Brit. J. Phil. Sci. **51** (2000), 807–835

Varieties of Propensity

Donald Gillies

ABSTRACT

The propensity interpretation of probability was introduced by Popper ([1957]), but has subsequently been developed in different ways by quite a number of philosophers of science. This paper does not attempt a complete survey, but discusses a number of different versions of the theory, thereby giving some idea of the varieties of propensity. Propensity theories are classified into (i) long-run and (ii) single-case. The paper argues for a long-run version of the propensity theory, but this is contrasted with two single-case propensity theories, one due to Miller and the later Popper, and the other to Fetzer. The three approaches are compared by examining how they deal with a key problem for the propensity approach, namely the relationship between propensity and causality and Humphreys' paradox.

1 Introduction

The propensity theory of probability was introduced by Popper ([1957]),[1] and subsequently expounded and developed by him in a series of papers and books ([1959], [1983], [1990]). In his *Logic of Scientific Discovery* ([1934], Ch. VIII, pp. 146–214), Popper advocated a version of the frequency theory. He continued to support an objective interpretation of probability, but subsequent reflection convinced him that the frequency theory was inadequate, and that therefore a new objective interpretation of probability was needed. This he sought to provide with his propensity theory. The main drawback of the frequency theory, according to Popper, was its failure to provide objective probabilities for single events. Yet he thought that these were needed for quantum mechanics.

[1] Popper first presented the propensity theory at a conference in the University of Bristol. However, as he could not attend himself, his paper ([1957]) was read by his then student Paul K. Feyerabend.

Popper's suggestion of a propensity theory of probability was taken up by quite a number of philosophers of science who developed the idea in different ways. As a result there are now several different propensity theories. As Miller puts it ([1994], p. 175):

> One of the principal challenges confronting any objectivist theory of scientific knowledge is to provide a satisfactory understanding of physical probabilities. The earliest ideas here, known collectively as the frequency interpretation of probability, have now been all but abandoned, and have been replaced by an equally diffuse set of proposals all calling themselves the propensity interpretation of probability.

My aim in this paper is not to give a complete account of this 'diffuse set of proposals', but I will describe accounts of propensity due to Popper, Miller and Fetzer, as well as one of my own. This should give a feeling for the varieties of propensity. Naturally I will argue for my own propensity theory, but I hope to show as well some of the pros and cons of different accounts of propensity. The whole situation is quite intricate, and I will approach it historically by describing (in Section **2**) Popper's first version of the propensity theory. Then in Section **3**, I will consider whether this theory of Popper's really does solve the problem of providing objective probabilities for single events. My conclusion will be that it does not, and indeed that objective probabilities of single events may not be necessary at all. At this stage it may look as if I am abandoning the propensity theory altogether, but this is not the case. In Section **4**, I will suggest that we use the term 'propensity theory' not just for Popper's own theory, but for any theory which tries to develop an objective, but non-frequency, interpretation of probability. It seems to me that such an interpretation is needed for reasons that have nothing to do with the question of whether there are objective probabilities of single events. This analysis of propensity leads to a classification of propensity theories. In Section **5**, I consider the propensity theories of Miller and the later Popper, and of Fetzer. This leads to a further refinement of the classification introduced in Section **4**. In Section **6**, I consider how the three kinds of propensity theory which have been introduced cope with one of the main problems confronting the whole approach. This problem concerns the relation between propensity and causality, and involves what is known as 'Humphreys' Paradox'.

2 Popper's introduction of the propensity theory

The problem that gave rise to the propensity theory had already been considered by Popper ([1934]). The question was whether it was possible to introduce probabilities for single events, or *singular probabilities* as Popper called them. Von Mises, assuming of course his frequency theory of probability, had denied that such probabilities could validly be introduced. The example he considered was the probability of death. We can certainly introduce the probability of death before 41 in a sequence of say 40-year-old Englishmen. It is simply the limiting frequency of those in the sequence who die before 41. But can we consider the probability of death before 41 for a particular 40-year old Englishman (Mr Smith, say)? Von Mises answered: 'no!' ([1928], p. 11):

> We can say nothing about the probability of death of an individual even if we know his condition of life and health in detail. The phrase 'probability of death', when it refers to a single person has no meaning at all for us. This is one of the most important consequences of our definition of probability [. . .]

Of course it is easy to introduce singular probabilities on the subjective theory. All Mr Smith's friends could, for example, take bets on his dying before 41, and hence introduce subjective probabilities for this event. Clearly, however, this procedure would not satisfy an objectivist like Popper. The key question for him was whether it was possible to introduce objective probabilities for single events.

Popper ([1934]) disagreed with von Mises' denial of the possibility of objective singular probabilities, partly because he wanted such probabilities for his interpretation of quantum mechanics. Popper therefore considered a single event which was a member of one of von Mises' collectives, and made the simple suggestion that its singular probability might be taken as equal to its probability in the collective as a whole. Popper ([1957], [1959]) presented an objection, which he had himself invented, to this earlier view of his, and this led him to his new theory of probability.

Popper's argument is as follows. Begin by considering two dice: one regular, and the other biased so that the probability of getting a particular face (say the 5) is 1/4. Now consider a sequence consisting almost entirely of throws of the biased die but with one or two throws of the regular die interspersed. Let us take one of these interspersed throws and ask what is the probability of getting a 5 on that throw. According to Popper's earlier suggestion this probability must be 1/4 because the throw is part of a collective for which prob(5) = 1/4. But this is an intuitive paradox, since it is surely much more reasonable to say that prob(5) = 1/6 for any throw of the regular die.

One way out of the difficulty is to modify the concept of collective so that the sequence of throws of the biased die with some throws of the regular die interspersed is not a genuine collective. The problem then disappears. This is just what Popper did ([1959], p. 34):

> All this means that the frequency theorist is forced to introduce a modification of his theory—apparently a very slight one. He will now say

> that an admissible sequence of events (a reference sequence, a 'collective') must always be a sequence of repeated experiments. Or more generally, he will say that admissible sequences must be either virtual or actual sequences which are *characterised by a set of generating conditions*—by a set of conditions whose repeated realisation produces the elements of the sequences.

He then continued a few lines later ([1959], p. 34):

> Yet, if we look more closely at this apparently slight modification, then we find that it amounts to a transition from the frequency interpretation to the propensity interpretation.

In this interpretation, the generating conditions are considered as endowed with a propensity to produce the observed frequencies. As Popper put it ([1959], p. 35):

> But this means that we have to visualise the conditions as endowed with a tendency or disposition, or propensity, to produce sequences whose frequencies are equal to the probabilities; which is precisely what the propensity interpretation asserts.

There is an ambiguity in this formulation. Popper does not make it clear whether, when speaking of sequences, he means infinite sequences or long, but still finite, sequences. One piece of evidence in favour of the former interpretation is that Popper speaks of 'frequencies' being 'equal to the probabilities'. Now limiting frequencies in infinite sequences would be exactly equal to the probabilities, but frequencies in long finite sequences would only be *approximately* equal to the probabilities. There are however two pieces of evidence against the view that Popper had infinite sequences definitely in mind.

