An Argument By Stove Against Inductive Scepticism

Introduction

In *The Rationality of Induction*, Stove presents an argument against scepticism about inductive inference—where, for Stove, inductive inference is inference from the observed to the unobserved. The argument is appealing because it attempts to squeeze through a lacuna in Hume's sceptical argument.[1] Recall Hume's argument against the reasonableness of expecting nature to be uniform. First, there is no a priori argument to the uniformity of nature, since it is conceivable that nature will change her course; second, there is no empirical argument to the probability of the uniformity of nature, because any prospective argument can be shown to be circular; third, since a priori (to a certain conclusion) and empirical (to a probable conclusion) are the only sorts of argument that there are, there is no argument to either the uniformity of nature or the probability of the uniformity of nature.

The third part of Hume's argument rules out an a priori argument to the probability of the uniformity of nature. Stove thinks that this sort of argument is possible, and attempts to give one. More precisely, he attempts to give an a priori argument that shows that there is a large class of inductive inferences whose members have highly probable conclusions.

In Section 1, I present Stove's argument. In Section 2, I argue that the argument must undergo a modification, and propose one. In Section 3, I show that the proposed modification requires an additional modification, and in Section 4 I propose one. In Section 5, I take a pause from the main theme of the essay and use the results obtained thus far to show how to meet one of the common objections to Stove's argument. In Section 6, I explain why I suspect that the argument (after two modifications) will still be unsuccessful.

1. Presenting Stove's argument

Let U be a finite collection of *n* particulars such that each member of U either has property F-ness or does not. If *s* is a natural number less than *n*, define an *s*-fold sample of U as *s* observations of distinct members of U each either having F-ness or not having F-ness.[2] Let p_U denote the proportion of members of U that are Fs[3] and, if S is an *s*-fold sample of U, let p_S denote the proportion of members of S that are Fs. Call S representative iff $|p_S - p_U|<0.01$.

Stove's argument is built on the following statistical fact:

> (SF) As *s* gets larger the proportion of all possible *s*-fold samples of U that are representative gets closer to 1 (regardless of the size of U or of the value of p_U).

Some readers will find (SF) more intuitive than others. It is perhaps easiest to see when U is small and p_U high. For example, let U be a collection of 100 balls, 95 white and 5 black. In the context of this example only, redefine "S is representative" as $|p_S - p_U|\leq 0.05$. Then a 10-fold sample of U is unrepresentative iff it contains at least 2

[1] See (Millican 2002, pp.136-138)

[2] Since the observations can be paired with the particulars that are observed, I henceforth write as if the sample is composed of particulars.

[3] In this essay, "x has property F-ness" is used interchangeably with "x is an F".

black balls; a 30-fold sample is unrepresentative iff it contains at least 4 black balls; a 50-fold sample is unrepresentative iff it contains at least 6 black balls. Since there are only 5 black balls in U, it is clear that a lower proportion of 30-fold samples are representative than 50-fold samples (all 50-fold samples are representative). Similarly, since an unrepresentative 30-fold sample must take at least 4 of 5 black balls, whereas an unrepresentative 10-fold sample need only take at least 2 of 5 black balls, it should be fairly plausible that the proportion of 10-fold samples that are representative is lower than the proportion of 30-fold samples that are representative.

I will not give a proof of (SF) because I do not think that it will be illuminating. I leave the reader to fully convince themselves of (SF) by trying examples.[4] I will now show how Stove uses (SF) in his argument. I have broken the argument into three steps:[5]

Step One

Let p be any real number between 0 and 100 inclusive; F-ness and G-ness any properties; x any particular.

Step One of the argument is to assert that the proposition "x is a G" has probability at least p/100 in relation to the conjunctive proposition "At least p% of the Fs are G and x is an F".[6]

Step Two

Let U be defined as in the beginning of this section. Also specify that U has at least 250,000 particulars. (SF) says that as *s* gets larger the proportion of possible *s*-fold samples of U that are representative gets closer to 1. In particular, it can be shown that:

At least 99% of all 250,000-fold samples of U are representative.

Step Two is to prove this result.[7]

Step Three

Let S be a 250,000-fold sample of U. An instance of the generalisation asserted in Step One tells us that the proposition "S is representative" has probability at least 0.99 in relation to the conjunctive proposition "At least 99% of 250,000-fold samples are representative and S is a 250,000-fold sample". By Step Two, the conjunctive proposition is true.

