

OBSERVER-RELATIVE CHANCES IN ANTHROPIC REASONING?*

(c) Nick Bostrom

Dept. of Philosophy, Logic and Scientific method
London School of Economics; Houghton St.; WC2A AE; London; UK

Email: n.bostrom@lse.ac.uk

Homepage: <http://www.analytic.org>

11 October, 1999

ABSTRACT

John Leslie presents a thought experiment to show that chances are sometimes observer-relative in a paradoxical way. The pivotal assumption in his argument – a version of the weak anthropic principle – is the same as the one used to get the disturbing Doomsday argument off the ground. I show that Leslie's thought experiment trades on the sense/reference ambiguity and is fallacious. I then describe a related case where chances *are* observer-relative in an interesting way. But not in a paradoxical way. The result can be generalized: At least for a very wide range of cases, the weak anthropic principle does *not* give rise to paradoxical observer-relative chances. This finding could be taken to give new indirect support to the doomsday argument.

1. INTRODUCTION

In a recent paper [1], John Leslie argues that a version of the weak anthropic principle gives rise to a paradoxical kind of observer-relative chances:

Estimated probabilities can be observer-relative in a somewhat disconcerting way: a way not depending on the fact that, obviously, various observers often are unaware of truths which other observers know. (p. 435)

The anthropic assumption that is used in deriving these observer-relative chances is that: Any observer should regard herself as a random sample from (some suitable subset) of the set of all observers. We can call it the *self-sampling assumption*. [2]

Leslie is well-known as the most prolific proponent of the so-called Doomsday argument [3]. The Doomsday argument purports to show that the risk that humankind will go extinct soon is much greater than has previously been thought. The self-sampling assumption is what is used to get the Doomsday argument off the ground.

Leslie's own reaction to his discovery that the self-sampling assumption generates chances that are observer-relative is to bite the bullet and accept this consequence: "Any air of paradox must not prevent us from accepting these things." (p. 428).

A person less convinced than Leslie about the validity of the Doomsday argument might rather view the finding as a *reductio*, showing that the self-sampling assumption should be rejected. This outcome would also have the pleasant feature that it would allow us to dispose of the disturbing Doomsday argument.

Unfortunately, it is not true that the self-sampling assumption gives rise to paradoxically observer-relative chances in the way Leslie imagines. In the next section I will try to show that Leslie's argument rests on an ambiguity and that when this ambiguity is eliminated, so are the observer-relative chances. The remainder of this paper will then investigate whether there might be other situations where chances *are* observer-relative in some interesting sense. We shall answer this in the negative: at least for a very general set of cases, the self-sampling assumption does *not* give rise to problematic observer-relative chances. If this result is correct it could be taken to give some new indirect support for the Doomsday argument.

2. LESLIE'S ARGUMENT AND WHY IT FAILS

Leslie's argument for observer-relative chances takes the form of a thought experiment. Suppose we start with a batch of one hundred women and divide them randomly into two groups, one with ninety-five and one with five women. By flipping a fair coin, we then assign the name 'the Heads group' randomly to one of these groups and the name 'the Tails group' to the other. According to Leslie it is now the case that an external observer, i.e. a person not in either of the two groups, ought to derive radically different conclusions than an insider:

All these persons – the women in the Heads group, those in the Tails group, and the external observer – are fully aware that there are two groups, and that each woman has a ninety-five per cent chance of having entered the larger. Yet the conclusions they ought to derive differ radically. The external observer ought to conclude that the probability is fifty per cent that the Heads group is the larger of the two. Any woman actually in [either the Heads or the Tails group], however, ought to judge the odds ninety-five to five that her group, identified as 'the group I am in', is the larger, regardless of whether she has been informed of its name. (p. 428)

Even without knowing her group's name, a woman could still appreciate that the

external observer estimated its chance of being the larger one as only fifty per cent – this being what his evidence led him to estimate *in the cases of both groups*. The paradox is that she herself would then have to say: 'In view of my evidence of being in one of the groups, ninety-five per cent is what I estimate.' (p. 429)

Somewhere within these two paragraphs a mistake has been made. It's not hard to locate the error if we look at the structure of the reasoning. Let's say there is a woman in the larger group who is called Chris. The "paradox" then takes the following form:

- (1) $\Pr_{\text{Chris}}(\text{"The group that Chris is in is the larger group"}) = 95\%$
- (2) The group that Chris is in = the Heads group
- (3) Therefore: $\Pr_{\text{Chris}}(\text{"The Heads group is the larger group"}) = 95\%$
- (4) But $\Pr_{\text{External observer}}(\text{"The Heads group is the larger group"}) = 50\%$
- (5) Hence chances are observer-relative.

