SCOTT CAMPBELL and JAMES FRANKLIN

RANDOMNESS AND THE JUSTIFICATION OF INDUCTION

1. INTRODUCTION

In 1947 Donald Cary Williams claimed in *The Ground of Induction* to have solved the Humean problem of induction, by means of an adaptation of reasoning first advanced by Bernoulli in 1713. Later on David Stove defended and improved upon Williams' argument in *The Rationality of Induction* (1986). We call this proposed solution of induction the 'Williams-Stove sampling thesis'.

There has been no lack of objections raised to the sampling thesis, and it has not been widely accepted. In our opinion, though, none of these objections has the slightest force, and, moreover, the sampling thesis is undoubtedly true. What we will argue in this paper is that one particular objection that has been raised on numerous occasions is misguided. This concerns the randomness of the sample on which the inductive extrapolation is based.

2. THE SAMPLING ARGUMENT

Williams' justification of induction is an *a priori* one, that appeals only to certain basic statistical principles. Because his justification is *a priori*, Williams avoids the circularity that Hume showed attaches to any attempted justification of induction that relies on supposed contingent facts (such as the 'uniformity of nature'), for these principles can themselves only be justified – if they can be justified at all – by induction.

The fact that Williams' justification is *a priori*, though, does not mean that he claims that inductive inferences guarantee their conclusions, for this is certainly false, as Hume also showed. Rather, Williams' argument is that we have *a priori* grounds for believing that (good) inductive inferences provide some degree of support to their conclusions – and this is a claim which Hume's arguments do not touch. That is, the conclusion of a (good)

Synthese **138:** 79–99, 2004.

inductive inference is likely to be true, to a greater or lesser degree, given the evidence in the premises. It should not be supposed, Williams says, that entailment is the only kind of support that premises can give to a conclusion. Williams claims that there is rather a spectrum of support that premises can provide. Entailment is the extreme case, where the support given to the conclusion is 100% certainty. But there are other cases where the support is less than this, but is nevertheless still strong.

The sort of case that Williams uses to support this claim is what he calls (following Peirce) the statistical (or proportional) syllogism, where we infer non-demonstratively from a knowledge of the general composition of a population to a conclusion about the composition of a subset or 'sample' of the population. Suppose we have the knowledge that there are one hundred balls in a barrel, and that ninety-nine of them are red. In such a case, it is clear that (assuming this is all the evidence we have about the matter) the statement 'the next ball to be picked out of the barrel will be red' has 99% support from the premises. (This support is, it should be noted, an entirely objective matter.) In other words, there is a 99% probability or likelihood of the conclusion being true, given the evidence we have. The conclusion may well turn out to be false, but it would be rational to have 99% confidence in it being true. This much is relatively uncontroversial. More controversial, though, is Williams' claim that the same sort of support is given to the conclusions of inductive inferences. That is, an inductive inference, where we infer from sample to population, can also provide a strong degree of support to its conclusion, just as an inference from population to sample can. The issue that arises now is whether it can be shown that this is true.

The justification for the assignment of a 99% probability in the statistical syllogism example above is clear enough. There are one hundred balls that could be pulled out of the barrel, and 99% of them are red, so the chances of getting a red one are 99/100. The logic underlying this reasoning is entirely *a priori* – no contingent truths are presumed. But can any such *a priori* justification be provided for inductive cases? Williams shows that the answer to this question is yes.

Consider another barrel case where we infer from population to sample. Suppose the population comprises 100 000 balls in a barrel. One third of these balls are blue, one third are yellow, and one third are black. What is the likelihood that a sample of 1000 balls that we draw out of the barrel will have roughly that proportion (given that we have no other relevant information)? The likelihood of this is extremely high. This can be proven from the mathematical fact that of all the possible subsets of 1000 balls that can be created from the population, the vast majority have roughly the same proportion of colours as the population. Let us say that one set 'matches' another (in respect to a property F) when their composition (in respect to F) is roughly the same. Most subsets of the population, then, will match it in regard to their proportion of colours. Unless, then, we are very unlucky (and assuming that we have no reason to think that our sample is untypical), our sample is very likely to have roughly the same proportion of colours as the population does, and so it is rational to suppose that most likely the 1000 balls we pick out will have roughly the same proportion of colours as the population does.

Williams goes on to show that this sort of reasoning not only allows us to infer from population to sample, as we have done here, but inductively from sample to population. We can do this because matching is a symmetrical relation. As most possible subsets of a population of x's match that population, it follows that a population matches most of its possible subsets. Suppose we then think of induction as like reaching into an enormous barrel containing all the possible subsets of the x's of a particular size – Williams calls this imagined set the 'hyper-population' – and bringing one such subset out. As most of the subsets in the hyper-population match the population, the chances are high that the subset we pull out will match the population. The situation is, in essence, no different from the situation where we pull a ball out of a barrel, where most of the balls are red. In that case, the odds are high that we will get a red ball; in induction, the odds are high that we will get a matching or 'representative' subset, because most of the subsets are representative.

Suppose that we are dealing with ravens, and that we have a sample of 1000 ravens (and no other information about ravens). Consider now the number of possible subsets of 1000 ravens that can be created from the total raven population. (This set of possible subsets is the raven 'hyper-population', for 1000-fold subsets). This number will be extremely large, but even so, it is a statistical fact that the vast majority of these possible subsets will have roughly the same colour composition as the total raven population. That is, the vast majority of possible 1000-fold subsets of ravens match the raven population in respect to colour composition. (This applies whatever the colour composition of the raven population is, and no matter how large the population is, as long as it is not infinite, which it is not, and as long as the sample is not small.[1]) It is therefore rational to believe that most likely our sample matches the population. We cannot hold that our sample will *definitely* match the population, because there is a small chance that it will not, but we are entitled to hold that *most likely* it will match. Hence, the inductive inference from sample to population in this case is justified.