2. A modification of Stove's argument

Stove ends the argument where I have ended it in Step Three. The point of the

---

[4] Or by reference to a proof. For instance, see (Parzen 1960, pp.228-230).

[5] Stove does not present his argument as I have done. See (Stove 1986, pp.55-75) for his presentation.

[6] Stove endorses the logical conception of probability; Step One asserts a necessary relation between two propositions.

[7] Stove does not provide a proof. A proof can be worked out from a result in (Parzen 1960, pp.228-230).

argument, however, is to argue that certain inductive inferences are reasonable. In particular, the argument is meant to show that the inductive inference from the fact that the sample proportion is p_S to the conclusion that the population proportion is within 0.01 of p_S is reasonable, because it is reasonable to infer that the proposition "S is representative" has probability at least 0.99. But if it is reasonable to infer that some proposition has some probability, then it is reasonable for some epistemic agent. Call this agent E. Extending Step Three in this way would involve adding "Therefore, it is reasonable for E to infer that the proposition "S is representative" has probability at least 0.99".

Step Three is now implausible. Essentially it says that:

(i) q has probability at least 0.99 in relation to p (by Step One)
(ii) p is true (by Step Two)
therefore, it is reasonable for E to infer that q has probability at least 0.99.[8]

The problem is that neither Step One nor Step Two tells us anything about E's beliefs or about what E is justified in believing. Suppose that E is certain that p is false, and also has a good reason for believing that p is false. Then it would not, I submit, be reasonable for E to infer that q has probability at least 0.99. On some sorts of externalist conception of reasonable inference, Step Three may be acceptable. But I do not endorse these conceptions, and neither does Stove.

There are many possible candidate modifications to Step Three. For example, we might replace "(ii) p is true" by "(iia)E has a justified belief that p" or "(iib)E knows that p" or "(iic)E is certain that p" etc.

Notice that when we replace (ii) with one of these conditions, we will also have to replace (i) to ensure that Step Three is successful. For example, if we replace (ii) by "(ii)* p is true and E has excellent reasons for believing that p" then Step Three will become:

(i) q has probability at least 0.99 in relation to p (by Step One)
(ii)* p is true and E has excellent reasons for believing that p
therefore, it is reasonable for E to infer that q has probability at least 0.99.

The conclusion does not obviously follow unless we replace (i) by

(i)* If p is true and E has excellent reasons for believing that p, then it is reasonable for E to infer that q has probability at least 0.99.

The replacement of (ii) that I have settled on is "(ii)* p is true and E has excellent reasons for believing that p". There are two reasons I have selected this replacement rather than (iia), (iib), or (iic). First, by making (ii)* a relatively strong condition, (i)* need only be a relatively weak condition. In the course of this essay I will argue against (i)*; by showing that a relatively weak version of (i)* fails I will have shown more than I would have by showing that a relatively strong version of (i)* fails. Second, the relatively strong version of (ii)* is actually quite plausible.

Since (i) was established by Step One and (ii) was established by Step Two, replacing (i) and (ii) requires modifying Steps One and Two. Here is the proposed modification of Stove's argument. I call it **Argument 1**:

---

[8] p stands for "At least 99% of 250,000-fold samples are representative and S is a 250,000-fold sample" and q stands for "S is representative".

Step One*

Let p be any real number between 0 and 100 inclusive; F-ness and G-ness any properties; x any particular; E any epistemic agent. Step One* is to assert:

If the proposition "At least p% of the Fs are G and x is an F" is true and E has excellent reasons for believing the proposition, then it is reasonable for E to infer that the proposition "x is a G" has probability at least p/100.

Step Two*

This step is identical to Step Two except that "E has excellent reasons for believing that at least 99% of 250,000-fold samples are representative and S is a 250,000-fold sample" has been added.

Step Three*

"At least 99% of 250,000-fold samples are representative and S is a 250,000-fold sample" is true and E has excellent reasons for believing it (by Step Two*). Therefore, by an application of Step One*, it is reasonable for E to infer that "S is representative" has probability at least 0.99.