First of all, in his exposition of the frequency theory earlier in the same paper, Popper gives what is clearly an ambiguous formulation ([1959], p. 29):

> From the point of view of the frequency interpretation, the probability of an *event of a certain kind*—such as obtaining a six with a particular die—can be *nothing but* the relative frequency of this kind of event in an extremely long (perhaps infinite) sequence of events.

Secondly the formulation of the propensity theory in the 1957 paper seems to favour the finite sequence interpretation. Popper says ([1957], p. 67):

> [. . .] since the probabilities turn out to depend upon the experimental arrangement, they may be looked upon as *properties of this arrangement. They characterize the disposition, or the propensity*, of the experimental arrangement to give rise to certain characteristic frequencies *when the experiment is often repeated.*

Surely only finite sequences can be produced by experiments which are often repeated.

I do not intend to continue with further exegesis of Popper. I introduced the point mainly to stress that throughout what follows I will adopt the *long but finite* sequences interpretation, and correspondingly regard the conditions as having a propensity to produce frequencies which are *approximately* equal to the probabilities. This is because my aim is to make the propensity theory more scientific and empirical, and it is obvious that infinite sequences of repetitions are not to be found in the empirical world. It may be objected to this interpretation that it is very difficult to say when a sequence of repetitions is long, or how close two numbers must become in order to be approximately equal. There is indeed a problem here, but I believe that it can be solved. It would be too much of a digression to consider the matter further here, but I can refer the reader to my earlier discussion ([1973], Chs 4 & 5, pp. 77–118).

Popper's suggestion that probabilities should be related to the outcomes of sets of repeatable conditions (**S**) rather than collectives (**C**) had in fact already been made by Kolmogorov ([1933]). In the section which discusses the relations of his theory to experimental data (Ch. 1, section 2), Kolmogorov says in a footnote ([1933], p. 3):

> In establishing the premises necessary for the applicability of the theory of probability to the world of actual events, the author has used, in large measure, the work of R. v. Mises.

In point of fact, however, Kolmogorov does not follow von Mises in associating probabilities with collectives, but rather associates them with repeatable conditions, as the following quotation shows ([1933], pp. 3–4):

> There is assumed a complex of conditions, **S**, which allows of any number of repetitions [. . .] If the variant of the events which has actually occurred upon realization of conditions **S** belongs to the set A (defined in any way), then we say that the event A has taken place [. . .] Under certain conditions [. . .] we may assume that to an event A which may or may not occur under conditions **S**, is assigned a real number P(A) [. . .]

Kolmogorov did not, however, give any argument for his abandonment of von Mises' concept of collective, and such an argument was supplied by Popper.

There is nevertheless rather more to Popper's notion of propensity than is involved in the change from collectives to conditions. The word 'propensity' suggests some kind of dispositional account, and this marks a difference from the frequency view. A useful way of looking into this matter will be to consider some earlier views of Peirce which were along the same lines.[2] These are contained in the following passage ([1910], pp. 79–80):

[2] On this topic see Fetzer [1993].

> I am, then, to define the meaning of the statement that the *probability*, that if a die be thrown from a dice box it will turn up a number divisible by three, is one-third. The statement means that the die has a certain 'would-be'; and to say that the die has a 'would-be' is to say that it has a property, quite analogous to any *habit* that a man might have. Only the 'would-be' of the die is presumably as much simpler and more definite than the man's habit as the die's homogeneous composition and cubical shape is simpler than the nature of the man's nervous system and soul; and just as it would be necessary, in order to define a man's habit, to describe how it would lead him to behave and upon what sort of occasion—albeit this statement would by no means imply that the habit *consists* in that action—so to define the die's 'would-be' it is necessary to say how it would lead the die to behave on an occasion that would bring out the full consequence of the 'would-be'; and this statement will not of itself imply that the 'would-be' of the die *consists* in such behavior.

Peirce then goes on to describe 'an occasion that would bring out the full consequence of the "would-be" '. Such an occasion is an infinite sequence of throws of the die and the relevant behaviour of the die is that the appropriate relative frequencies fluctuate round the value 1/3, gradually coming closer and closer to this value and eventually converging on it.

Peirce is of course mistaken in speaking of the 'would-be' as a property of the die. Obviously it depends on the conditions under which the die is thrown, as is shown by the following two interesting examples of Popper's. Suppose first we had a coin biased in favour of 'heads'. If we tossed it in a lower gravitational field (say on the Moon), the bias would very likely have less effect and prob(heads) would assume a lower value. This shows an analogy between probability and weight. We normally consider weight loosely as a property of a body whereas in reality it is a relational property of the body with respect to a gravitational field. Thus the weight of a body is different on the Moon whereas its mass (a genuine property of the body) is the same. For the second example we can use an ordinary coin but this time, instead of letting it fall on a flat surface, say on a table top, we allow it to fall on a surface in which a large number of slots have been cut. We now no longer have two outcomes 'heads' and 'tails' but *three*, viz. 'heads', 'tails', and 'edge', the third outcome being that the coin sticks in one of the slots. Further, because 'edge' will have a finite probability, the probability of 'heads' will be reduced. This example shows that not only do the probabilities of outcomes change with the manner of tossing but even that the exact nature of the outcomes can similarly vary.

Despite this error, Peirce has made what seems to me a valuable point in distinguishing between the probability of the die as a dispositional quantity, a 'would-be', on the one hand, and an occasion that would bring out the full consequence of the 'would-be' on the other. The importance of making this distinction is that it allows us to introduce probabilities as 'would-be's' even on occasions where the full consequences of the 'would-be' are not manifested, where in effect we do not have a long sequence of repetitions. On the other hand, if we regard probabilities as 'consisting in such behavior' then it will only make sense to introduce probabilities on 'occasions of full manifestation', i.e. only for long sequences of repetitions. All this will become clearer if we now return to von Mises and Popper.

It is a consequence of von Mises' position that probabilities ought only to be introduced in physical situations where we have an empirical collective, i.e. a long sequence of events. If we adopt Popper's propensity theory, however, it becomes perfectly legitimate to introduce probabilities on a set of conditions *even though these conditions are not repeated a large number of times*. We are allowed to postulate probabilities (and might even obtain testable consequences of such a postulation) when the relevant conditions are only repeated once or twice. Thus Popper's propensity theory provides a valuable extension of the situations to which probability theory applies, as compared to von Mises' frequency view. But does Popper's propensity provide at the same time a solution to the problem of introducing objective probabilities for single events? We shall consider this question in the next section.

3 Can there be objective probabilities of single events?

What is perhaps the major difficulty in the way of introducing objective probabilities for single events was discussed by Ayer ([1963], pp. 188–208), though the problem has an earlier history. The difficulty is this. Suppose we are trying to assign a probability to a particular event, then the probability will vary according to the set of conditions which the event is considered as instantiating—according, in effect, to how we describe the event. But then we are forced to consider the probabilities as attached to the conditions that describe the event rather than to the event itself.