The reasoning here, from sample to population, is just as 'direct' as the reasoning employed in the statistical syllogism that went from population to sample. There is no need to 'invert' the population-to-sample reasoning, as Bernoulli and Laplace (as well as later writers like Carnap (1945a)) thought must happen.[2] This is because Williams treats the possible subsets (of the relevant size) of a population as a gigantic 'hyper-population', from which one subset will be chosen, and so the same sort of reasoning can be employed here as in the population to sample case.

To see this more clearly, consider again our first statistical syllogism. You reach into a barrel which contains 100 balls, 99 of which are red. The odds are very high that you will get a red ball. You could be unlucky and get a non-red one, so you cannot be certain of getting a red one, but you can very confident of doing so. And getting a sample of a population is like reaching into this barrel and hoping for a red ball, for getting a sample is like reaching into the hyper-population and taking out one of the possible subsets, when you know that most of the possible subsets or samples match the population. So, just as you are likely to get a red ball when you reach into the barrel, you are likely to get a representative sample when you reach into the hyper- population. So if your sample contains 95% black ravens, then it is highly likely that roughly 95% of all ravens are black. Of course, it is possible that you could be unlucky, and get an untypical, or as Williams' puts it, an 'unrepresentative' sample, just as you may be unlucky and get one of the few non-red balls from the barrel. But just as you would be irrational not to have a high confidence in getting a red ball, so you would be irrational not to have a high confidence that your sample is representative.[3]

3. THE RANDOMNESS OBJECTION

One common objection that has been made to this argument is that you are only justified in inferring from sample to population if you know that the sample was a 'random' or 'unbiased' sample, one drawn by a 'fair' method. Furthermore, the critics have added, we cannot know that to be the case in induction, so Williams' sampling argument does not justify induction after all. (Or, at least, we cannot know that to be the case in most of the significant real-life situations where we use inductive reasoning, so most of our important inductive inferences are not justified.)

Here are some instances of this complaint from Williams' time:

> ... about the presented sample ... one would seem to need some assumption about its method of selection in order to infer that probably it very nearly matched the population.

One would seem to need assurance that any sample is equally as likely to be drawn as any other. Only thus would there be any ground for supposing the proportion of samples drawn having a certain distribution of P was the same as the proportion of such samples in the population (Ambrose 1948, p. 516).

And:

... unless there is some further evidence, as in the idea of random distributions, that those types of sample which have the higher ratio among the possible samples will be more frequently realized, there is no reason to suppose that this means that this type of sample can be expected with any more assurance than any other (Will, 1948, p. 239).

And from Stove's time:

There is no need to go into the details of the matter [ie. of the sampling argument]; the obvious difficulty provides an immediate and telling refutation of the proof. We are given absolutely no grounds for the assertion that our observations are gathered through a fair sampling procedure. And as Hume noted, the only evidence we can offer for this claim is itself inductive evidence (Brown 1987, pp. 117–118).

And:

The fallacy is explicit. Consider the question 'why should I believe that the sample with which nature happens to have furnished me is a representative sample?' The ... answer , Stove says, is in essence: "... It probably is because most large samples are representative ..." [But] A lot depends on how the sample is obtained. If we know that our sampling procedure is truly random so that we are just as likely to end up with one (large) sample as any other sample of the same size, we may reasonably infer that the sample obtained is probably representative. Suppose now that a high proportion of ravens are not black but all of these live in very remote regions, are difficult to spot, and are easily mistaken for birds of another species when they are spotted. In this circumstance we are likely to obtain a sample which is biased in favour of black; our sample is probably unrepresentative, even though most large samples are representative. In Stove's responses to anticipated objections this one is not considered (Giaquinto 1987, p. 614).[4]

4. WHY RANDOMNESS IS UNNECESSARY

Contrary to what these authors believe, though, the sampling argument is not threatened by any failing in regard to randomness. In fact, the demand for randomness, as articulated by our objectors, leads to nothing but absurdity.

Consider again the second statistical syllogism given earlier. The barrel contained 100 000 balls, one third of which were blue, one third yellow, and one third black. We had to pick out 1000 balls, and the odds of getting a representative sample were very high. But suppose we took the balls from the top of the barrel, and before we have looked at their colour, someone who we know to be reliable, and who had inspected the barrel just beforehand, tells us that the top layer contained mostly blue balls. In that case, we have good reason to think that our sample was one of the small proportion of possible subsets that are not representative, and so our confidence that our sample is representative should be reduced drastically, and we should think instead that it is more likely that the balls are blue, or mostly blue.

This, though, does not change the fact that in a situation where you have no reason to think that your sample is not representative, you should think that probably it is. [5] It would be irrational for you to think, for example, that your sample would contain all yellow balls, in the absence of any reason to think that your way of picking the balls out of the barrel will produce yellow balls, rather than blue or black. Unless you have a positive reason to think that you will get an unrepresentative sample, you should suppose that you will probably get a representative one, simply because there are so many more representative ones than unrepresentative ones.