In the next section I will show that Step One* is unfortunately false. However, I have argued that if Stove's argument is to have any practical import at all—that is, if Stove's argument is going to establish that it is reasonable for some epistemic agent to infer that S is representative—the replacement of Step One by Step One*, or something similar, is required. Thus, after showing that Step One* is false, I will try to patch up Argument 1 with a second modification of Step One.

3. Why Argument 1 needs a modification

Step One* is a generalisation and can therefore be shown false by a single counterexample (there are, in fact, many counterexamples). Suppose that E is presented with an urn and that E has excellent reasons for believing the following facts about the urn:

(1) it has 1000 balls inside, each either black or white;
(2) 999 balls are white;
(3) the ball that E is about to select is in the urn;
(4) the urn is partitioned into two compartments; the black ball is in one compartment and all the white balls are in the other; the ball that E is about to select is in the compartment containing the black ball.

We have that "At least 99.9% of the balls in the urn are white and the ball that E is about to select is in the urn" is a true proposition and E has excellent reasons for believing it. But it is not the case that it is reasonable for E to infer that the probability that "E will select a white ball" is at least 0.999. Thus, Step One* is false.

Recall that Step One essentially says that for certain pairs of propositions, (P,Q), Q has probability at least p/100 in relation to P (p depends on P). Step One* essentially says, for the same pairs of propositions, that if P is true and E, where E is some epistemic agent, has excellent reasons for believing that P, then it is reasonable for E to infer that Q has probability at least p/100.

At first sight, Step One* appears to be a plausible corollary of Step One. Here are two reasons why: first, in the special case where p=100, and Q is therefore entailed by P, Step One* very naturally follows from Step One; one might therefore suppose that Step One* follows for other values of p. Second, it may seem that if asserting that Q has probability at least p/100 in relation to P is to have any practical consequences at all—that is, any consequences for reasonable inference—then it should have the consequence expressed by Step One*.

However, since Step One* is false, and Step One is not obviously false, I do not think that the former is a corollary of the latter. There are, however, claims similar to Step One*, which are not obviously false, and which may be thought to be easy corollaries of Step One, and also to capture its practical consequences. For example:

> (+) If P is true, and E, where E is some epistemic agent, has excellent reasons for believing that P, *and further, E has no beliefs other than her belief that P and her beliefs in the reasons for belief that P*, then it is reasonable for E to infer that Q has probability at least p/100.

(+) is not obviously false. (Notice that it is not shown to be false by the urn example; the reason is that if E believes the fourth fact about the urn, then it is not true that E has no beliefs other than her belief that P and her beliefs in the reasons for belief that P.) But it should be clear that (+) is not a useful replacement of Step One. We are interested in whether it is reasonable for actual epistemic agents—such as ourselves—to infer that Q. Actual epistemic agents will never satisfy the conditions of (+), because actual epistemic agents will always believe more than P and their reasons for believing that P.

(+) is the same as Step One* except that the embedded antecedent[9] has been strengthened. The problem with (+) is not that it is obviously false, but that it cannot be used to complete an argument against inductive scepticism, because it cannot be applied to actual epistemic agents. I now want to look for alternative modifications of Step One* to see if Argument 1 can be salvaged. There are two aims: first, to strengthen the embedded antecedent of Step One* in a way that gives us a true generalisation; second, to strengthen the embedded antecedent in a way that for at least some triples (E,U,S)[10] the antecedent of the relevant instance of the reformed generalisation can be shown to by satisfied. If we can fulfil these two aims then we will have a weak argument against inductive scepticism; we will have shown that for at least some triples (E,U,S) it is reasonable for E to infer that the proposition "S is representative" has probability at least 0.99.

4. A modification of Argument 1: strengthening the embedded antecedent

In the reviews of Stove (1986) the most common criticism was, roughly translated into the language of this essay, that Step One* is false and that it is not possible to achieve both of the aims set in the last paragraph. For example, Giaquinto wrote:

[9] Step One* has the form: for all x, If P_x then Q_x. By the embedded antecedent I mean P_x.

[10] E is an epistemic agent (not a theoretical one—it is best to think of E as a present-day philosopher). U is a collection of particulars and S is a 250,000-fold sample of U (as in Section 1).

> 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.[11]

The raven example employed by Giaquinto can be viewed as an attempt to refute Step One*; the idea is that we might have a large sample of black ravens, and yet it may not be reasonable for us to infer that the sample is probably representative. The claim made in the second sentence of the quote can be viewed as a proposed strengthening of the embedded antecedent of Step One*; namely, the proposal to add the condition "E knows that x is just as likely to be one F as any other" to the embedded antecedent.