To illustrate this, let us return to our example of the probability of a particular man aged 40 living to be 41. Intuitively the probability will vary depending on whether we regard the individual merely as a man or more particularly as an Englishman; for the life expectancy of Englishmen is higher than that of mankind as a whole. Similarly the probability will alter depending on whether we regard the individual as an Englishman aged 40 or as an Englishman aged 40 who smokes two packets of cigarettes a day, and so on. This does seem to show that probabilities should be considered as dependent on the properties used to describe an event rather than as dependent on the event itself.

It is natural in the context of the propensity theory to consider the problem in terms of the conditions used to describe a particular event, but we could equally well look at the problem as being that of assigning the event to a reference class. Instead of asking whether we should regard Mr Smith as a man aged 40, or as an Englishman aged 40, or as an Englishman aged 40 who smokes two packets of cigarettes a day, we could ask equivalently whether we should assign him to the reference class of all men aged 40, of all Englishmen aged 40, or of all Englishmen aged 40 who smoke two packets of cigarettes a day. The reference class formulation is more natural in the context of the frequency theory where the problem first appeared. Although we are discussing the propensity theory, we will continue to use the traditional terminology, and refer to this fundamental problem as *the reference class problem*.

Howson and Urbach's reaction to the reference class problem is to argue that single case probabilities are subjective rather than objective. However, they also suggest that singular probabilities, though subjective, may be based on objective probabilities. Suppose, for example, that the only relevant information that Mr B has about Mr A is that Mr A is a 40-year-old Englishman. Suppose Mr B has a good estimate (p, say) of the objective probability of 40-year-old Englishmen living to be 41. Then it would be reasonable for Mr B to put his subjective betting quotient on Mr A's living to be 41 equal to p, and thereby making his subjective probability objectively based. This does not, however, turn Mr B's subjective probability into an objective one, for consider Mr C, who knows that Mr A smokes two packets of cigarettes a day, and who also has a good estimate of the objective probability (q, say) of 40-year-old Englishmen who smoke two packets of cigarettes a day living to be 41. Mr C will put his subjective probability on the same event (Mr A living to be 41) at a value q different from Mr B's value p. Once again the probability depends on how the event is categorized rather than on the event itself. Howson and Urbach put the point as follows ([1989], p. 228):

> . . . single-case probabilities . . . are not themselves objective. They are subjective probabilities, which considerations of consistency nevertheless dictate must be set equal to the objective probabilities just when all you know about the single case is that it is an instance of the relevant collective. Now this is in fact all that anybody ever wanted from a theory of single-case probabilities: they were to be equal to objective probabilities in just those conditions. The incoherent doctrine of objective single-case probabilities arose simply because people failed to mark the subtle distinction between the values of a probability being objectively based and the probability itself being an objective probability.

I am inclined to accept this criticism of Howson and Urbach, and so to adopt the following position. We can certainly introduce objective probabilities for events A which are the outcomes of some sets of repeatable conditions **S**. When, however, we want to introduce probabilities for single events, these probabilities, though sometimes objectively based, will nearly always fail to be fully objective because there will in most cases be a doubt about the way we should classify the event, and this will introduce a subjective element into the singular probability. I will now try to elaborate this position, and to discuss some further arguments which can be given in favour of objective singular probabilities. The first of these (the Ali–Holmes example) is due to Robert Northcott.

In 1980 Muhammad Ali, aged 38, fought Larry Holmes for the world heavyweight title. Because Muhammad Ali was a famous and popular figure, the majority of people accepted betting quotients in his favour which were too high, and so the punters made a lot of money by betting in favour of Larry Holmes, who won easily. Does this not indicate that there was an objective probability of Muhammad Ali winning which was much lower than most people thought?

I think that this argument does indeed establish something, but the conclusion is rather weaker than the existence of an objective singular probability. What it does show is that some subjective probabilities (betting quotients) may be preferable to others as a basis for action, but the existence of better subjective probabilities does not establish the existence of a single objective probability. The example is an instance of a general principle that may be roughly stated as follows. On the whole, it is better to use as the basis for action a subjective probability (betting quotient) based on more evidence rather than one based on less evidence. Thus in the Ali–Holmes example, the punters knew a great deal more than the general public about the effects of age on a boxer's performance, on the relative form of Ali and Holmes, etc. So the subjective probability of Ali winning assigned by a punter was likely to have been a better basis for action than one assigned by an ignorant member of the public.

Let us next look at a particular instance of this general principle. Suppose a particular event E can be classified as an instance of a series of conditions **S**, **S**′, **S**″, . . . , where the set of conditions **S** is a subset of **S**′ which is a subset of **S**″ and so on. Suppose further that statistical data enables us to obtain good estimates of the objective probability of E's occurring relative to **S**, **S**′, **S**″, . . . , say p, p′, p″, ... Then common sense suggests that it would, when considering the occurrence of E, be better to adopt as our probability p′ rather than p, p″ rather than p′, and so on. If instead of the conditions **S**, we consider the reference class of the set of instances of **S**, then the principle here could be called the *principle of the narrowest reference class*. It is regarded by Ayer as 'rational to accept'. He states it as follows ([1963], p. 202):

> The rule is that in order to estimate the probability that a particular individual possesses a given property, we are to choose as our class of reference, among those to which the individual belongs, the narrowest class in which the property occurs with an extrapolable frequency.

Again we can illustrate this by our example of the probability of a particular 40-year-old man living to be 41. This individual can be put in the following reference classes: the class of 40-year-old men, the class of 40-year-old Englishmen, the class of 40-year-old Englishmen who smoke two packets of cigarettes a day. Now suppose we have good statistical data for all three classes, then the principle of the narrowest reference class suggests that we should base our probability of the particular individual living to be 41 on the frequency in the third of these three reference classes.

The principle of the narrowest reference class certainly seems to be a sound one, but there are some problems with it. First of all there may not be a single narrowest reference class for which statistics are available.[3] Suppose Mr Smith in addition to smoking two packets of cigarettes a day, plays football once a week. Let us suppose we have statistical data regarding death within a year for the class of 40-year-old Englishman who smoke two packets of cigarettes a day, and for the class of 40-year-old Englishmen who play football once a week, but *not* for the class of 40-year-old Englishman who both smoke two packets of cigarettes a day and play football once a week. We thus have not one but two narrowest reference classes for which statistical data are available and the frequency estimates of the probability of Mr Smith living to be 41 on the bases of these two classes (p'', p''', say) may well be different.

Even if there is a single narrowest reference class, however, there may be, as Keynes pointed out, a danger in its uncritical use. Suppose we adopt the policy of taking as our probability for a single event, the frequency ratio in the narrowest reference class to which that event belongs and for which good statistical data exist, if there is such a class. This policy, according to Keynes, may well lead us astray because we may know things about the event which do not constitute statistical data in a reference class, but which, none the less, give us very good reasons for adjusting our probability. If we neglect such qualitative evidence and use only quantitative evidence, we may often be led to a probability that is a less satisfactory basis for action than might otherwise have been obtained. Keynes puts the point as follows ([1921], p. 322):

> Bernoulli's second axiom, that in reckoning a probability we must take everything into account, is easily forgotten in these cases of statistical probabilities. The statistical result is so attractive in its definiteness that it

[3] This was pointed out to me by David Corfield and Jon Williamson.