Where it goes wrong is in step (3). The group that Chris is in is indeed identical to the Heads group, but Chris doesn't know that. $\Pr_{\text{Chris}}(\text{"The Heads group is the larger group"}) = 50\%$, not 95% as claimed in step (3). There is nothing paradoxical or surprising about this, at least not any longer after Frege's discussion of Hesperus and Phosphorus. One need not have rational grounds for assigning probability one to the proposition "Hesperus = Phosphorus", even though as a matter of fact Hesperus = Phosphorus. For one might not know that Hesperus = Phosphorus. The expressions 'Hesperus' and 'Phosphorus' present their denotata under different modes of presentation: they denote the same object while connoting different concepts. There is disagreement over exactly how to describe this difference and what lesson to learn from it; but the basic observation that you can learn something from being told " $a = b$ " (even if $a = b$) is neither new nor has it very much in particular to do with self-sampling assumption.

Let's see if there is some way we could rescue Leslie's conclusion by modifying the thought experiment.

Suppose that we change the example so that Chris now knows that the sentence "Chris is in the Heads group" is true. Then step (3) will be correct. However, we will now run into trouble when we try to take step (5). It will no longer be true that Chris and the external observer know about the same facts. Chris's information set now contains the sentence "Chris is in the Heads group". The external observer's information set doesn't contain this piece of information. So no interesting form of observer-relative chances has been established.

What if we change the example again and assume that the external observer, too, knows that Chris is in the Heads group? Well, if Chris and the external observer agreed on the chance that the Heads group is the large group before they both learnt that Chris is in the Heads group, then they will continue to be in agreement about this chance after they have received that information – *provided* they agree about the conditional probability $\Pr(\text{The Heads group is the larger group} \mid \text{Chris is in the Heads group})$. This, then, is what we have to examine to see if any paradoxical conclusion can be wrung from Leslie's set-up: we have to check whether Chris and the outside observer agree on this conditional probability.

First look at it from Chris' point of view. Let's go along with Leslie and assume that she should think of herself as a random sample from the batch of one hundred women. Suppose she knows that her name is Chris (and that she's the only woman in the batch with that name). Then, before she learns that she is in the Heads group, she should think that the probability of that being the case is 50%. (Recall that what group should be called 'the Heads group' was determined by tossing of a fair coin.) She should think that the chance of the sentence "Chris is in the larger group" is 95%, since ninety-five out of the hundred women are in the larger group, and she can regard herself as a random sample from these hundred women. When learning that she is in the Heads group, the chance of her being in the larger group remains 95%. ('the Heads group' and 'the Tails group' are just arbitrary labels at this point; randomly calling one group the Heads group doesn't change the likelihood that it is the big group.) Hence the probability she should give to the sentence "The Heads group is the larger group" is now 95%. Therefore the conditional probability which we were looking for is $\Pr_{\text{Chris}}(\text{"The Heads group is the larger group"} \mid \text{"Chris is in the Heads group"}) = 95\%$.

Next consider the situation from the external observer's point of view. What is the probability for the external observer that the Heads group is larger given that Chris is in it? Well, what's the probability that Chris is in the Heads group? In order to answer these questions, we need to know something about how this woman Chris was selected.

Suppose that she was selected as a random sample (with uniform sampling density) from among all the hundred women in the batch. Then the external observer would arrive at the same conclusion as Chris: if the random sample 'Chris' is in the Heads group then there is a 95% chance that the Heads group is the bigger group.

If we instead suppose that Chris was selected randomly from some subset of the hundred women, then it might happen that the external observer's estimate diverges from Chris'. For example, if the external observer randomly selects one individual x (whose name happens to be 'Chris') from the large group, then, when he finds that x is in the Heads group, he should assign a 100% probability to the sentence "The Heads group is the larger group." This is indeed a different conclusion than the one which the insider Chris

draws. *She* thought the conditional probability of the Heads group being the larger given that Chris is in the Heads group was 95%.