Giaquinto's proposed strengthening of the embedded antecedent of Step One* looks as though it might turn Step One* into a true generalisation. Consider again the urn example of Section 3. Giaquinto's suggestion is that it is not reasonable for E to infer that the probability that "E will select a white ball" is at least 0.999, precisely because E does not have excellent reasons for believing that she is just as likely to select one ball from the urn as any other, because she has excellent reasons for believing that the ball that she will select is in the compartment containing the black ball. The suggestion is prima facie plausible, and should be investigated.

Label the condition "E has excellent reasons for believing that x is just as likely to be one F as any other" condition (C). We are interested in whether adding (C) to the embedded antecedent of Step One* turns it into a true generalisation.

The remainder of Section 4 consists of two subsections. In 4.1, I will show that (C) admits of different readings. In 4.2, I will consider various readings, and I will show that on one of these readings (C) does not turn Step One* into a true generalisation. I will also show, however, that on another reading it does.

4.1.

Look again at the urn example of Section 3. Suppose that in addition to facts (1)-(4), E also has excellent reasons for believing the following facts:

(5) Each ball has a distinct number from {1,…,1000} written on it;
(6) The number written on the black ball has been chosen by some sort of random process; perhaps, for example, by drawing numbers from a hat.

Does E have excellent reasons for believing that she is just as likely to select one ball from the urn as any other? The answer might be no, since E has excellent reasons for believing that she is about to select the black ball (by virtue of (4)). And the answer might be yes, since E has excellent reasons for believing that she is just as likely to select ball number i as ball number j, $\text{for all } i, j \in \{1,...,1000\}$ (by virtue of (6)). This is why I say that condition (C) admits of different readings.

[11] (Giaquinto 1987, p.614)

4.2.

I shall now introduce the notion of a numbering, which will provide an opportunity to propose various readings of condition (C).

Definition. A *numbering* for a finite collection of Fs is a 1-1 and onto function from {1,...,n} to the collection of Fs (where n is the number of Fs).

Let N_1 be the numbering of the balls in the urn that is given by the numbers written on the balls; that is, $N_1(a)$=ball with number a written on it. Let N_2 be any numbering of the balls that assigns number 1000 to the black ball. Then E has excellent reasons for believing that the about to be selected ball is just as likely to be ball number i as ball number j, $\forall \mathrm{i}, \mathrm{j} \in \{1,...,1000\}$, *with respect to N_1* (by virtue of (6)) but does not have excellent reasons for believing that the about to be selected ball is just as likely to be ball number i as ball number j, $\forall \mathrm{i}, \mathrm{j} \in \{1,...,1000\}$, *with respect to N_2* (by virtue of (4)).[12]

Using numberings, here are two possible readings of (C):

> (C)’ E has excellent reasons for believing that there exists a numbering of the Fs with respect to which x is just as likely to be F number i as F number j, $\forall \mathrm{i}, \mathrm{j} \in \{1,...,\mathrm{n}\}$;
>
> (C)’’ E has excellent reasons for believing that for any numbering of the Fs, x is just as likely to be be F number i as F number j, $\forall \mathrm{i}, \mathrm{j} \in \{1,...,\mathrm{n}\}$, with respect to that numbering.

(C)’ is a weak reading and (C)’’ is a strong reading. I shall now show that (C)’ is too weak by which I mean that if (C) is read as (C)’, then adding (C) to the embedded antecedent of Step One* does not result in a true conditional.

In the urn example (of subsection 4.1), the proposition “At least 99.9% of the balls in the urn are white and the about to be selected ball is in the urn” is true and E has excellent reasons for believing it. And, as we have already seen, E has excellent reasons for believing that the about to be selected ball is just as likely to be ball number i as ball number j, $\forall \mathrm{i}, \mathrm{j} \in \{1,...,1000\}$, with respect to the numbering N_1. Therefore (C) (when read as (C)’) is satisfied. But it is not the case that it is reasonable for E to infer that the proposition “the about to be selected ball is white” has probability at least 0.999.

Another option is to use (C)’’ as our reading of (C). But (C)’’ is unnecessarily strong. I shall now show, by stating and arguing for a theorem, that there is a condition, weaker than (C)’’ but stronger than (C)’, which turns Step One* into a true conditional if added to the embedded antecedent.

