas here we are reasoning inductively from an evidence statement – a simple enumeration – and drawing a conclusion about what the probability in the population is. The presence of the word 'observed' is significant and marks this distinction; in (1), (2), and (3) there is no such qualification.

Is it inductively valid? Or more simply: is it a good argument?

This depends, I think, on whether we treat the argument as an example of inductive generalization or as confirmation of a hypothesis. As a confirmation it is not generally valid, and here it does not seem to matter whether we read "every raven is black" in the conclusion as a universal material conditional or as a probability statement. Szymanek and Hitchcock say that it is invalid when these are all the premises available, but when we can fill in unexpressed premises with the priors, it is implied that it can be inductively valid. But if Popper is right then this is never a valid argument; it is simply irrelevant, as far as he is concerned, that every observed raven is black.

If it is nonetheless likely that every raven is black, that is entirely in virtue of the priors alone. I have tried to distinguish between two kinds of situation and argued that in the second of them Szymanek and Hitchcock would be right and that the argument is inductively valid. The situations Szymanek and Hitchcock themselves discuss are not of this second kind, however, but are analogous to that Popper discusses. So, although I think that Popper is wrong that this is *never* a valid argument, I think he is right in the kinds of situations he discusses, and Szymanek and Hitchcock are wrong to consider arguments, once completed, to be valid in these situations, which they seem to do.

However, I am not quite sure that Szymanek and Hitchcock have proved quite what they need to. Even if we do need the prior probabilities, this does not necessarily mean that they are unexpressed premises of (4). I think Bermejo-Luque and Thomas may be able to answer by incorporating the prior probabilities into a sense of how 'carefully' (to borrow Weddle's phrase) we make the inference. These priors are conditions that must be satisfied for us to draw the conclusion from the data, yet it is from the data (of observed ravens) and not from anything else that the conclusion is drawn – concerns about priors and about representativeness of samples are comments on the data and are not part of the data. At least, I think that this is the kind of thing that Bermejo-Luque might say.[13] Also, she might say that as long as these conditions *are* satisfied the argument is good even if you do not *know* that they are satisfied and consequently do not include them among the premises, although the argument may not be assertible.

However, when treated instead as an example of inductive generalization, the argument does seem to be good. In fact, the argument "Every observed raven is black, therefore every raven is black" is also valid, because here "every raven is black" means that the probability is 1.0 and the 'likely' is redundant here, though helpful perhaps to indicate that the conclusion is a probability statement and not the universal material conditional. If we wish to *assert* the probability statement, or make a modally qualified statement, then we need to consider the weight of evidence, but this does not affect the goodness of the argument itself, or render it permissible to adjust the probability value from the limit as it currently stands, irrespective or whether that is the limit after one million trials or after one trial. Thus, the argument is still good even if only one raven has been observed. It is not good, though, if no ravens have been observed, that is to say, if the universal material conditional "every observed raven is black" is vacuously true. Here there is simply no probability to be had, although

I would say that you could still have an epistemic probability for the next instance in virtue of *a priori* factors.

Acknowledgements
FCT-funded individual post-doctoral project SFRH/BPD/77687/2011 and FCT-funded institutional project in ArgLab PTDC/FIL-FIL/110117/2009

NOTES

[1] Question: what is the probability of heads coming up in a series of coin tosses, before the first toss? Answer: there is no such probability and it is impossible to attribute one. The answer is not, as one might tend to think, 0.5. Question: what is the probability that heads will come up *the next time* the coin is tossed? Answer: here I think it does make sense to exploit the fact that we know there are two possible outcomes, and when this is all we know the rational choice is to bet equally (supposing that we have to bet at all) on each possibility; in other words, we would in this case attribute the probability 0.5 to the next instance. This is one reason why, as I will explore later, the probability of a series or sample can differ from the probability of the next instance (and incidentally why I think that singular probabilities cannot be considered to be disguised statements of the probability in the series).

[2] I deny the common thought, then, that wherever the conclusion does not follow with certainty from the premises the argument is non-deductive. It must only follow to the extent given it by the premises, and this extent (whatever it is) does follow with certainty and in virtue of the truth-values of the propositions.

[3] Carnap takes a different view as we will see later.

[4] High confirmation is still relevant, but how large a role should be allowed to this kind of empirical adequacy is unclear; he rightly remarks that it is outside the scope of his paper, which is a discussion of Bermejo-Luque's paper.