In this case, however, we have to question whether Chris and the external observer know about the same evidence. (If they don't, then the disparity in their conclusions doesn't signify that chances are observer-relative in any paradoxical sense.) But it is clear that their information sets *do* differ in a relevant way. For suppose Chris got to know what the external observer is stipulated to already know: that Chris had been selected by the external observer through some random sampling process from among a certain subset of the hundred women. That implies that Chris is a member of that subset. This information would change her probability estimate so that it once again becomes identical to the external observer's. In the above case, for instance, the external observer selected a woman randomly from the large group. Now, evidently, if Chris gets this extra piece of information, that she has been selected as a random sample from the large group, then she knows with certainty that she is in that group; so her conditional probability that the Heads group is the larger group given that Chris is in the Heads group should then be 100%, the same as what the outside observer should believe.

We see that as soon as we give the two people access to the same evidence, their disagreement vanishes. There are no paradoxical observer-relative chances in this thought experiment.[\[4\]](#)

3. OBSERVER-RELATIVE CHANCES: ANOTHER TRY

In this section I shall give an example where chances could actually be said to be observer-relative in an interesting – though by no means paradoxical – sense. What philosophical lessons we should or shouldn't draw from this phenomenon will be discussed in the next section.

Here is the example:

Suppose the following takes place in an otherwise empty world. A fair coin is flipped by an automaton and if it falls heads, ten humans are created; if it falls tails, one human is created. Suppose that in addition to these people there is one additional human that is created independently of how the coin falls. This latter human we call *the bookie*. The people created as a result of the coin toss we call *the group*. Everybody knows about these facts. Furthermore, the bookie knows that she is the bookie and the people in the group know that they are in the group.

The question is what are the fair odds if the people in the group want to bet against the bookie on how the coin fell? One could think that everybody should agree that the chance of it having fallen heads is fifty-fifty, since it was a fair coin. That overlooks the fact that the bookie obtains information from finding that she is the bookie rather than one of the people in the group. This information is relevant to her estimate of how the coin fell. It is more likely that she should find herself being the bookie if one out of two is a bookie than if the ratio is one out of eleven. So finding herself being the bookie, she obtains reason to believe that the coin probably fell tails, leading to the creation of only one other human. In a similar way, the people in the group, by observing that they are in the group, obtain some evidence that the coin fell heads, resulting in a large fraction of all observers observing that they are in the group.

It is a simple exercise to use Bayes' theorem to calculate what the posterior probabilities are after this information has been taken into account.

Since the coin is fair, we have $\Pr(\text{Heads}) = \Pr(\text{Tails}) = \frac{1}{2}$.

By the self-sampling assumption, $\Pr(\text{I am bookie} \mid \text{Heads}) = \frac{1}{11}$ and $\Pr(\text{I am bookie} \mid \text{Tails}) = \frac{1}{2}$. Hence,

$$\begin{aligned} \Pr(\text{I am bookie}) &= \Pr(\text{I am bookie} \mid \text{Heads}) \cdot \Pr(\text{Heads}) + \Pr(\text{I am bookie} \mid \text{Tails}) \cdot \Pr(\text{Tails}) \\ &= \frac{1}{11} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{44}. \end{aligned}$$

From Bayes' theorem we then get:

$$\begin{aligned} \Pr(\text{Heads} \mid \text{I am bookie}) &= \Pr(\text{I am bookie} \mid \text{Heads}) \cdot \Pr(\text{Heads}) / \Pr(\text{I am bookie}) \\ &= \frac{\frac{1}{11} \times \frac{1}{2}}{\frac{3}{44}} = \frac{2}{3}. \end{aligned}$$

In exactly the same way we get the odds for the people in the group:

$\Pr(\text{I am in the group} \mid \text{Heads}) = 10/11$ and $\Pr(\text{I am in the group} \mid \text{Tails}) = \frac{1}{2}$.

$$\Pr(\text{I am in the group}) = \frac{10}{11} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{31}{44}.$$

$$\Pr(\text{Heads} \mid \text{I am in the group}) = \Pr(\text{I am in the group} \mid \text{Heads}) \cdot \Pr(\text{Heads}) / \Pr(\text{I am in the group})$$

am in the group)

$$= \frac{\frac{10}{11} \times \frac{1}{2}}{\frac{31}{44}} = \frac{20}{31}.$$

We see that the bookie should think there is a $\frac{2}{13}$ chance that the coin was heads while the people in the group should think that the chance is $\frac{20}{31}$. This is a consequence of the self-sampling assumption.

Shall we conclude that the self-sampling assumption has paradoxical implications and should be rejected? That would be too rash.