The same principle applies in induction. If you have no reason to believe that your sample is not representative, then probably it is representative, because most possible samples are. However, if you have reason to believe that your sample is unrepresentative, then your confidence in the conclusion must be reduced. For example, suppose you get a 1000-fold sample of a new type of bird, all of which are orange, but you then find out that this sample of ravens comes from a valley where all birds are orange, no matter what colour they are outside of that valley. In that case, you have reason to think that your sample is not representative.

But if you have no such reason to think that your sample is not representative, then it is rational to think that probably it is, *because most samples are*. It would be irrational of you to think that your sample is unrepresentative (or to think that we must sit on the fence) when there is no evidence to support this. To think otherwise - to think, as the 'randomness' critics do, that we cannot conclude that a sample is likely to match the population unless we know that its method of selection is random or unbiased – is equivalent to holding that we cannot suppose that we are likely to draw a red ball out of a barrel of a hundred balls, 99 of which are red, unless we know that the method of drawing out the ball is random or unbiased. But of course we do not need to use any randomizing procedure in such a case. If we have no reason to think that the non-red ball is more likely to be in any one part of the barrel than any other part, then any method for extracting a ball is as good as any other, and no matter what method we use (presuming that we have no good reason to think that the method we use will somehow produce one of the few non-red balls), we are entitled to think that most likely we will get a red ball.

The same applies to the inductive case. There exists a vast number of possible subsets of the population, most of which match it. Unless you have a reason to think that your sample is unrepresentative or biased, the odds are that it is representative.[6] You do not need to have gained the sample by some 'random' procedure, any more than you do with the ball in the barrel case. Nor do you need a positive reason to think that your sample probably is representative, any more than you need a positive reason to think that the ball you choose from the barrel probably is red, because the sheer fact that most samples in the hyper-population are representative is positive reason in itself for this belief.[7]

So, contrary to what these critics suggest, you do not need to know that your sample is not biased – all you need to know is that the sample is not known to be biased (or have no reasons to think that it is likely that it is biased). In fact, as Williams has pointed out, the demand that we know that the sample is not biased is 'extortionate and self-contradictory', for it is nothing but the demand that we know in advance that the sample matches the population, and to know that, we would have to know the composition of the population, in which case we would not need induction (Williams, 1947, p. 67 ff).[8]

5. 'ONE SAMPLE AS LIKELY AS ANY OTHER'

It might be replied by our critics that the randomness demand is not that the sample be known to be unbiased in a sense that implies that we must know already know the composition of the population, but only that we need to know that any one possible sample is as likely to be drawn as any other.

However, it is not at all clear what this means. Perhaps it means that our method of drawing a sample of x's must not be such that the very method itself is incapable of drawing in x's of a certain type F (or unlikely to do so), where F is the sort of property that is relevant to the inductive inference in question. For example, if we are interested in the colour of ravens, our method of drawing a sample of ravens must not be such that we know that it is only capable of getting ravens of one particular colour. Nor must our method be one that we know favours one type of colour over another.

If something like this is meant, then the demand is a most reasonable one, but then Williams and Stove would agree with it, and it presents no problem for their argument. Giaquinto might demur here, though. He might say this. Suppose your method for observing the colour of ravens is to look in the valleys and plains. But it could be that most ravens are white and live in the mountains. In that case, you will fail to observe all the white ravens in the mountains. But such an objection would be beside the point, for there is nothing about observing ravens in the valleys and plains in itself that prevents you from observing any white ravens that happen to be there. That is, this method in itself does not prevent you from seeing white ravens. You will, as it happens, see no white ravens if you employ this method, but that is only due to the contingent circumstance that there are (we are supposing) no white ravens in those areas to be seen. But that does not show that induction is not justified. That just means that in this particular case we will unluckily get an unrepresentative sample. That is not a problem for induction, though, for induction does not *guarantee* that the sample we actually get will be representative. It is unreasonable to demand that the method we use must draw in a representative sample, for it is always possible that our sample is untypical. We can only demand that the method be such that we know of no reason to think that it will favour the few unrepresentative samples over the far more numerous representative ones.[9]

If we think back to the barrel case, we can see that it is reasonable to demand that the method we employ for pulling out a ball from the barrel not be known to be one that favours one type of colour over another. For example, suppose we are using a dog to pick out a ball, and we know that he has been trained to avoid picking red balls. In that case, we could not have confidence that he will pick out a red ball, despite the fact that the red balls comprise 99% of all the balls in the barrel. But if we do not know that the dog has been so trained, and we have no reason to even be suspicious that someone has, say, secretly been training the dog to avoid red balls in order to make us go wrong, then we are entitled to think that we will probably get a red ball by using the dog. We do not need to know that the dog does not avoid red balls, in order to think that a red ball will probably be picked out by the dog. The mere possibility that the dog may avoid red balls is not enough to make us doubt that he will most likely pick out a red ball; after all, the dog could equally well have been trained to avoid yellow balls, or blue balls, or black, orange and pink balls, and so on. All that is required is that we do not know of any reason for thinking that the dog avoids red balls.

Any stronger interpretation of the claim that ‘we need to know that any one sample is as likely to be drawn as any other’ will be, it seems to us, unreasonable. For example, it may be felt that what is needed is that we must know that our method of drawing a sample is such that if we applied it again and again, we would eventually draw out every possible subset an equal (or roughly equal) number of times. Or it may be felt that what is needed is that we know that our method of drawing a sample is such that in the long run, the set of observed x's will match the total population of x's in the relevant respect.[10]

These demands are too strong, though, and unnecessary. Consider the first demand. It may be that our method would only ever pick out a small class of possible subsets. But as long as we have no reason to think that those subsets are not representative, then we are entitled to think that probably they are, because there is a far greater range of representative subsets than unrepresentative subsets, and so the odds that our subset belongs to the bigger range rather than to the smaller range are far greater. Consider the barrel case again. Suppose we decide that our method will be to pick out ball number 85 (suppose that we can feel the ball numbers with our hand, without seeing their colour). This method, if it were to be applied over and over (with replacement), would not in the long run pick out every ball an equal (or even a roughly equal) number of times. But this fact will not dent our confidence that the ball we pick out will most likely be red, as long as we have no reason to think that ball number 85 is not red.