> leads us to forget the more vague though more important considerations which may be, in a given particular case, within our knowledge. To a stranger the probability that I shall send a letter to the post unstamped may be derived from the statistics of the Post Office; for me those figures would have but the slightest bearing upon the question.

Keynes obviously considered either that he was more likely than average to post an unstamped letter (perhaps through absent-mindedness or unconscious avarice), or less likely (through being very meticulous in his habits). He does not say which.

We can illustrate Keynes's point with our familiar example as follows. We are trying to assign a probability that our particular individual Mr Smith will live to be 41. Let us suppose that Mr Smith does not, after all, play football once a week and that there is a narrowest reference class for which we have good statistics, namely the class of 40-year-old Englishmen who smoke two packets of cigarettes per day. We accordingly estimate the probability of his living to be 41 as the frequency r say of those in this class who have lived to be 41. Suppose, however, that we learn that Mr Smith comes from a very numerous family who all smoke two packets of cigarettes per day, but none of whom has contracted lung cancer or any other smoking related disease or indeed died before the age of 80. No statistical data are available concerning individuals who belong to such unusual families, but surely, in the light of this extra information, it would be reasonable to change our probability to a value somewhat higher than r.

The general procedure for assigning probabilities to single events then becomes something like the following. We first assign the event to the narrowest reference class for which reliable statistical data exists (if there is such a class), and calculate the relative frequency (r say) of the event's occurring in this class. We then consider any further information of a non-statistical character which is relevant to the event's occurring on this particular occasion, and adjust r either up or down in the light of this information to obtain our probability. If there happen to be several narrowest reference classes with relative frequencies r, r′, r″, . . . , say, we then have to use the non-statistical information to choose a particular r-value as well as to adjust it. If there is no suitable reference class at all, we have to rely exclusively on the non-statistical information to decide on a subjective probability. Such a procedure is surely a reasonable and practical one, but it involves many subjective elements, and is therefore unlikely to produce an objective singular probability in most cases. I will now give one further example of this procedure—*the Francesca argument*.

My wife is from Rome, and her sister has a daughter called Francesca. To explain how Francesca came to formulate the argument, some background on the social customs of Rome is needed. It seems that when a schoolchild reaches the age of 16 in Rome, it becomes necessary, for emotional well-being and maintaining status with the peer group, to own a motor scooter. Naturally, however, this causes great alarm to the parents (and even uncles) who are concerned about the possibility of a road accident. Francesca, when she became 16, was no exception to the general rule. So I had an argument with her on this subject. I pointed out that the frequency of 16-year-old Roman motor scooter riders who had accidents was quite high, and therefore that it might be better not to get a scooter. In her reply, Francesca accepted the truth of the statistics, and even added that two members of her class at school had already been taken to hospital in a coma as a result of motor scooter accidents. One girl had gone on her scooter without wearing a crash helmet to buy a pizza. She was returning balancing the pizza in one hand, and steering the scooter with the other, when the accident occurred. However, Francesca commented that this girl was extremely stupid, and that she (Francesca) would never do a thing like that. She would drive her scooter well and carefully, wear a crash helmet, and take all the other recommended precautions, so that the probability of her having an accident was much lower than average. Although I was trying to support the opposite conclusion, it seemed to me that this argument of Francesca's could not be faulted. Indeed it is a particular instance of Keynes's general principle. To one who knew her well, it did seem likely that she would drive well and carefully, and would therefore be less likely to have an accident than the average 16-year-old Roman. The only criticism that might have been made is that accidents are sometimes the fault of the other party against whose errors even the very best and most careful driving offers no protection. Thus the reduction in the probability of an accident for a good and careful driver below the average level should perhaps not be too great.[4]

I will now consider one final argument in favour of objective probabilities for single events.[5] It might be conceded that it is difficult to assign such singular probabilities in cases like an individual dying before 41, or a 16-year-old Roman having an accident with her motor scooter. However, it could still be claimed that such singular probabilities are more plausible in cases like games of chance, or scientific experiments such as the quantum-mechanical two-slit experiment with electrons. Let us take games of chance first. Certainly in examples such as coin tossing or dice rolling, it does seem quite reasonable to say that on each toss or roll there is an objective singular probability equal to the objective probability in the sequence of tosses or rolls. Our earlier discussion shows why objective singular probabilities are more plausible here than in human cases involving individuals dying before

[4] The reader may be interested to know that Francesca did get her motor scooter, and has ridden it about Rome since the mid-1980s without having an accident.

[5] This argument was suggested to me by Ladislav Kvasz.

41 or having road accidents. In the human case there are many facts about the individual under consideration, which do not take the form of statistical data relating to long sequences, but which seem relevant to assessing the probability. Perhaps there are strong indications that the individual in question has such a character as to make him or her a more careful driver than average, and so on. In the case of standard coin-tossing, however, if fraud and malpractice are excluded, it is part of our background knowledge that additional facts about the toss do not influence the result. Thus it does not matter whether the coin was heads uppermost or tails uppermost before it was tossed, whether it was allowed to fall on the table or on the floor, and so on. Our background knowledge therefore suggests that we should make the singular probabilities of each toss equal, and so equal to the objective probability in the sequence as a whole. Thus we could in this special case introduce objective singular probabilities, but there seems little point in doing so rather than accepting the Howson and Urbach analysis of a subjective probability based on an objective probability. After all we would normally be interested in a particular toss (as opposed to a sequence of tosses) only if we wanted to gamble on the result, and thus the subjective probability analysis in terms of betting quotients seems quite appropriate.

Let us now consider scientific experiments. In my view there is a weaker case for introducing objective singular probabilities here than in the case of games of chance for the following reason. As just observed, it is characteristic of coin-tossing and dice-rolling that they can be carried out in a wide variety of ways without affecting the probability of getting a particular result. Indeed it is difficult, if not impossible, to toss a fair coin in such a way as to favour one side rather than the other.[6] With scientific experiments, however, the situation is quite different. It is often very difficult to perform the experiment correctly without extraneous factors disturbing the result. Great skill and care is needed to ensure that outside influences do not have an effect. Consider, for example, the quantum-mechanical two-slit experiment with electrons. Suppose that two scientists Mr A and Ms B are betting on where an electron will impinge in a particular repetition of the experiment. Mr A sets his probabilities equal to those calculated by the standard theory. Ms B, however, has noticed that there was a thunderstorm nearby, and knows from experience that the resulting electrical disturbances in the atmosphere often affect an experiment of this sort. She therefore adjusts her probabilities in the light of this factor. Once again it seems better to analyse the singular probabilities in a particular repetition of the experiment as subjective

[6] This at least has been my experience of coin-tossing, but David Miller assures me that there is a mechanical coin-tossing apparatus which is guaranteed to produce heads each time. I have not seen such an apparatus to check the claim myself.

probabilities rather than as objective probabilities exactly equal to the objective probabilities in a sequence of repetitions of the experiment.