4. PHILOSOPHICAL DISCUSSION

While it might be slightly noteworthy that the bookie and the people in the group are rationally required to disagree in the above scenario, it isn't the least bit paradoxical. Their information sets are not identical. For instance, the bookie knows that "I am the bookie.". That is clearly a different proposition from the corresponding one – "I am in the group." – known by the people in the group. So chances have *not* been shown to be observer-relative in the sense that people with the same information can be rationally required to disagree. And if we were to try to modify the example so as to make the participants' information sets identical, we would see that their disagreement disappears; as it did when we attempted various twists of the Leslie example.

There is a sense, though, in which the chances in the present example can be said to be observer-relative. The information sets of the bookie and the people in the group, while not identical, are quite similar. They differ only in regard to such indexical facts as "I am the bookie." or "I am in the group.". We could say that the example demonstrates, in an interesting way, that chances can be relative to observers in the sense that: people with information sets that are identical *up to indexical facts* can be rationally required to disagree about non-indexical facts.

This kind of observer-relativity is not particularly counterintuitive and it should not be taken to cast doubt on the self-sampling assumption from which it was derived. That indexical facts can have implications for what we should believe about nonindexical facts shouldn't surprise us. It can be shown by a very simple example: from "I have blue eyes." it follows that somebody has blue eyes.

The rational odds in the example above being different for the bookie than for the people in the group, we might begin to wonder whether it is possible to formulate some kind of bet for which all parties would calculate a positive expected payoff? This would not necessarily be an unacceptable consequence since the bettors have different information. Still, it could seem a bit peculiar if we had a situation where purely by applying the self-sampling assumption rational people were led to start placing bets against each other.

So it is worth calculating the odds to see if there are cases where they do indeed favour betting. This is done in the appendix. The result turns out to be negative – no betting. In the quite general class of cases considered, there is no combination of parameter values for which a bet is possible in which both parties would rationally expect a positive non-zero payoff.

This result is reassuring for the proponent of anthropic reasoning. Yet we are still left with the fact that there are cases where observers come to disagree with each other just because of applying the self-sampling assumption. While it is true that these disagreeing observers will have different indexical information, and while there are trivial examples in which a difference in indexical information implies a difference in non-indexical information, it might nonetheless be seen as objectionable that anthropic reasoning should lead to this kind of disagreements. Does not that require that we ascribe some mysterious quality to the things we call 'observers', some property of an observer's mind that cannot be reduced to objective observer-independent facts?

The best way to resolve this scruple is to show how the above example, where the "observer-relative" chances appeared, can be restated in purely physicalistic terms:

A coin is tossed and either one or ten human brains are created. These brains make up 'the group'. Apart from these there is only one other brain, the 'bookie'. All the brains are informed about the procedure that has taken place. Suppose Alpha is one of the brains that have been created and that Alpha remembers recently having been in the brain states A_1, A_2, \dots, A_n .

At this stage, Alpha should obviously think the probability that the coin fell heads is 50%, since it was a fair coin. But now suppose that Alpha is informed that he is the bookie, i.e. that the brain that has recently been in the states A_1, A_2, \dots, A_n is the brain that is labeled 'the bookie'. Then Alpha will reason as follows:

"Let A be the brain that was recently in states A_1, A_2, \dots, A_n . The conditional probability that A is labeled 'the bookie' given that A is one of two existing brains is greater than the conditional probability that A is the brain labeled 'the bookie' given that A is one out of eleven brains. Hence, since A does indeed turn out to be the brain labeled 'the bookie', there is a greater than 50% chance that the coin fell tails, creating only one brain."

A parallel line of reasoning can be pursued by a brain labeled 'a brain in the group'. The argument can be quantified in the same way as in the earlier example and will result in the same "observer-relative" chances. This shows that anthropic reasoning can be understood in a purely physicalistic framework.

The observer-relative chances in this example too are explained by the fact that the brains have access to different evidence. Alpha, for example, knows that (P_{Alpha}) *the brain that has recently been in the states A_1, A_2, \dots, A_n is the brain that is labeled 'the bookie'*. A brain, Beta, who comes to disagree with Alpha about the probability of heads, will have a different information set. Beta might for instance rather know that (P_{Beta}) *the brain that has recently been in the states B_1, B_2, \dots, B_n is a brain that is labeled 'a member of the group'*. P_{Alpha} is clearly not equivalent to P_{Beta} .