Theorem 1. Let p be any real number between 0 and 100 inclusive; F-ness and G-ness any properties; x any particular; E any epistemic agent. Theorem 1 asserts that:
if the proposition “At least p% of the Fs are G and x is an F” is
(1) true
and (2) E has excellent reasons for believing the proposition

---

[12] We say that E has excellent reasons for believing that x is just as likely to be F number i as F number j with respect to N_1 iff E has excellent reasons for believing that x is just as likely to be $N_1(i)$ as $N_1(j)$.

and (3) E has excellent reasons for believing that there exists a numbering of the Fs by G[13] with respect to which x is just as likely to be F number i as F number j, $\forall i, j \in \{1,...,n\}$;

then it is reasonable for E to infer that the proposition "x is a G" has probability equal to p/100.

Proof. (3) is equivalent to "E has excellent reasons for believing that there exists a numbering of the Fs by G, N, such that x is just as likely to be N(i) as N(j), $\forall i, j \in \{1,...,n\}$ (where n is the number of Fs). Let A_i denote the proposition "N(i) is a G". Let $Pr_E(p) = \alpha$ (where p is a proposition) be shorthand for "it is reasonable for E to believe that the probability of p is α". By the law of total probability, we have that:

$$\begin{aligned} Pr_E(\text{x is a G}) &= \sum_{i=1}^{n} Pr_E(\text{x is a G} \mid \text{x is N(i)}) Pr_E(\text{x is N(i)}) \\ &= \sum_{i=1}^{n} Pr_E(A_i \mid \text{x is N(i)}) Pr_E(\text{x is N(i)}) \qquad \text{(by definition of } A_i\text{)} \end{aligned}$$

Lemma.

Let m = number of Fs that are Gs.

Then $Pr_E(A_i \mid \text{x is N(i)}) = \begin{cases} 1 \text{ if } i \in \{1,...,m\} \\ 0 \text{ if } i \in \{m+1,...,n\} \end{cases}$

Proof of lemma. If $i \in \{m+1,...,n\}$, then $Pr_E(A_i) = 0$. Thus,

$$Pr_E(A_i \mid \text{x is N(i)}) = \frac{Pr_E(\text{x is N(i)} \mid A_i) Pr_E(A_i)}{Pr_E(\text{x is N(i)})} = 0$$

If $i \in \{1,...,m\}$, then $Pr_E(A_i) = 1$ and $Pr_E(\neg A_i) = 0$. Thus,

$$Pr_E(\neg A_i \mid \text{x is N(i)}) = \frac{Pr_E(\text{x is N(i)} \mid \neg A_i) Pr_E(\neg A_i)}{Pr_E(\text{x is N(i)})} = 0$$

and since $Pr_E(A_i \mid \text{x is N(i)}) = 1 - Pr_E(\neg A_i \mid \text{x is N(i)})$, $Pr_E(A_i \mid \text{x is N(i)}) = 1$.

Applying the lemma to the above yields, $Pr_E(\text{x is a G}) = \sum_{i=1}^{m} Pr_E(\text{x is N(i)})$

[13] A numbering of the Fs by G is defined to be a numbering of the Fs, using the numbers {1,...,n}, such that each F that is a G has a number in {1,...,m} and each F that is not a G has a number in {m+1,...,n} (where n=the number of Fs, and m=number of Fs that are Gs).

I take it that condition (3) of Theorem 2 implies that $\Pr_E(\text{x is N(i)}) = \Pr_E(\text{x is N(j)})$ for all i, $j \in \{1,...,n\}$. Therefore, since $\sum_{i=1}^{n} \Pr_E(\text{x is N(i)}) = 1$, we have that

$\Pr_E(\text{x is N(i)}) = \frac{1}{n}$ for each $i \in \{1,...,n\}$.

Thus, $\Pr_E(\text{x is a G}) = \sum_{i=1}^{m} \Pr_E(\text{x is N(i)}) = \frac{m}{n} = \frac{p}{100}$.

□[14]

Thus, if (C) is read as condition (3), then adding (C) to Step One* does result in a true generalisation. If we replace Step One* by Theorem 1, and make the necessary adjustment[15] to Step Two*, then we will have another possible argument against inductive scepticism. Call this **Argument 2**.