My general conclusion from the discussion of this section is as follows. It is reasonable in some cases to assign objective probabilities to events A which are the outcomes of sets S of repeatable conditions. Suppose $\mathrm{Prob}(A|S) = p$. Popper claims that there is an objective singular probability p of A occurring on a particular instantiation of the conditions S. We have argued, however, that such a claim is justified, if at all, only in the case of simple games of chance such as coin-tossing or dice-rolling. In all other cases, and perhaps in the case of games of chance as well, it is more reasonable to analyse singular probabilities as subjective probabilities, which may, however, as Howson and Urbach have emphasized, be based at least partly on objective probabilities. As we have seen, Popper's propensity theory was developed in order to permit the introduction of objective singular probabilities. I have argued that it does not succeed in doing so, and it seems therefore as if I have thereby rejected the propensity theory. Certainly if propensity theory is used in a strict sense to describe Popper's precise views, then I do indeed reject the theory.[7] However, since Popper's introduction of the term, 'propensity theory' has come to have a wider significance and to mean roughly 'an objective, but non-frequency, theory'. In the next section I will examine this broader sense of the term 'propensity', and show how it leads to a classification of propensity theories. In the subsequent sections of the paper (**5–6**), the pros and cons of these various approaches to propensity will be considered.

4 Classification of propensity theories

A frequency theory of probability may be characterized as one in which probability is defined in terms of frequency either in the mathematical formalism or in an informal supplement designed to tie the theory in with experience. Now this indicates that frequency theories are based on an operationalist philosophy of science. Operationalism is the view that the theoretical terms of a science should be defined in terms of observables.

[7] This was in fact the position which I adopted in my 1973 book, *An Objective Theory of Probability*. The theory developed in that book was objective and non-frequency, but yet I argued against calling it a propensity theory, partly because it differed in some respects from Popper's theory. Indeed I had at that time some general doubts about the use of the term 'propensity' even for Popper's own views (cf. [1973], pp. 149–50). Subsequently, however, the term 'propensity' became well established in the literature, and has taken on the broader meaning of an objective but non-frequency view of probability. I would therefore now reclassify my earlier position as one particular example of a propensity theory. I had some discussions with Popper on this point after my book had appeared. Interestingly Popper favoured using the term 'propensity' in a general sense rather than as specifically referring to his own views.

Frequency theories of probability are examples of the operationalist approach, because the theoretical term 'probability' is defined in terms of observable frequencies.

Now operationalism was very widely held in the 1920s but subsequently has been much criticized by philosophers of science. The alternative view that has come to prevail is that the theoretical terms of a natural science may often be introduced as undefined primitives, and then connected to experience in a somewhat indirect fashion—not directly through an operational definition. If this more recent view is applied to probability theory, it ties in very nicely with the modern mathematical treatment of probability based on the Kolmogorov axioms. In von Mises' mathematical treatment of probability, probability is explicitly defined in the mathematical formalism as limiting frequency. Kolmogorov abandons this approach, and in his mathematical development takes probability as a primitive undefined term which is characterised axiomatically. Admittedly Kolmogorov's approach is still compatible with a frequency theory, if we take probability as defined in terms of frequency in an informal supplement designed to connect the theory with experience. Indeed Kolmogorov himself seems to adopt a theory of this general character (see his [1933], section 2, pp. 3–5, including fn. 4). Yet although Kolmogorov's mathematics is, in this sense, compatible with the frequency theory of probability, it seems to fit more naturally with recent non-operationalist philosophies of science, in which key theoretical concepts are often taken as undefined, and then connected only somewhat indirectly with observation.

These considerations, which, by the way, have nothing to do with the question of whether there are objective probabilities of single events, suggest that there is a need to develop an objective, but non-frequency, theory of probability. Such a theory would agree with von Mises' view that probability theory is a mathematical science concerned with observable random phenomena. It would also agree with von Mises' view that probability is an objective concept like mass in theoretical mechanics, or charge in electromagnetic theory. It would, however, differ from von Mises' view that probability should be given an operational definition in terms of frequency. Probability would rather be introduced as a primitive undefined term characterized by a set of axioms,[8] and then connected with observation in some manner more indirect than a definition in terms of frequency. My suggestion is that we should use Popper's term 'propensity theory' to describe any objective, but non-frequency, theory of probability having the general character just described.

[8] Obviously this set of axioms would normally be the Kolmogorov axioms, but, as we shall see, there are some versions of the propensity theory, notably Fetzer's, in which propensities do not satisfy Kolmogorov's axioms, but a different set of axioms.

Propensity theories in the above general sense can now be classified into (i) *long-run propensity theories*, and (ii) *single-case propensity theories*.[9] A long-run propensity theory is one in which propensities are associated with repeatable conditions, and are regarded as propensities to produce in a long series of repetitions of these conditions frequencies which are approximately equal to the probabilities. A single-case propensity theory is one in which propensities are regarded as propensities to produce a particular result on a specific occasion. As we have seen, Popper's original propensity theory was, in a sense, *both* long-run *and* single-case. His characterization of propensities corresponds to our long-run propensities, and yet he wanted these propensities to apply to the single-case as well. This position ran into difficulties connected with the reference class problem, and so there has been a tendency for the two halves of Popper's account to separate, producing two different types of propensity theory. My own preference is for a long-run propensity theory, and for dealing with the single-case by subjective probabilities which may, however, be objectively based. But we will next examine more closely the other possibility of sticking to single-case propensities and modifying Popper's original account in other ways.

This analysis explains the fact, pointed out by Runde ([1996]), that Popper's later views on propensity, particularly in his ([1990]), differ considerably from his earlier views. This later position is also developed by Miller ([1994], [1996]). It retains from the earlier Popper objective singular probabilities, but abandons the association of propensities with repeatable conditions. Instead propensities are associated with states of the universe. A single-case propensity theory was developed earlier by Fetzer ([1981]), but it differs significantly from the view of the later Popper and Miller. Instead of associating propensities with the complete state of the world at a given time, he associates them with a complete set of (nomically and/or causally) relevant conditions, which are subject to replication whether or not they are ever replicated. In the next section I will give a fuller account of these single-case propensity theories, and criticize them from the point of view of the long-run propensity theory advocated here.

5 The propensity theories of Miller and the later Popper, and of Fetzer

The main difference between the earlier and later Popper on propensity is that the earlier Popper associates propensities with repeatable conditions, while

[9] The distinction between long run and single case propensity theories is taken from Fetzer ([1988], pp. 123 & 125–6). However, I am using the terminology in a slightly different sense from Fetzer. Fetzer takes the 'long run' to refer to infinite sequences, while, as already explained, I am using 'long-run' to refer to long, but still finite sequences of repetitions.

the later Popper says ([1990], p. 17): '[. . .] propensities in physics are properties of *the whole physical situation* and sometimes of the particular way in which a situation changes.' One reason for this change may have been the desire to preserve objective probabilities for single events. If propensities are associated with repeatable conditions, then, as we argued in detail in Section **3**, it is difficult to carry them over to particular instances of these conditions. At all events Miller is determined to retain objective singular probabilities. He writes ([1994], p. 175): 'The principal virtue of the propensity interpretation in any of its variants is supposed to be that, unlike the frequency theory, it renders comprehensible single-case probabilities as well as probabilities in ensembles and in the long run', and again ([1994], p. 177): '[. . .] the propensity interpretation [. . .] is an objectivist interpretation where single-case probabilities are supreme [. . .]' Naturally I disagree with this point of view since I want to develop a version of the propensity theory in which objective single-case probabilities are abandoned.