It is instructive to see what happens if we take a step further and eliminates from the example not only all non-physicalistic terms but also its ingredient of indexicality:

In the previous example we assumed that the proposition P_{Alpha} which Alpha knows but Beta does not know was a proposition concerning the brain states A_1, A_2, \dots, A_n of Alpha itself. Suppose now instead that Alpha does *not* know what label the brain Alpha has (whether it is 'the bookie' or 'a brain in the group') but that Alpha has been informed that there are some recent brain states G_1, G_2, \dots, G_n of some other existing brain, Gamma, and that Gamma is labeled 'the bookie'.

At this stage, what conclusion Alpha should draw from this piece of information is underdetermined by the specifications we have given. *It would depend on what Alpha would know or guess about how this other brain Gamma had been selected to come to Alpha's notice.* Suppose we specify the thought experiment further by stipulating that, as far as Alpha's knowledge goes, Gamma can be regarded as a random sample from the set of all existing brains. Alpha may know, say, that one ball for each existing brain was put in an urn and that one of these balls was drawn at random and it turned out to be the one corresponding to Gamma. Reasoning from this information, Alpha will arrive at the same conclusion as if Alpha had learnt that Alpha was labeled 'the bookie' as was the case in the previous version of the thought experiment. Similarly, Beta may know about another random sample, Epsilon, that is labeled 'a brain in the group'. This will lead Alpha and Beta to differ in their probability estimates, just as before. In this version of the thought experiment no indexical evidence is involved. Yet Alpha's probabilities differ from Beta's.

What we have here is hardly distinct from any humdrum situation where John and Mary know different things and therefore estimate probabilities differently. The only difference between the present example and a commonplace urn game is that here we are dealing in brains whereas there we are dealing in raffle tickets – surely not philosophically relevant.

But what exactly did change when we removed the indexical element? If we compare the two last examples we see that the essential disparity is in how the random samples were produced.

In the second of the two examples there was a *physical selection mechanism* that generated the randomness. We said that Alpha knew that there was one ball for each brain in existence, that these balls had been put in an urn, and that one of these balls had then been selected randomly and had turned out to correspond to a brain that was labeled 'the bookie'.

In the other example, by contrast, there was no such physical mechanism. Instead, there the randomness did somehow *arise from each observer considering herself as a random sample from the set of all observers*. Alpha and Beta observed their own states of mind (i.e. their own brain states). Combining this information with other, non-indexical, information allowed them to draw conclusions about non-indexical states of affairs that they could not draw without the indexical information obtained from observing their own states of mind. But there was no physical randomization mechanism at work analogous to selecting a ball from an urn.

Now, it is indeed problematic how such reasoning can be justified or explained. The principle that an observer should regard herself as a random sample from (a suitable subset of?) the set of all observers who will ever have existed is not trivial – how can there be a random sample in the absence of any randomizing physical selection procedure? What, precisely, is the class from which I am supposed to be a random sample? Does it include observers that will in fact be born in the future even if the laws of physics have not yet determined whether these future observers will come to exist? However, this principle – the self-sampling assumption – is what is used to get anthropic reasoning [5] off the ground in the first place. So the discovery that the weak anthropic principle leads to "observer-relative" chances, and that these chances arise from this problematic principle, is not something that should be taken to add new reasons for being skeptical about anthropic reasoning. It is rather a matter of restating an assumption from which we started.

5. CONCLUSIONS

Leslie's argument that there are cases where anthropic reasoning gives rise to paradoxical observer-relative chances does not hold up to scrutiny. I have argued that it rests on a sense/reference ambiguity and that when this ambiguity is resolved then the purported observer-relativity disappears. Several ways that one could try to rescue Leslie's conclusion were examined and it turned out that none of them would

work.

We then considered an example where observers applying the self-sampling assumption end up disagreeing about the outcome of a coin toss. The observers' disagreement depends on a difference in their information sets and is not of a paradoxical nature; there are completely trivial examples of the same kind of phenomenon. I also showed that (at least for a wide range of cases) this disparity in beliefs cannot be marshaled into a betting arrangement where all parties involved would expect to make a gain.

This example was given a physicalistic reformulation showing that the observers' disagreement does not imply some mysterious irreducible role for the observers' consciousness. What does need to be presupposed, however, unless the situation be utterly trivialized, is the self-sampling assumption (i.e. that each observer should regard herself as a random sample from among (some suitable subset) of all observers who will ever have existed). This is not a finding that should be taken to cast *new* doubt on the weak anthropic principle. Rather, it's a restatement of what that principle really means.