And we can see that the second demand (that in the long run the set of x's that we observe with our method must match the total population of x's – a similar demand is that the conclusions of our inductive inferences are mostly correct in the long run) is also unreasonable.[11] We cannot possibly have any such guarantee, because it is always possible that no matter how many x's we observe in the long run (presuming that we have not seen all or most of the x's), the rest of the population of the x's is completely different, just as it is possible that we could pick out balls, with replacement, from the barrel for forever and a day, and get the one non-red ball every time. But we do not need a 100% long-run guarantee for induction to be justified, any more than the statistical sampling thesis requires 100% certainty for its justification. Williams shows us that induction gives us a high *probability* of success in the long run (and in the short run, for that matter), and this is enough to make it rational to reason inductively, just as a high probability of getting the red ball from the barrel is enough to make it reasonable to have a high confidence in the conclusion that you will get a red ball. We do not need 100% certainty in either case.[12]

6. SUSPICIOUS CIRCUMSTANCES

Another author who has recently objected to the sampling thesis on the grounds of randomness is Bipin Indurkhya (1990). He argues as follows.

If our inductive situation involves us picking out a large number of balls from an urn, then, he concedes, we would be entitled to inductively infer to the composition of the whole population of balls in the urn, because our selection in such a case is, in some unspecified sense, 'random' (103). But, he says, consider the following scenario. You are standing outside a sealed room which is known to contain a million balls, and each ball can only be black or white. Someone inside the room passes out a ball to you through a slot, you examine it and return it, and are then given another ball, and so on. Suppose you go through this procedure 3000 times, and 95% of your sample is black. Are you entitled to infer that probably roughly 95% of all the balls in the room are black? Indurkhya says

> I claim that in the absence of any other information, any rational person will be very much reluctant to make such sweeping generalization. For instance, there seems to be no reason to suppose that the person is giving a new ball to be examined each time, instead of just using one black and one white ball (102).

The problem with this example, though, is that any reluctance to make an inductive extrapolation here seems to be due to the suspicious nature of the circumstances - the scenario seems to be one that has been set up with the express purpose of deceiving people outside the room. Or, at least, the strange circumstances would make it reasonable to think that the person inside has a reason for distributing colours that might not match up to what is inside. As Indurkhya himself says, 'there seems to be no reason to suppose that the person is giving a new ball to be examined each time, instead of just using one black and one white ball'. By this he presumably does not just mean that this is a bare possibility, however unlikely, for if it were merely a bare possibility, it would not be worth taking seriously. He means, rather, that it is a distinct likelihood, perhaps just as likely as not. But this means that we *do* have some reason to think that the sample we have been given is unrepresentative. So, just as Indurkhya says, we should be very reluctant to generalize from our sample here. But this fact does not conflict with Williams' sampling thesis - it is, on the contrary, just what his thesis tells us we should do in such a case, for it tells us that we should not accept the conclusion of a sample-to-population extrapolation if we have some reason to think that the sample is unrepresentative.

Indurkhya says that one objection to his argument might be 'that it is the presence of a person, a free agent, that makes all the difference' (102). To overcome this, he says, he can simply replace the person with a machine. And 'in the absence of any information about how the machine operates one would still be reluctant to conclude anything about the balls in the room' (102). We are not sure that we would agree that one would or should be reluctant to conclude *anything* about the balls in the room in such a case. But the main point to be made to Indurkhya here is that the circumstances are still suspicious, and what's more, they carry a strong residue of suspicion from the previous case which involved the person. So if one is reluctant to come to any strong conclusions in such a case, it will be because one thinks there is a reasonable chance that one is being deliberately set up with an unrepresentative sample, or that the machine has orders to distribute colours according to certain criteria which may not match what is inside, and as Williams tells us, we are not entitled to generalize from a sample if we have reason to think it is unrepresentative.

It should also be pointed out that it would be reasonable to think the following. The person who filled up the room may have done so not by buying bags that contained both black balls and white balls, but by buying bags of white balls, and then bags of black balls. They may have firstly tipped all the bags of white balls into the room, and after that tipped all the black balls into the room. In that case, there may be mostly black balls in the front part of the room, and these may be the balls that the person or machine inside the room is passing out, as they are the ones closest to them. Such a scenario is not just a bare possibility (bare possibilities are, after all, incredibly numerous, and furthermore they cancel each other out), but something that we know from experience has a moderately reasonable likelihood of being true. Given this, we have a further reason to reduce our confidence in the conclusion that roughly 95% of the balls in the room are black, and this reasoning is entirely in accord with Williams' thesis. So we do not think that Indurkhya's examples present any problem for the sampling thesis.[13]

7. SPATIAL PROXIMITY

Indurkhya also relies on the claim that we need randomness because

> it is very easy to imagine that due to environmental differences, ravens in different geographical regions have different colours... Having drawn one's sample all from Australia it would be unreasonable to conclude that 95 percent of the ravens are black all over Earth (103).