In his earlier period, Popper wrote in a passage already quoted ([1959], p. 35):

> But this means that we have to visualise the conditions as endowed with a tendency, or disposition, or propensity, to produce sequences whose frequencies are equal to the probabilities; which is precisely what the propensity interpretation asserts.

As already explained, I would replace 'equal' here with 'approximately equal', but otherwise accept this passage as part of my own version of the propensity theory. Miller, however, criticizes the view that propensities are propensities to produce frequencies. He regards them instead as propensities to realize particular outcomes. As he says ([1994], p. 182):

> In the propensity interpretation, the probability of an outcome is not a measure of any frequency, but (as will be explained) a measure of the inclination of the current state of affairs to realize that outcome.

In a significant passage, Miller relates these changes from the position of the earlier Popper with the need to solve the problem of singular probabilities. As he says ([1994], pp. 18–56):

> It is to be regretted, therefore, that [. . .] we find remarks [in the earlier Popper—D. G.] that [. . .] depict propensities as 'tendencies to produce relative frequencies on repetition of similar conditions or circumstances' [. . .] Propensities are not located in physical things, and not in local situations either. Strictly, every propensity (absolute or conditional) must be referred to the complete situation of the universe (or the light-cone) at the time. Propensities depend on the situation today, not on other situations, however similar. Only in this way do we attain the specificity required to resolve the problem of the single case.

That concludes my account of this version of the propensity theory, and I will now proceed to criticize it.

The main problem with the 1990s views on propensity of Popper and Miller is that they appear to change the propensity theory from a scientific to a metaphysical theory. If propensities are ascribed to a set of repeatable conditions, then by repeating the conditions we can obtain frequencies that can be used to test the propensity assignment. If, on the other hand, propensities are ascribed to the 'complete situation of the universe [. . .] at the time', it is difficult, in view of the unique and unrepeatable character of this situation, to see how such propensity assignments could be tested. Miller seems to agree with this conclusion since he writes ([1996], p. 139):

> The propensity interpretation of probability is inescapably metaphysical, not only because many propensities are postulated that are not open to empirical evaluation [. . .]

Popper, too, writes in similar vein ([1990], p. 17):

> But in many kinds of events [. . .] the propensities cannot be measured because the relevant situation changes and cannot be repeated. This would hold, for example, for the different propensities of some of our evolutionary predecessors to give rise to chimpanzees and to ourselves. Propensities of this kind are, of course, not measurable, since the situation cannot be repeated. It is unique. Nevertheless, there is nothing to prevent us from supposing such propensities exist, and from estimating them speculatively.

Of course we can indeed estimate the propensities speculatively, but if these speculations cannot be tested against data, they are metaphysical in character.

Now there is nothing wrong with developing a metaphysical theory of propensities, and such a theory may be relevant to the discussion of old metaphysical questions such as the problem of determinism. However, my own aim is to develop a propensity theory of probability that can be used to provide an interpretation of the probabilities that appear in such natural sciences as physics and biology. For a theory of this kind, probability assignments should be testable by empirical data, and this makes it desirable that they should be associated with repeatable conditions.

Fetzer's single case propensity theory differs from that of Miller and the later Popper in that he does not associate propensities with the complete state of the universe. As Fetzer says ([1982], p. 195):

> [. . .] it should not be thought that propensities for outcomes [. . .] depend, in general, upon the complete state of the world at a time rather than upon a complete set of (nomically and/or causally) relevant conditions [. . .] which happens to be instantiated in that world at that time.

This seems to me a step in the right direction relative to Miller and the later Popper, but some doubts remain in my mind about how scientifically testable such propensities can be. If propensities are associated with a set of repeatable conditions as in the long-run propensity view, then it is always in principle possible to test a conjectured propensity value by repeating the conditions. If, as Fetzer suggests, we ascribe propensities to a complete set of (nomically and/or causally) relevant conditions, then in order to test a conjectured propensity value we must make a conjecture about the complete list of the conditions that are relevant. This necessary conjecture might often be difficult to formulate and hard to test, thereby rendering the corresponding propensities metaphysical rather than scientific. Once again, then, I have a doubt as to whether single-case propensities give an appropriate analysis of the objective probabilities that appear in the natural sciences.

On the other hand, it should be said in favour of Fetzer's view that, if the problem of finding the complete set of relevant conditions could be solved, his theory would provide an elegant and unified account. If we can define single-case propensities relative to some complete set $\mathbf{S_C}$ say of relevant conditions, we can then extend this to long-run sequences (whether finite or infinite) produced by repetitions of $\mathbf{S_C}$. My own account relates propensities to long-run repetitions of sets of conditions **S** which may not be complete. This makes propensity assignments easily testable, but means, for the reasons explained in Section **2**, that they cannot in general be extended to the single case where subjective probabilities are needed. Thus Fetzer's theory, if it could be made to work, would lead to a unified monistic account, whereas the alternative long-run propensity approach leads necessarily to a more complicated dualism.

We can now extend our classification of propensity theories by subdividing single-case propensity theories into (i) *state of the universe* where the propensity depends on the complete state of the universe at a given time, and (ii) *relevant conditions* where the propensity depends on a complete set of relevant conditions. Miller and the later Popper opt for (i), and Fetzer for (ii). If we add the long-run propensity theory here advocated, we have three different propensity theories. In the next section I will test out these three theories by seeing how well they deal with another major problem connected with propensity. This is the problem of relating propensity to causality, a problem that leads to what is known as Humphreys' paradox.

6 Propensity and causality: Humphreys' paradox

To introduce the difficulty, let me begin by observing that propensities can quite plausibly be considered as generalisations of causes. For example, a massive dose of cyanide will definitely cause death. A suitably small dose of cyanide might only give rise to a propensity of, say, 0.6 of dying. Here propensity appears to be a kind of weakened form of causality. There is, however, a problem with this idea. Causes have a definite direction in time. So if A causes B and A occurs before B, then B does not cause A. Apart from a few speculations in theoretical physics, it is universally conceded that causes do not operate backwards in time. The situation is very different with probabilities. In general if $P(A|B)$ is defined, then so is $P(B|A)$. Probabilities have a symmetry where causes are asymmetrical. It thus seems that propensity cannot after all be a generalization of cause.