[5] Hitchcock then wonders if we might supply prior probabilities on a priori grounds by partitioning over all the logical alternatives, in this case all the possible colours the ball might be, but soon abandons the attempt as useless (Hitchcock 1999, pp. 203–207) because there is no unique way to partition the alternatives. Although Hitchcock does not mention this explicitly it can make a big difference to the probabilities how the alternatives are described. Is non-blue, for example, to be considered as one alternative and equiprobable with blue, or is it lots of different colours, each of which is equiprobable with blue, in which case the conditional probability of the last ball being blue is low? This kind of problem plagues Carnap's continuum of inductive methods.

[6] The complications of defending the Straight Rule as *the* rule will be swept under the carpet here.

[7] True, it is not formally valid as it stands, but it is a trivial exercise to make it so.

[8] If the prior probabilities are assigned using a priori using a Principle of Indifference, it is difficult to see how they ever could be higher than 0.5 for there must always be at least one epistemic alternative and dividing the probability space equally over both epistemic alternatives will attribute a probability of 0.5 to each. In consequence, sampling is a waste of time and you could do just as well by guessing.

[9] Without those prior probabilities I think that Szymanek and Hitchcock are right to say that it is invalid.

[10] This is $(0.5\cdot 1.0)+(0.5\cdot 0.95)$, i.e., the probability of choosing the bag with all blue balls (viz., 0.5) multiplied by the probability of choosing a blue ball from that bag (viz., 1.0) plus the probability of choosing the other bag (viz. 0.5) multiplied by the probability of choosing a blue ball from that bag (viz., $19/20 = 0.95$).

[11] I am tempted to think that the hypothesis that the bag contains only blue balls as simpler than the hypothesis that it contains 19 blue balls and 1 non-blue ball because the latter posits two qualitatively different kinds of ball and we may not know of any non-blue balls existing in the world. However, since we acknowledge the logical possibility of a non-blue ball we can never discount the second hypothesis altogether in the extrapolation. If so, then at least some explanatory features may be relevant and constitute a logical factor in cases of extrapolation, though not in generalizations or confirmations of the first kind; for reasons already given generalizations depend only on the evidence, and confirmations of the first kind depend only on prior probabilities, empirical adequacy being irrelevant. I am less sure about confirmations of the second kind. This might save inference to the best explanation of some kind for some cases. Elsewise, inference to the best explanation just seems to be invalid.

[12] Carnap's "*if* we bet" tends to obviate the willingness to bet by presupposing from the start that we are considering a situation where we do bet.

[13] It is the kind of thing that Toulmin (1958, pp. 129–130) says when he says that the information that no relevant information has been left out of the premises or that the instance in question was chosen at random (or, perhaps, that the sample was representative) is "not so much the statement *of* a datum but a statement *about the nature of* our data" and "a passing comment about the applicability to this particular man of a warrant based only on statistical generalities" (p. 130 italics original). Admittedly, this does not mention priors, and in their case this maneuver is more questionable since they do figure in the calculations, e.g., as Hitchcock's r*.

REFERENCES

Bermejo-Luque, L. (2009). Argumentation theory and the concept of epistemic justification. *Studies in Logic, Grammar and Rhetoric*, *16*(29), 285–303

Carnap, Rudolf. 1952. *The continuum of inductive methods*. Chicago: University of Chicago Press

Hitchcock, D. (1999). The Thomas-Nolt dispute: some lessons about induction. *Informal Logic*, *19*(2), 201–212

Popper, Karl. (1962). On Carnap's version of Laplace's rule of succession. *Mind*, 69–73

Nolt, J. (1984). Possible worlds and imagination in informal logic. *Informal Logic*, *6*(2), 14–17

Nolt, J. (1985). More on induction and possible worlds: replies to Thomas and Kahane. *Informal Logic*, *7*(1), 55–60

Szymanek, K. (2014). Justification and argumentation. *Studies in Logic, Grammar and Rhetoric*, *16*(29), 231–239

Thomas, S. (1984). Degrees of validity and ratios of conceivable worlds. *Informal Logic*, *6*(3), 31–34

Toulmin, S.E. (1958). *The uses of argument.* Cambridge: Cambridge University Press

Weddle, Perry. (1979). Inductive, deductive. *Informal Logic*, *2*(1), 1–5

Weisberg, Jonathan. (2011). Varieties of Bayesianism. In D. Gabbay, S. Hartmann and J. Woods (eds.) *Handbook of the History of Logic vol.10: Inductive Logic* (Oxford, UK: Elsevier, 477–552)