We also saw that Giaquinto put forward such an argument – he imagined that most ravens were white, and lived on high mountains. Thus, our belief that most ravens are black, acquired inductively from our sample of observed ravens, would be false.

But the mere fact that there *could* be differences between the x's we observe and the rest of the x's is not a reason to doubt the rationality of induction. That there *could* be such a difference is the very fact which prevents us from being certain when using induction. This possibility is what is being allowed for when we say that our inductive conclusion is highly probable, but not certain. But unless we have a positive reason to think that there actually *is* such a difference, or that such a difference has some non-negligible likelihood of being true, we are justified in thinking that most probably our sample is representative, because most subsets of the population are.

Indurkhya, though, thinks that what is required is (at the very least) that our sampling method is such that it prevents the elements of our sample from being spatially proximate. He says, for instance, that an inductive inference about ravens is only justified if our sample is drawn from all over the face of the Earth. Moreover, he claims that this would only justify an inference to present unobserved ravens, but not to future or past ravens. Suppose that we restrict the time scale we wish to infer to, to between 8 000 B. C. and 12 000 A. D. In that case, we would need to have a sample whose members are drawn evenly across this time span. As this is impossible, at least in practice, then induction to the future and past is impossible, at least in practice.

But what could justify such demands? There can be no *a priori* justification for them, for there is no *a priori* reason to think that a sample that contains spatially and/or temporally proximate elements is likely to be unrepresentative, or that it is any more likely to be unrepresentative than a sample that contains elements that are more widely separated in space and/or time. If there is a justification for such a demand, then, it can only be from an empirically acquired fact (or facts). But this will not harm the sampling thesis, because we will therefore have a reason for thinking that these samples are unrepresentative, and the sampling thesis tells us that we should not generalize from a sample if we have reason to believe that it is unrepresentative. So even if it is true that we should not generalize from samples that are spatially and/or temporally proximate (to some specified degree), this presents no threat to the sampling thesis.

The critic might respond that while this response shows that Indurkhya's claims do not defeat the sampling thesis itself, his claims do at least show that there are empirical reasons why we are not entitled to use it in making inductive inferences from samples whose elements are spatially and/or temporally proximate, and this will greatly restrict the amount of inductive inferences we can make, because many of our inductive inferences rely on such samples.

But is it true that it has been empirically discovered that samples with spatially (and/or temporally) proximate elements are often unrepresentative?[14] It is true that we have discovered that there exist some unrepresentative 'clumps', and sometimes our inductive conclusions have turned out to be false because of this. Most such cases come from the animal kingdom, and from the human population, e.g., in some locations, swans are white, and in some locations they are black. And sometimes the colours of animals change over time. (Indurkhya mentions the classic case of the Midlands moths whose colour darkened with the increase in soot during the industrial age (104).[15]) And in some suburbs, more people may vote one way, while in other suburbs more people vote the other way.

Because of cases such as these, when we are dealing with a sample whose elements are spatially (and/or temporally) proximate, we should perhaps reduce our confidence to some minor degree when making an inductive generalization, although mainly so only in cases that involve animal or human populations. But most of the inductive generalizations that have been made on the basis of samples whose elements are spatially (and/or temporally) proximate have not turned out to be false. We have not discovered that most ravens are not black. Nor have we discovered that most grass is not green, or that most snow is not white, or that most people do not have red blood. Nor do most things change colour over time (and many of those that do, do so in short, easily-observable time spans anyway, for example, apples and tomatoes). The law of gravity hasn't ever been observed to differ anywhere or at any time, nor has fire ever not burned.

So it is not true that a large (or even a moderate) proportion of our inductive generalizations which have been made on the basis of samples whose elements are spatially (and/or temporally) proximate have turned out to be false. This has happened only occasionally, and only in regard to certain fields. Besides, because inductive generalizations are not certain, but probabilistic, we should expect to find that the occasional sample is not representative after all, as we did with, for instance, swans. So there is in fact very little reason (whether *a priori* or empirical) to lose confidence in generalizing from spatially (and/or temporally) proximate samples.

Anyway, the only way we could decide that generalizations drawn from samples whose elements are spatially (and/or temporally) proximate are often false is by getting further samples whose elements are not spatially (and/or temporally) proximate to the specified degree. But if we possess such samples, we can thus use *them* to extrapolate inductively from, because we will have no empirical (or *a priori*) reason to believe that these larger samples are unrepresentative.

Moreover, once we discover that our inductive inferences in some particular field have been going wrong, we can use this knowledge to develop techniques for stratifying our samples, in order to eliminate samples that we know are likely to be unrepresentative.[16] Consider political polling.

We have discovered that if we just take a survey of attitudes in one suburb, we may get a sample that is unrepresentative of the larger population. To overcome this, various techniques have been developed which allow us to avoid getting such a sample. It is still always possible that there are further factors at play which mean that such a method will still mostly produce unrepresentative samples in the case at hand, but in the absence of any knowledge that this is the case (such as a failure in the past of such stratified sampling methods), we are entitled to think that this is unlikely.

It might be said that the fact that we are obliged to apply such stratified sampling techniques to the field of political polling means that we are only justified in inductively inferring from a sample in *any* field if we have employed such stratified sampling techniques in acquiring it. But there are no grounds for making such a claim. It is only when we have a reason for thinking that an unstratified sample would be unrepresentative that we are required to use stratifying techniques. If we are, for example, surveying opinions on music, then we would do well to use stratified sampling techniques to make sure that our sample took in many different types of people, and this is because we have reason to think that if we do not, we will probably get an unrepresentative sample. Even if we knew nothing about musical opinion, the known fact that peoples' opinions on cultural matters change from area to area, or from social type to social type, gives us a reason to think that they may well also vary in such ways when it comes to music, and this in itself would justify us in using stratifying techniques.