APPENDIX

In this appendix it is shown for a quite general set of cases that adopting and applying the self-sampling assumption does not lead rational agents to bet against each other.

Consider again the case where a fair coin is tossed and a different number of observers are created depending on how the coin falls. The people created as a result of the coin toss make up 'the group'. In addition to these there exists a set of people we call the 'bookies'. Together, the people in the group and the bookies make up the set of people who are said to be 'in the experiment'. To make the example more general, we also allow there to be (a possibly empty) set of observers who are not in the experiment (i.e. who are not bookies and are not in the group); we call these observers 'outsiders'.

We introduce the following abbreviations:

Number of people in the group if coin falls heads $=h$
 Number of people in the group if coin falls tails $=t$
 Number of bookies $=b$
 Number of outsiders $=u$
 For "*The coin fell heads.*", write H
 For "*The coin fell tails.*", write $\neg H$
 For "*I am in the group.*", write G
 For "*I am a bookie.*", write B
 For "*I am in the experiment (i.e. I'm either a bookie or in the group)*", write E

First we want to calculate $\Pr(H|G\&E)$ and $\Pr(H|B\&E)$, the probabilities that the groupies and the bookies, respectively, should assign to the proposition that the coin fell heads. Since G implies E , and B implies E , we have $\Pr(H|G\&E) = \Pr(H|G)$ and $\Pr(H|B\&E) = \Pr(H|B)$. We can derive $\Pr(H|G)$ from the following equations:

$$\begin{aligned}\Pr(H|G) &= \Pr(G|H) \Pr(H) / \Pr(G) \text{ (Bayes' theorem)} \\ \Pr(G|H) &= h / (h + b + u) \text{ (Self-Sampling Assumption)} \\ \Pr(G|\neg H) &= t / (t + b + u) \text{ (Self-Sampling Assumption)} \\ \Pr(H) &= \Pr(\neg H) = 1/2 \text{ (Fair coin)} \\ \Pr(G) &= \Pr(G|H) \Pr(H) + \Pr(G|\neg H) \Pr(\neg H) \text{ (Theorem)}\end{aligned}$$

This gives us

$$\Pr(H|G\&E) = \frac{h \cdot (t + b + u)}{h \cdot (t + b + u) + t \cdot (h + b + u)}.$$

In an analogous fashion, using $\Pr(B|H) = b / (h + b + u)$ and $\Pr(B|\neg H) = b / (t + b + u)$, we get

$$\Pr(H|B\&E) = \frac{t + b + u}{(h + b + u) + (t + b + u)}.$$

We see that $\Pr(H|B\&E)$ is not in general equal to $\Pr(H|G\&E)$. The bookies and the people in the group will arrive at different estimates of the probability that the coin was heads. For instance, if we have the parameter values $\{h = 10, t = 1, b = 1, u = 10\}$ we get $\Pr(H|G\&E) \mid 85\%$ and $\Pr(H|B\&E) \mid 36\%$. In the limiting case when the number of outsiders is zero, $\{h = 10, t = 1, b = 1, u = 0\}$, we have $\Pr(H|G\&E) \mid 65\%$ and $\Pr(H|B\&E) \mid 15\%$ (which coincides with the result in section 3). In the opposite limiting case, when the number of outsiders is large, $\{h = 10, t = 1, b = 1, u \infty\}$, we get $\Pr(H|G\&E) \mid 91\%$ and $\Pr(H|B\&E) = 50\%$. In general, we should expect the bookies and the groupies to disagree about the outcome of the coin toss.

Now that we know the probabilities can check whether a bet occurs. There are two types of bet that we will consider. In a type1 bet a bookie bets against the group as a whole, and the group members bet against the set of bookies as a whole. In a type 2 bet an individual bookie bets against an individual group member.

Let's look at the type 1 bet first. The maximal amount \$ that a person in the group is willing to pay to each bookie if the coin fell heads in order to get \$1 from each bookie if it was tails is given by

$$\Pr(H|G)(-x)b + \Pr(\neg H|G)b = 0.$$

When calculating the rational odds for a bookie we have to take into account the fact that depending on the outcome of the coin toss, the bookie will turn out to have betted against a greater or smaller number of group members. Keeping this in mind, we can write down a condition for the minimum amount \$y that a bookie has to receive (from every groupie) if the coin fell heads in order to be willing to pay \$1 (to every groupie) if it fell tails:

$$\Pr(H|G)y \cdot h + \Pr(\neg H|G)(-1)t = 0.$$

Solving these two fairness equations we find that $x = y = \frac{t(h+b+u)}{h(t+b+u)}$, which means that nobody expects to win from a bet of this kind.