This problem was first noticed by Humphreys, and first published by Salmon who gave it a memorable formulation which is worth quoting ([1979], pp. 213–4):

> As Paul W. Humphreys has pointed out in a private communication, there is an important limitation upon identifying propensities with probabilities, for we do not seem to have propensities to match up with 'inverse' probabilities. Given suitable 'direct' probabilities we can, for example, use Bayes's theorem to compute the probability of a particular cause of death. Suppose we are given a set of probabilities from which we can deduce that the probability that a certain person died as a result of being shot through the head is 3/4. It would be strange, under these circumstances, to say that this corpse has a propensity (tendency?) of 3/4 to have had its skull perforated by a bullet. Propensity can, I think, be a useful causal concept in the context of a probabilistic theory of causation, but if it is used in that way, it seems to inherit the temporal asymmetry of the causal relation.

The problem was named 'Humphreys' paradox' by Fetzer ([1981]), and has given rise to much interesting discussion. A statement by Humphreys himself of the paradox is to be found in his ([1985]), which is critically discussed by McCurdy in his ([1996]). There are also important contributions by Fetzer and Miller that will be discussed later on. My aim, however, is not to give a complete review of the literature on the subject, but rather to carry out the following more limited strategy. I will begin by giving what seem to me the two simplest illustrations of the paradox. I will then examine how these cases can be dealt with by the three propensity theories that have been introduced earlier.

The first illustration comes from Milne ([1986]). Let us consider rolling a standard die, let $A = 6$, and $B = \text{even}$. Then, according to standard probability theory, $P(B|A) = 1$, and $P(A|B) = 1/3$. $P(B|A)$ raises no problems. If the result of a particular roll of the die is 6, then that result must be even. B is completely determined by A, which corresponds satisfactorily to a propensity of 1. But now Milne ([1986], p. 130) raises the question of how $P(A|B)$ is to be interpreted by a single-case propensity theory. There is indeed a problem here, for we cannot interpret it as saying that the occurrence of the outcome B partially causes with weight 1/3 outcome A to appear. In fact if outcome B has occurred, then the actual result must have been 2, 4, or 6. In the first two cases it has been determined that 6 will not occur on that roll, while in the third case it has been determined that 6 will definitely occur. In neither case does it make any sense to say that A is partially determined to degree 1/3 by the occurrence of B.

Milne's example concerns two events A and B which occur simultaneously, and so raises difficulties for a causal interpretation since characteristically causes come before their effects. The situation becomes worse in a case where we need to consider $P(A|B)$ and A actually occurs before B. Cases of this sort are a little more complicated than Milne's simple die-rolling example. I have chosen what seems to me the simplest and most elegant. This is the frisbee example (Earman and Salmon [1992], p. 70). It is essentially the same as an example involving electric light bulbs given by Salmon ([1979], p. 214).

Let us suppose then that there are two machines producing frisbees. Machine 1 produces 800 per day with 1% defective. Machine 2, an older and less efficient machine, produces 200 per day with 2% defective. Let us suppose that, at the end of each day, a frisbee is selected at random form the 1,000 produced by the two machines. Let D = the selected frisbee is defective. Let M = it was produced by machine 1, and N = it was produced by machine 2. Let us consider the two conditional probabilities $P(D|M)$ and $P(M|D)$. $P(D|M) = 0.01$, while $P(M|D)$ can be calculated using Bayes theorem as follows.

$$P(M|D) = \frac{P(D|M)P(M)}{P(D|M)P(M) + P(D|N)P(N)}$$

$$= \frac{0.01 \times 0.8}{0.01 \times 0.8 + 0.02 \times 0.2}$$

$$= \frac{8}{12} = \frac{2}{3}$$

As far as the standard operations of the calculus of probabilities are concerned, there is nothing problematic about these two conditional probabilities. But how are they to be interpreted in terms of single-case propensities, if D, M, and N refer to a particular day?

Of course as regards $P(D|M)$, there is no problem. This is just the propensity for machine 1 to produce a defective frisbee. But what of $P(M|D)$? This is the propensity for the actual defective frisbee drawn at the end of a particular day to have been produced by machine 1. If we think of propensities as partial causes, this becomes the following. The drawing of a defective frisbee in the evening is a partial cause of weight 2/3 of its having been produced by machine 1 earlier in the day. Such a concept seems to be nonsense, because by the time the frisbee was selected, it would either definitely have been produced by machine 1 or definitely not have been produced by that machine. We can make the point more vivid by supposing that machine 1 produces blue frisbees, and machine 2 red frisbees. If the defective frisbee drawn at the end of the day was blue, it would definitely have been produced by machine 1, and it is not clear what would be the sense of saying that it had a propensity 2/3 to have been been produced by machine 1. It is clear that examples of this sort pose a problem for the propensity view of probability. Let us next examine how the three propensity theories so far discussed cope with the difficulty.

I will begin with the long-run propensity theory. Here propensities are associated with sets of repeatable conditions. Let **S** be such a set. Let the specific outcomes of **S** be members of a class Ω. Then propensities are assigned to events A, B, . . . which are taken to be subsets of Ω. So, for example, $P(A|\mathbf{S}) = p$ means that there is a propensity if **S** were to be repeated a large number of times for A to appear with a relative frequency approximately equal to p. This view of propensity does not deal in any way with single repetitions of **S**. These are handled using subjective probabilities.

It is obvious from the above brief summary of the long-run propensity theory that basic propensities are conditional, and have the form $P(A|\mathbf{S})$ where **S** is a set of repeatable conditions. Note that here we cannot reverse the order because $P(\mathbf{S}|A)$ does not make sense. Now often for brevity the reference to **S** is not made explicit, and we abbreviate $P(A|\mathbf{S})$ to P(A). Probabilities like P(A) are often called *absolute probabilities*, but really since P(A) is an abbreviation for $P(A|\mathbf{S})$, it would be more accurate to refer to then as *fundamental conditional probabilities*, or *conditional probabilities in the fundamental sense*. Such fundamental conditional probabilities can be contrasted with conditional probabilities of the form $P(A|B)$ where B is not a set of repeatable conditions but an event. It is these conditional probabilities which can be reversed to produce $P(B|A)$. Let us call such conditional probabilities *event conditional probabilities*. This gives rise to two questions. What do such conditional probabilities mean in the given interpretation? And how do they relate to fundamental conditional probabilities? I will now try to answer these questions.

My suggestion is that, just as P(A) should be considered as an abbreviation of $P(A|\mathbf{S})$, so $P(A|B)$ should be considered as an abbreviation of $P(A|\mathbf{B\&S}$ where **B&S** stands for a new set of repeatable conditions defined as follows. We repeat **S** just as before, but only note the result if it is a member of B. Results which do not lie in B are simply ignored. To say that $P(A|\mathbf{B\&S} = q$ means that there is a propensity if this new set of conditions **B&S** is repeated a large number of times for A to appear with frequency approximately equal to q. My next claim is that, with this interpretation of event conditional probabilities, all the conditional probabilities in both the Milne example and the frisbee example make perfect sense and do not raise any problems.

Let us start with Milne's example in which the problem lay in interpreting $P(A|B) = 1/3$, where A = the result of a roll of the die was 6, and B = the result of a roll of the die was even. The meaning of this probability on our long run propensity view is the following. Suppose we roll the die a large number of times, but ignore all odd results. There is a propensity, under these conditions, for 6 to appear with a frequency approximately equal to 1/3. This is both true and straightforward. Note that Milne's difficulties disappear because we are considering the long-run and not a single roll of the die.