But if we are interested, say, in testing the specific gravity of a new substance, there is no point in making sure that we test it in different parts of the world, for we have no reason to think that unless we do, we will get an unrepresentative sample (and all our experience so far has told us that the laws of physics do not change from area to area). The fact that we have gone wrong at times with unstratified sampling techniques in regard to human opinion provides no reason for thinking that we will also go wrong in using unstratified sampling techniques in physics. The two fields are too unlike to justify us in inferring from one to the other in this way. The more similar two fields are, then the more we are entitled to think that what is true of one is likely to be true of another, but the less alike two fields are, the less reason we have for thinking this.

So despite Indurkhya's claims, it is not true that we have any reason to doubt the majority of the inductive inferences that we make.

8. EXPERIMENTAL RANDOMISATION

In medical and agricultural trials, experiments that have not been deliberately 'randomized' in certain ways are usually considered to be worthless. It might then be argued that this shows that our denial of the claim that one must only extrapolate from a random sample is mistaken.[17]

Consider the classic agricultural trial, where two varieties of a crop are being tested to see which grows faster. It is usually not considered adequate to grow all the seeds of one variety in one half of the field, and the seeds of the other variety in the other half (or to grow one variety in one field, and the other in another field). The reason for this is that there may be differences between the halves of the field (or between the two fields), for example, one half may have more fertile soil than the other, or more pests, etc., and so any difference in the growth rates of the two varieties may well be due to differences such as these. Fisher's solution to this problem was to divide the field up into pairs of plots (each pair being called a block), and using a randomizing device to decide which plot in each block receives which type of seed. In the light of this, shouldn't it also be held that we can only inductively infer from sample to population when the elements of the sample are gathered using a randomizing process?

The answer to this is no. First of all, there are a large number of factors that may *possibly* affect the growth rates of the two varieties of seeds. Suppose, for example, that variety A comes in a red packet, and variety B in a blue packet. It is possible that seeds that come in blue packets grow faster than seeds that come in red packets. And there are a multitude of other such factors which could possibly have a differential effect on the growth rates. Shouldn't we then also randomize for these factors as well, or at least try to control them in some way?

We could, but it would obviously be a complete waste of time to do so. In the case of the colour of the packets, we know that it is overwhelmingly likely that packet colour has no effect on the growth capacity of the seeds inside the packet, so there is no point in practice in randomizing for it, or controlling for it in any way. There is only a point in randomizing or controlling for factors which we have reason to believe may be operational in the actual situation, that is, which have some non-negligible probability of affecting the result. And that is why we randomize for the plot allocation, because we have some reason to believe that the growth rates may be differentially affected by soil fertility, which may well vary spatially, if we just use a left-right allocation. That is, the chances of this allocation affecting the growth rates are not negligible.

This is all in accord with the sampling thesis, though. If we performed the experiment without randomizing the plot allocations, then we could not have much confidence in the result, or at least, not as much confidence as we could have had the randomization been carried out. And this is because we have some reason, based on our knowledge of the world, to believe that our sample was unrepresentative, this reason being that there is a fair chance that one side of the field has more fertile soil than the other. So it is because we have a reason to think that our sample here may well have been unrepresentative that we do not place much confidence in it. We do not take account of bare possibilities, though – we cannot decide that a sample is likely to be unrepresentative simply because of a bare possibility. For example, we are not entitled to deny the result of a randomized trial by saying that the different colours of the packets may have made the difference, because the odds of this being true are so low. The reason must be one that has a non-negligible degree of likelihood.

So experimental randomness reveals no problem for the sampling thesis. In the absence of any reason to believe that a sample is unrepresentative, we are entitled to take it that it probably is representative. But where we have a reason to suspect that our sample is unrepresentative, as we do when the seeds are allocated in the left-right pattern, we must reject the result.[18] And this is all in accordance with the sampling thesis.

We should also note that the fact which we take into consideration here, namely that one half of a field may well be more fertile than the other half, is one that we have acquired by empirical means. We have discovered that it is a contingent fact about the world that the fertility of the soil can sometimes vary from one side of a field to another. This fact is not something we know *a priori* – it could have turned out that the world was such that soil fertility usually formed completely different types of patterns than in the actual world. So *a priori*, this is no more than a bare possibility, just like the claim that the colours of the packets differentially affect growth rates is. And we do not bother to randomize for factors which may *possibly* affect growth rates. We only bother to randomize for factors which we have reason to think do often, or do sometimes, make a difference in the actual world.[19] Besides, *a priori*, there are many mutually exclusive factors which may possibly affect growth rates, so it is impossible even in principle to randomize for all possible factors which may, *a priori*, affect growth rates.

9. RANDOM AND NON-RANDOM INDUCTIVE INFERENCES

Our last argument is a general one. Suppose the randomness supporter claims to have a sample that has been acquired by a randomized method.

How, then, does he get to the conclusion that this sample probably is representative? He can only do so, it seems, by means of a premise to the effect that most samples chosen by such a method match the population. Hence, his argument will be as follows.

Argument I
1. This is a large sample that has been chosen by a randomized method.
2. Most large samples chosen by a randomized method match the population.
Probably, then,
3. This sample matches the population (and is thus representative).

But if this method of reasoning is acceptable, then the randomness supporter can have no objection to the following argument, which concerns a sample that was not acquired by a random method.