It may be difficult to understand the content of condition (3), so here is another condition, perhaps easier to grasp, that implies (3):

> E has excellent reasons for believing that x is just as likely to be y as z, for any pair (y,z) such that y is an F that is a G and z is an F that is not a G.

5. One of the objections to Stove's argument rebutted

There is an objection that is directed against a particular kind of application of Stove's argument. The kind of application is that which seeks to ground inferences to generalisations about the future. Here is an example: let U be the collection of the 500,000 consecutive days of which tomorrow is the first day, and let E be some epistemic agent. Suppose that E today decides to observe for each of the earliest 250,000 days of U whether or not the day has the property of having a sunrise.[16] Call the collection of 250,000 observations S. Since S is a 250,000-fold sample of U, Stove wants to be able to say that it is reasonable for E to infer that the proposition "S is representative" has probability at least 0.99.

Returning to the above quote from Giaquinto, and the raven example he gives, Giaquinto is putting forward something like the claim: "If E does not know that she is just as likely to end up with one sample as any other of the same size, then it is not reasonable for E to infer that her sample is probably representative." Other reviewers have held this view, and have used it to argue that, in the case where we are trying to ground an inference to a generalisation about the future, since no part of our sample can come from the future, we know that we are not just as likely to end up with one sample as any other sample of the same size. Therefore, we cannot reasonably infer that our sample is probably representative.

The work of Section 4—particularly Theorem 1—allows us to show that this argument fails. Applying Theorem 1 to the sunrise example, the result is:

> *If* E has excellent reasons for believing that there exists a numbering of the

---

[14] The proof assumes that $\Pr_E$(x is N(i)) is non-zero $\forall i \in \{1,...,n\}$. This is reasonable because if $\Pr_E$(x is N(i))=0 for some i, then $\Pr_E$(x is N(i))=0 for all i (by condition (3)).

[15] Namely, include in Step Two* an argument for the relevant instance of condition (3).

[16] Ignore that a person cannot live for 250,000 days. E could be a succession of individuals lasting for at least 250,000 days.

250,000-fold samples by the property of being representative with respect to which S is just as likely to be sample number i as sample number j, $\forall i, j \in \{1,...,n\}$;

then it is reasonable for E to infer that the proposition "S is representative" has probability at least 0.99.

Call this conditional (†). The objection that we are considering fails, because the antecedent of (†) may be satisfied, even though E knows that S is the sample that consists of the earliest 250,000 days of U, and so E knows that in one sense S is not just as likely to be one 250,000-fold sample as any other. It may be hard to see why this is so, partly because the number of 250,000-fold samples is so large. For this reason, I shall now give an example that can be viewed as a micro-model of the sunrise example.

Suppose that E is presented with a tube, the ends of which are labelled A and B, and that E has excellent reasons for believing the following facts about the tube:

(1) It has 100 balls inside, arranged in single file, each either black or white;
(2) 99 balls are white and 1 ball is black;
(3) E is about to select the ball from end A of the tube.

The balls play the role of samples, and the property of being white plays the role of the property of being representative. Just as at least 99% of 250,000-fold samples are representative, at least 99% of the balls are white. Now consider the condition:

(4) E has excellent reasons for believing the following facts about the way in which the tube was loaded:
first, 100 white balls were put into the tube;
second, a number between 1 and 100 was chosen by some sort of random process; perhaps, for example, by drawing a number from a hat (call the number n);
third, the ball in position n was painted black (where position 1 is next to end A and position 100 is next to end B).

(4) is consistent with (1), (2), and (3). In particular, (4) is consistent with E having excellent reasons for believing that in one sense she is not just as likely to select one ball as any other, because it is consistent with her having excellent reasons for believing that she is going to select the ball at end A. If (4) is true, then the following is satisfied:

E has excellent reasons for believing that there exists a numbering of the balls in the tube by the property of being white with respect to which the ball that E is about to select (that is, the ball at end A) is just as likely to be ball number i as ball number j, $\forall i, j \in \{1,...,100\}$.[17]

Similarly, the sunrise example is consistent with:

[17] Therefore, by Theorem 1, it is reasonable for E to infer that the proposition “the about to be selected ball is white” has probability at least 0.99.