Turning now to the type 2 bet, where individual bookies and individuals in the group bet directly against each other, we have to take into account an additional factor. To simplify things, we assume that it is assured that all of the bookies get to make a type 2 bet and that no person in the group bets against more than one bookie. This implies that the number of bookies isn't greater than the smallest number of group members that could have resulted from the coin toss; for otherwise there would be no guarantee that all bookies could bet against a unique group member. But this means that if the coin toss generated more than the smallest possible number of group members then a selection has to be made as to which of the group members get to bet against a bookie. Consequently, a group member who finds that she has been selected obtains reason for thinking that the coin fell in such a way as to maximize the proportion of group members that get selected to bet against a bookie. (The bookies' probabilities remain the same as in the previous example.)

Let's say that it is the tails outcome that produces the smallest group. Let s denote the number of group members that are selected. We require that $s = t$. We want to calculate the probability for the selected people in the group that the coin was heads, i.e. $\Pr(H|G \& E \& S)$. Since S implies both G and E, we have $\Pr(H|G \& E \& S) = \Pr(H|S)$. Using

$$\Pr(H|S) = \Pr(S|H) \Pr(H) / \Pr(S) \text{ (Bayes' theorem)}$$

$$\Pr(S|H) = s / (h + b + u) \text{ (Self-sampling assumption)}$$

$$\Pr(S|\neg H) = s / (t + b + u) \text{ (Self-sampling assumption)}$$

$$\Pr(H) = \Pr(\neg H) = 1/2 \text{ (Fair coin)}$$

$$\Pr(S) = \Pr(S|H)\Pr(H) + \Pr(S|\neg H)\Pr(\neg H) \text{ (Theorem)}$$

we get

$$\Pr(H|G \& E \& S) = \frac{t + b + u}{(t + b + u) + (h + b + u)}.$$

Comparing this to the result in the previous example, we see that $\Pr(H|G \& E \& S) = \Pr(H|B \& E)$. This means that the bookies and the group members that are selected now agree about the odds. So there is no possible bet between them for which both parties would calculate a positive non-zero expected payoff.

We conclude that adopting the self-sampling assumption does not lead observers to place bets against each other. Whatever the number of outsiders, bookies, group members and selected group members, there are no bets, either of type 1 or of type 2, from which all parties should expect to gain.

REFERENCE

Barrow, J. D. & Tipler, F. J.: 1986, *The Anthropic Cosmological Principle*, Clarendon Press, Oxford.

Bostrom, N.: 1999, 'The Doomsday Argument is Alive and Kicking', *Mind* 108, 539-50.

Carter, B.: 1993, 'The anthropic selection principle and the ultra-Darwinian evolution', pp. 38-63 of F. Bertola & U. Curi (eds) *The Anthropic Principle*, Cambridge University Press, Cambridge.

Earman, J.: 1987, 'The SAP also rises: a critical examination of the anthropic principle', *Am. Phil Quart.*, Vol. 24, No. 4, pp. 307-17.

Eckhardt, W.: 1992, 'A Shooting-room view of doomsday', *Journal of Philosophy*, Vol. 44, No. 5, p. 248.

Korb, K. and Oliver, J.: 1999, 'A Refutation of the Doomsday argument', *Mind*, 107, 403-10.

Leslie, J.: 1997, 'Observer-relative Chances and the Doomsday Argument', *Inquiry*, 40, pp. 427-36.

Leslie, J.: 1996, *The End of the World: The Ethics and Science of Human Extinction*, Routledge, London.

Leslie, J.: 1993, 'Doom and Probabilities', *Mind*, Vol. 102, No. 407, pp. 489-91.

Leslie, J.: 1992, 'Doomsday Revisited', *Phil. Quart.* Vol. 42, No. 166, pp. 85-87.

Leslie, J.: 1989, *Universes*, Routledge: London.

NOTES

* I am grateful to John Leslie, Colin Howson and Craig Callender for some very helpful comments on an earlier draft.