Argument II
1a. This is a large sample.
2a. Most large samples match the population.
Probably, then,
3. This sample matches the population (and is thus representative).

Neither premise of this argument can be doubted. The first is what we have been given. The second is a mathematical truth. Any doubt, therefore, must centre on the (non-demonstrative) inference from 1a and 2a to 3. But this inference seems to be the same, in essence, as the inference in Argument I from 1 and 2 to 3. If the inference in I is justified, it is hard to see why the inference in II is not. So if inductive inference with randomness is justified, so must be inductive inference without randomness.

This is clearer still if we put the arguments more formally. Argument I becomes

Argument III
1b. This is a sample that has properties X and Y.
2b. Most samples that have properties X and Y match the population.
Probably, then,
3. This sample matches the population (and is thus representative).

(X and Y here replace 'large' and 'chosen by a random method' respectively.) Argument II becomes:

> *Argument IV*
> 1c. This is a sample that has property X.
> 2c. Most samples with property X match the population.
> Probably, then,
> 3. This sample matches the population (and is thus representative).

It cannot be that the inference in Argument III is acceptable, and that of Argument IV unacceptable. Hence, if 1 and 2 'probabilify' 3, then so do 1a and 2a. The randomness objection will therefore fail, unless the randomness objector can demonstrate that the inference that the randomness objector makes from 1 to 3 depends on an entirely different premise, or an additional premise (or premises) than 2.[20] No such additional premise seems to be in the offing, though, and so we conclude that it is likely that the randomness objection to the sampling thesis fails.[21]

10. CONCLUSION

We conclude that there is no threat to the Williams-Stove sampling thesis from any objection that concerns randomness. Induction is justified by the fact that most possible subsets of a population match the population, so unless you have a reason to believe that your sample comes from the very small proportion of unrepresentative subsets (or has been produced by a method which you have reason to believe produces samples from this small subset), then, relative to what you know, the odds are that it comes from one of the very large proportion of representative subsets, and thus it is most likely that the population matches the sample. You do not need to know that your sample is not biased; all that is required is that you have no reason to think that it is biased.

This means that the inductive sceptic can easily create imaginary scenarios in which your sample is unrepresentative. But this of no consequence; all it does it to highlight the fact that in using induction, you could always be wrong. (Such unlikely circumstances are in fact *what is allowed for* by saying that induction does not provide certainty, but probability.) So unless you have a reason (of non-negligible strength; mere *a priori* possibilities do not count) to think that your sample is unrepresentative, the odds are that it is representative.

A reason for thinking that your sample is unrepresentative need not be anything very definite. If, for example, you have some reason to suspect that you are being tricked, then that counts as a reason to reduce your confidence in the generalization from your sample. But if you have no such reason, then it is rational to have a high confidence in it.

We can conclude, then, that the sampling thesis demonstrates that induction is justified. This is not to say that it clears up all the troublesome details involved in inductive logic, but it is to say that it defeats the skeptical Humean view that induction can never be rationally justified.[22]

NOTES

[1] The mathematical details of these matters can be found in Williams 1947, as can the details of such terms as 'matches', 'most', 'probably', etc. An exposition of Bernoulli's proof (which Williams relies upon) can be found in Kneale (1949a), §29 (pp. 136–142). (A proof using Stirling's theorem appears in Keynes (1921), 338–340).

[2] Nor is there any hidden inversion or 'indirectness' here, as some commentators on Williams, such as Kneale (1949b) and Black (1947), have thought (see Williams 1953 for details, and also McGrew 2001).

[3] For some recent defences of the sampling thesis against objections, see Campbell (2001), Franklin (2001), and McGrew (2001).

[4] Similar views can also be found in Ernest Nagel (1947), A. J. Ayer (1972), pp. 40-42, and Anthony O'Hear (1989), pp. 147–149.

[5] It needs to be remembered that probability is always relative to the premises. Hence, if the information from your informant that the top layer contained mostly blue balls was added to your premises, then the probability of the conclusion relative to these premises is low. But that does not change the fact that relative to the original premises, which do not contain this information, the probability of the conclusion is high.

[6] In case our terminology confuses, let us stress that our use of terms such as 'the odds are', and 'the chances are', are to be taken in an epistemic sense, rather than in any frequentist or stochastic sense.

[7] It might be thought that we should at least hold that the odds of a sample being representative are equal to the odds of it not being so. But this would be no more rational than holding that the odds of pulling out the non-red ball from the barrel are 50–50.

[8] Henry Kyburg is one of the few authors who has argued over the years for similar conclusions as ours. See, for eg., Kyburg (1972), ch. 9; (1974). For an excellent recent defence of Williams against randomness-type objections, see McGrew (2001). We are in full agreement with McGrew's arguments.

[9] It also has to be suspected that Giaquinto is trading on our background knowledge that geographical distribution does tend to be slightly relevant to bird colour, while pretending not to. We consider the issue of spatial relevance in §7.

[10] What counts as the 'long run' is usually left unspecified by such objectors.

[11] Ambrose, Will and Nagel all seem to be demanding something along these lines (as does Thomas Mayberry 1968, who also puts forward the randomness objection to Williams).

[12] We suspect that the confusion of Ambrose, Will and Mayberry derives in part from the confusion about probability that is engendered by the frequency theory of probability. Williams brilliantly demolished the frequency theory in his contributions to a symposium

on this issue (1945a, b and 1946. See also Carnap's well-known contribution to this symposium (1945b)).