> (4)' E has excellent reasons for believing that the positions[18] of the unrepresentative samples were determined by some sort of random process.[19]

If (4)' is true, then the antecedent of (†) is satisfied. Thus, the antecedent of (†) may be satisfied even though E knows that in one sense S is not just as likely to be one 250,000-fold sample as any other.

6. Why I doubt that Argument 2 will be successful

At the end of Section 3, I set two aims for a modification of Step One*. My proof of Theorem 1 is my argument that Theorem 1 achieves the first aim. In this section, I sketch why I doubt that Theorem 1 achieves the second aim. This is disappointing, because it is only if Theorem 1 fulfils both aims that it can serve as a basis for an argument against inductive scepticism.

Recall that the second aim was to strengthen the embedded antecedent in a way that for at least some triples (E,U,S) the antecedent of the relevant instance of the reformed generalisation can be shown to be satisfied.

Let us look at a particular triple (E,U,S). Suppose that in 2050 a new planet is discovered—planet Yog. It is known that over a million birds live on Yog. E is a philosophy professor who is flown to Yog to investigate whether all birds on Yog in the year 2050 are black. U is the collection of birds on Yog in the year 2050. When E lands on Yog she starts collecting a 250,000-fold sample. S is this sample.

The only difference between the embedded antecedent of Step One* and of Theorem 1 is condition (3). Thus, we need to argue that:

> (*) E has excellent reasons for believing that there exists a numbering of the 250,000-fold samples by the property of being representative with respect to which S is just as likely to be sample number i as sample number j, $\forall i, j \in \{1,...,n\}$.

The best way to argue for (*) is by arguing for:[20]

> (**) E has excellent reasons for believing that Pr_E(S is A)=Pr_E(S is B) for any pair (A,B) such that A is a representative 250,000-fold sample and B is an unrepresentative 250,000-fold sample.[21]

I suggest that the only way that E might argue for Pr_E(S is A)=Pr_E(S is B) is by an

---

[18] To speak of positions presupposes an ordering of the 250,000-fold samples. The ordering I have in mind is chronological, and puts sample S in position 1. Formally, first observe that any 250,000-fold sample can be uniquely represented as $(a_1,a_2,\ldots,a_{250,000})$ where a_i is the ith earliest day observed in the sample. Then, for two samples, A = $(a_1,a_2,\ldots,a_{250,000})$ and B = $(b_1,b_2,\ldots,b_{250,000})$ say that A<B iff:

1. a_1 is earlier than b_1

or 2. $a_1=b_1$ and a_2 is earlier than b_2

or …

or 250,000. $a_1=b_1$ and $a_2=b_2$ and…and $a_{249,999}=b_{249,999}$ and $a_{250,000}$ is earlier than $b_{250,000}$.

Thus, just as the balls can be ordered by their position in the tube, the 250,000-fold samples can be ordered by this chronological ordering.

[19] I do not argue that (4)' is true or plausible; just that it is consistent with the sunrise example.

[20] See the very end of Section 4.

[21] It is important to realise that, since the conception of probability involved is logical, (**) is necessarily true if true.

application of the Principle of Indifference (henceforth referred to as (PI)). Assume that E knows nothing about the birds on Yog that she has not observed (except that there are at least 750,000 of them). Then it is plausible to suppose that given any pair (A,B) (such that A is a representative 250,000-fold sample and B is an unrepresentative 250,000-fold sample), she has no more reason to think that S is A than that S is B, and vice versa. If so, she can apply (PI) to conclude that Pr_E(S is A)=Pr_E(S is B).

It is well known that there are arguments that counsel us not to trust probability assignments recommended by (PI).[22] The central problem with (PI) is that it can often be applied in more than one way, yielding more than one probability assignment for a particular proposition. Probability assignments recommended by (PI) thus begin to look arbitrary.

We can see an example of this in the case of planet Yog: suppose that E dreams up a scenario. She imagines an omnipotent demon (she is, after all, a philosophy professor) who had never before paid any attention to her inductive inferences, but who decided to surreptitiously make sure that E's raven inference went wrong. The demon—being omnipotent—ensured that the birds presented to E made up an unrepresentative sample.

Let DS be shorthand for the demon scenario. Suppose that by an application of (PI), E decides that $\mathrm{Pr_E}(\mathrm{DS}) = \mathrm{Pr_E}(\neg\mathrm{DS}) = 0.5$. The problem is that this application precludes the application that yields (**), as we shall now see:

Claim. If Pr_E(DS)>0.01 then (**) does not obtain.