[1] John Leslie, 'Observer-relative Chances and the Doomsday Argument'. *Inquiry*, 40 (1997), pp. 427-36. Leslie uses 'chances' as synonymous with 'epistemic probabilities' and I will follow his usage in this paper.

[2] Eckhardt has defined what he calls the human randomness assumption: "We can validly consider our birth rank as generated by random or equiprobable sampling from the collection of all persons who ever live." (*Journal of Philosophy*, vol. XCIV, No. 5, May 1992, p. 248). I think the reference to 'human' is at best misleading and at worst erroneous. The anthropic principle, despite its name, does *not* refer specifically to human beings. (See e.g. Brandon Carter, "The anthropic selection principle and the ultra-Darwinian evolution", p.38, in *The Anthropic Principle*, ed. F. Bertola & U. Curi, 1989, Cambridge Univ. Press.)

[3] See e.g. *The End of the World: The Ethics and Science of Human Extinction*, 1996, Routledge; "Doom and Probabilities", 1993, *Mind*, 102, 407, pp. 489-91; and "Doomsday Revisited", 1992, *Phil. Quart.* 42 (166) pp. 85-87. For a recent critique of the Doomsday argument, see Korb & Oliver, "A refutation of the Doomsday argument", 1999, *Mind* 107, 403-10. For a rebuttal of this critique, see Bostrom, "The Doomsday Argument is Alive and Kicking", 1999, *Mind* 108, 539-50.

[4] The only way, it seems, of maintaining that there are observer-relative chances in a strong, nontrivial sense in Leslie's example is on pain of opening oneself up to systematic exploitation, at least if one is put one's money where one's mouth is. Suppose there is someone who insists that the odds are different for an insider than they are for an outsider, and not only because the insider and the outsider don't know about the same facts. Let's call this hypothetical person *Mr. L*. (John Leslie would not, I hope, take *this* line of defense.)

At the next major philosophy conference that *Mr. L* attends we select a group of one hundred philosophers and divide them into two subgroups which we name by means of a coin toss, just as in Leslie's example. We let *Mr. L* observe this event. Then we ask him what is the probability – for him as an external observer, one not in the selected group – that the large group is the Heads group. Let's say he claims this probability is *p*. We then repeat the experiment, but this time with *Mr. L* as one of the hundred philosophers in the batch. Again we ask him what he thinks the probability is, now from his point of view as an insider, that the large group is the Heads group. (*Mr. L* doesn't know at this point whether he is in the Heads group or the Tails group. If he did, he would know about a fact that the outsiders do not know about, and hence the chances involved would not be observer-relative in any paradoxical sense.) Say he answers *p'*.

If either *p* or *p'* is anything other than 50% then we can make money out of him by repeating the experiment many times with *Mr. L* either in the batch or as an external observer, depending on whether it is *p* or *p'* that differs from 50%. For example, if *p'* is greater than 50%, we repeat the experiment with *Mr. L* in the batch, and we keep offering him the same bet, namely that the Heads group is *not* the larger group, and *Mr. L* will happily bet against us at odds determined by $p^* = (50\% + p') / 2$ (the intermediary odds between what *Mr. L* thinks are fair odds and what we think are fair odds). If, on the other hand, $p' < 50\%$, we bet (at odds determined by *p**) that the Head's group is the larger group. Again *Mr. L* should willingly bet against us.

In the long run (with probability asymptotically approaching one), the Heads group will be the larger group approximately half the time. So we will win approximately half of the bets. Yet it is easy to verify that the odds to which *Mr. L* has agreed are such that this will earn us more money than we need pay out. We will be making a net gain.

It seems indisputable that chances cannot be observer-relative in *this* way. Somebody who thought otherwise would quickly go bankrupt in the suggested game.

[5] At least the form of anthropic reasoning underlying the doomsday argument and various other interesting applications discovered by Carter. There are also other propositions – completely unrelated – that all travel under the oxymoronic name of the Anthropic Principle, such as 'the Strong Anthropic Principle' (some versions of it), the 'Final Anthropic Principle', the 'Participatory Anthropic Principle' etc. Cf. John Earman "The SAP also rises: a critical examination of the anthropic principle (1987), *Am. Phil Quart.*, Vol. 24, 4, pp. 307-17; and Barrow, J. D. & Tipler, F. J. (1986) *The Anthropic Cosmological Principle*, Oxford: Clarendon Press, the introduction; and especially chapter 6 in Leslie, J. (1989) *Universes*, Routledge: London.