[13] It should be noted that in almost all cases of real inference there is relevant background information – it does not just occur in artificial cases such as Indurkhya's. Some discussion of this issue can be found in Franklin (2001).

[14] We will from now on assume that what is meant by 'spatially and temporal proximate' has been specified (noting, though, that the onus of specification is on the critic, not us).

[15] However, recent evidence has cast doubt on whether this really happened.

[16] We should also note that we are entitled to bring other, more general knowledge, to bear on any sample. That is, when making an inductive inference from a sample, we are entitled to modify our inference in the light of other scientific facts we have learned, e.g., facts from physics, biology, genetics, evolutionary adaptationism, etc.

[17] Sometimes randomness is used to prevent unconscious bias on the part of an experimenter, or, as in parapsychology, to prevent a subject from cheating or unconsciously detecting a pattern. Such uses of randomization are not related to our issue.

[18] Or infer with reduced confidence (although such a practice is frowned upon by classical statisticians – see, for example, Valliant (1987).

[19] This is the reason why physicists and chemists rarely bother with randomization – it is because they rarely have a reason to think that there exist any factors which may affect the results and which cannot be controlled for by non-random means.

[20] And in this case, he will have to hold that the arguments in I–IV are all invalid.

[21] A different but similar-sounding argument can be found in Kyburg (1956), 398–399.

[22] The authors would like to thank Tim McGrew, Nick Shackel, Lydia McGrew, Stephen Hetherington, and an anonymous referee. Dr. Campbell also wishes to thank the Leverhulme Trust for their financial support.

REFERENCES

Ambrose, Alice: 1948, 'Review of Williams 1947', *Mind* **57**, 514–516.

Ayer, A. J.: 1972, *Probability and Evidence*, Columbia University Press, New York.

Bernoulli, Jacques: 1713, *Ars Conjectandi*, Basileae.

Black, Max: 1947, 'Review of Williams 1947', *Journal of Symbolic Logic* **12**, 141–144.

Brown, M. B.: 1987, 'Review of Stove 1986', *History and Philosophy of Logic* **8**, 116–120.

Campbell, Scott: 2001, 'Fixing a Hole in the Ground of Induction', *Australasian Journal of Philosophy* **79**(4), 553–563.

Carnap, Rudolph: 1945a, 'On Inductive Logic', *Philosophy of Science* **12**(2), 72–97.

Carnap, Rudolph: 1945b, 'Two Concepts of Probability', *Philosophy and Phenomenological Research* **5**(4), 513–532.

Fisher, Ronald A.: 1935, *The Design of Experiments*, 7th edn, Oliver and Boyd, Edinburgh/London.

Franklin, James: 2001, 'Resurrecting Logical Probability', *Erkenntnis* **55**, 277–305.

Giaquinto, Marcus: 1987, 'Review of Stove 1986', *Philosophy of Science* **54**, 612–615.

Indurkhya, Bipin: 1990, 'Some Remarks on the Rationality of Induction', *Synthese* **85**(1), 95–114.

Keynes, John Maynard: 1921, *A Treatise on Probability*, Macmillan, London.

Kneale, William: 1949a, *Probability and Induction*, Oxford University Press, Oxford.

Kneale, William: 1949b, 'Review of Williams 1947', *Philosophy* **24**, 86–88.

Kyburg, Henry: 1956, 'The Justification of Induction', *Journal of Philosophy* **53**(12), 394–400.
Kyburg, Henry: 1974, *Logical Foundations of Statistical Inference*, D. Reidel, Dordrecht.
Kyburg, Henry: 1979, 'Tyche and Athena', *Synthese* **40**, 415–438.
Mayberry, Thomas C.: 1968, 'Donald Williams on Induction', *Journal of Thought* **3**, 204–211.
McGrew, Timothy: 2001, 'Direct Inference and the Problem of Induction', *The Monist* **84**(2), 153–178.
Nagel, Ernest: 1947, 'Review of *The Ground of Induction* (by D. C. Williams)', *Journal of Philosophy* **XLIV**(25), 685–693.
O'Hear, Anthony: 1989, *Introduction to the Philosophy of Science*, Clarendon Press, Oxford.
Stove, David: 1986, *The Rationality of Induction*, Clarendon Press, Oxford.
Urbach, Peter: 1985, 'Randomization and the Design of Experiments', *Philosophy of Science* **52**(2), 256–273.
Valliant, R.: 1987, 'Conditional Properties of Some Estimators in Stratified Sampling', *Journal of the American Statistical Association* **82**, 509–519.
Will, Frederick L.: 1948, 'Donald Williams' Theory of Induction', *Philosophical Review* **57**, 231–247.
Williams, Donald: 1945a, 'On the Derivation of Probabilities from Frequencies', *Philosophy and Phenomenological Research* **5**(4), 449–484.
Williams, Donald: 1945b, 'The Challenging Situation in the Philosophy of Probability', *Philosophy and Phenomenological Research* **6**(1), 67–86.
Williams, Donald: 1946, 'The Problem of Probability', *Philosophy and Phenomenological Research* **6**(4), 619–622.
Williams, Donald: 1947, *The Ground of Induction*, Russell and Russell (1963 reissue), New York.
Williams, Donald: 1953, 'On the Direct Probability of Inductions', *Mind* **62**, 465–483.

Scott Campbell
Department of Philosophy
University of Nottingham
University Park, Nottingham
NG7 2RD
U.K.

and

Institute for the Study of Genetics, Biorisks and
Society (IGBiS),
University of Nottingham

James Franklin
School of Mathematics,
University of New South Wales,
Sydney, NSW, 2052,
Australia