Proof. Let k = Pr_E(DS). DS implies that S is unrepresentative (and E, we can assume, is aware of this implication). Thus, Pr_E(S is representative|DS) = 0. Now, by the law of total probability:

$$
\begin{aligned}
&\mathrm{Pr_E}(\text{S is representative}) = \\
&\qquad \mathrm{Pr_E}(\text{S is representative} \mid \mathrm{DS})\mathrm{Pr_E}(\mathrm{DS}) + \mathrm{Pr_E}(\text{S is representative} \mid \neg\mathrm{DS})\mathrm{Pr_E}(\neg\mathrm{DS}) \\
&\quad = \mathrm{Pr_E}(\text{S is representative} \mid \neg\mathrm{DS})(1 - \mathrm{k}) \\
&\quad \leq (1 - \mathrm{k})
\end{aligned}
$$

We are assuming that Theorem 1 is true (because we want to apply it). Therefore, if (**) is satisfied then Pr_E(S is representative) = 0.99. But if k > 0.01, then Pr_E(S is representative) < 0.99.

□

Thus, if E wants to apply (PI) to argue for (**), she must explain why that application is warranted whereas the application to the $(\mathrm{DS}, \neg\mathrm{DS})$ pair is not.

There may be an explanation available, but I doubt that it will be convincing. One option is for E to dream up at least a hundred distinct scenarios that imply that S is representative. Applying (PI) to these scenarios together with DS would yield Pr_E(DS)<0.01. The problem with this option is that it is obviously ad hoc. Another option is to construct an argument, not relying on (PI), that $\mathrm{Pr_E}(\mathrm{DS}) \neq 0.5$. Since applying (PI) to the $(\mathrm{DS}, \neg\mathrm{DS})$ pair yields $\mathrm{Pr_E}(\mathrm{DS}) = 0.5$, such an argument would

[22] See (Gillies 2000, pp.37-42) for a survey of the well-known paradoxes associated with (PI).

provide a reason not to apply (PI) to the (DS,$\neg$DS) pair. The problem is that it is difficult to imagine that there is such an argument. An empirical argument does not appear possible as the demon scenario is empirically undetectable. And an a priori argument is likely to fall back on (PI).

Even if E can explain why the application of (PI) used to argue for (**) is warranted whereas the application to the (DS,$\neg$DS) pair is not, there will be other applications of (PI) that preclude the application that yields (**). E will also have to explain why those applications are not warranted.

In sum, there are reasons to suppose that (PI) will not provide a firm enough platform to convince the sceptic that (**) is satisfied. Yet using (PI) appears to be the only way to argue for (**).

Conclusion

The three main achievements of the essay are as follows: first, using numberings I showed how to formalise various readings of (C). Second, I proved that on one reading adding (C) to the embedded antecedent of Step One* results in a true generalisation (i.e. Theorem 1). Third, I showed in Section 5 that one of the commonest objections to Stove's argument misses the point. Sacrificing accuracy for brevity, we can say that the objection is unsuccessful because it fails to notice that selection can be random in one respect whilst determinate in another. In particular, selection of a sample can be random vis-à-vis the property of being representative (i.e. condition (3) of Theorem 1 can be satisfied) even though no part of a sample can come from the future.

In Section 6 I explained why I doubt that Theorem 1 can be used to complete an argument against inductive scepticism. But I do not claim that Stove's argument has been refuted; even if my doubts prove to be well founded, there may be other reconstructions of the argument that are successful. Rather I hope to have illuminated some of the issues involved in trying to make Stove's argument work.

Bibliography

Giaquinto, M. (1987), 'Book Review: The Rationality of Induction', *Philosophy of Science,* Vol. 54, No. 4, 612-615

Gillies, D. (2000), *Philosophical Theories of Probability* (London: Routledge)

Millican, P. (2002), 'Hume's Sceptical Doubts concerning Induction', in P. Millican (ed.), *Reading Hume on Human Understanding* (Oxford University Press)

Parzen, E. (1960), *Modern Probability Theory and Its Applications* (New York: John Wiley & Sons)

Stove, D. (1986), *The Rationality of Induction* (Oxford University Press)