The frisbee example is no more problematic than Milne's on this long-run propensity interpretation. Let **S** be the set of repeatable conditions specifying that the two machines produce their daily output of frisbees, and that, in the evening, one of these frisbees is selected at random and examined to see if it is defective. **S** can obviously be repeated each day. $P(M|D)$ is now interpreted as an abbreviation for $P(M|D\&\mathbf{S})$. The statement $P(M|D\&\mathbf{S} = 2/3$ means the following. Suppose we repeat **S** each day, but only note those days in which the frisbee selected is defective, then, relative to these conditions there is a propensity that if they are instantiated a large number of times, M will occur, i.e. the frisbee will have been produced by machine 1, with a frequency approximately equal to 2/3. Note that once again the difficulties disappear because we are considering the long-run rather than a single-case. In a specific instance it did not make sense to speak of the propensity of the selected frisbee having been produced by machine 1 as equal to 2/3. If the selected frisbee was blue, it would have been produced by machine 1. If red, by machine 2. In either case the situation would have already been determined so that a propensity of 2/3 would not make sense. If propensities are propensities to produce long-run frequencies, then a propensity of 2/3 makes perfect sense, even though we know that in each individual case the result has been definitely determined as either M or N by the time the selected frisbee is examined and found to be defective.

So far I have not mentioned the connection between causality and propensity, and the long run propensity theory does seem to sever this connection. But is it wrong to do so? It is after all standard in discussions of causality to distinguish between causes and correlations. My barometer's falling sharply is very well correlated with rain occurring soon, but no one supposes that my barometer's falling sharply is the cause of the rain. Now correlation is a probabilistic notion. So perhaps it is indeed correct that causes are different from probabilities. On the long-run propensity interpretation, there is a high propensity of rain occurring soon, given that my barometer has fallen sharply, but this propensity is not causal in character. This concludes my discussion of Humphreys' paradox in the context of the long-run propensity theory. I will next examine how the paradox might be resolved within the two single-case propensity theories.

Let us begin with the state of the universe version of the single-case propensity theory developed by Miller and the later Popper. In this approach the probability of a particular event A is considered as conditional on an earlier state of the universe U_t, say. $P(A|U_t) = p$ means that the state of the universe U_t has a propensity p to produce the event A. Here propensity is definitely thought of as a generalized cause. It will be sufficient to examine how this account applies in the frisbee case, since Milne's example does not raise any additional problems.

To deal adequately with the frisbee example within this propensity theory, it is important to add time subscripts to all the events involved. Let U_t be the state of the universe at the beginning of a particular day. Let D_v be the event that the frisbee drawn in the evening at time v was defective. Let M_u be the event that this defective frisbee was produced by machine 1 at time u during the day. Obviously we have $t < u < v$. The problem now is how to interpret the two conditional probabilities $P(D_v|M_u)$ and $P(M_u|D_v)$. As in the previous case, all probabilities are conditional. If we write P(A), this can only be an abbreviation for $P(A|U_t)$ where U_t is a state of the universe. So the fundamental conditional probabilities in this theory have the form $P(A|U_t)$. In the case of $P(D_v|M_u)$ and $P(M_u|D_v)$, however, neither M_u nor D_v are states of the universe but particular events. We are dealing with event conditional probabilities, and once again we have to examine what sense can be made of these in the present theory.

Let us first try interpreting $P(D_v|M_u)$ as $P(D_v|M_u \& U_t)$ by analogy with what we did last time. The problem is that $M_u \& U_t$ is not a state of the universe at a particular time. Suppose we wanted to turn it into a state of the universe at the later time u (U_u, say), then we would need to specify not just one event occurring at u such as M_u but all the other events occurring at u. Worse still, M_u might remain the same while these other events were different producing different values for U_u, and hence different values for $P(D_v|U_u)$. For example, in one U_u there might be severe oscillations in the supply of electricity between t and u increasing the proportion of defective frisbees produced by the two machines. In another U_u there might be no such disruption. The values of $P(D_v|U_u)$ would be different in the two cases.

There thus seems no hope of interpreting $P(D_v|M_u)$ along the lines we used in the previous case. The situation is even worse for $P(M_u|D_v)$, since if we extended $D_v \& U$ to U_v, then M_u would occur at a time earlier than v which disallows $P(M_u|U_v)$. As far as I can see, there is only one way of introducing event conditional probabilities in this theory, and that is by defining them formally thus.

$$\mathrm{P}(\mathrm{A}|\mathbf{B};\mathbf{U_t}) =_{\mathbf{def}} \frac{\mathbf{P(A\&B|U_t)}}{\mathbf{P(B|U_t)}} \qquad \textbf{for } \mathbf{P(B|U_t)?\ 0}$$

But these formal event conditional probabilities do not share the important properties of the fundamental conditional probabilities ($\mathrm{P}(\mathrm{A}|\mathrm{U_t})$) which underlie this version of the propensity theory. In particular '$\mathrm{P}(\mathrm{A}|\mathrm{B}) = \mathrm{p}$' does not imply that there is any link of a causal character between B and A. Thus Humphreys' paradox on this approach is again solved by denying that event conditional probabilities involve any kind of causal link, though, in contrast to the previous theory, it is maintained at the same time that conditional probabilities in the fundamental sense do involve a sort of causal link.

What I have just given is my own analysis of the situation, but it seems to agree quite well with what Miller says in the following passage ([1994], p. 189):

> [. . .] if *a* is my survival one year from today, and *c* is my taking up parachuting tomorrow [. . .] the causal influence that is measured by *$p(a|c)$ is an influence from today to a day one year hence [. . .] It is not an influence from the time recorded in c to the time recorded in a [. . .] What about the inverse conditional probability $p(c|a)$? This comes out as the propensity for today's world to develop into a world in which I take up parachuting tomorrow, given that it—today's world—will by the end of the year have developed into one of the worlds in which I am still alive [. . .] The causal pressure is from today to tomorrow, not from the remote future to tomorrow.*

I have some doubts about the ordinary language equivalent given here for *$p(c|a)$, but the point which agrees with my own analysis is that in $p(c|a)$ it is denied that there is causal pressure from a to c, and in $p(a|c)$ it is denied that there is influence from c to a. In effect it is denied that event conditional probabilities involve any causal type influences between the events. The influence is claimed to be from today to a day one year hence, or from today to tomorrow. In other words the causal type pressure runs from the state of the universe today to events lying in the future. It does not connect the future events that are involved in the event conditional probability.*

Let us finally examine Humphreys' paradox within the context of the relevant conditions single-case propensity theory due to Fetzer. As in the previous case, we will focus on the frisbee example using the time subscripts already introduced. As in the previous two cases, we can distinguish between conditional probabilities in the fundamental sense, and event conditional probabilities. Within this theory, fundamental conditional probabilities have the form $\mathrm{P}(\mathrm{A}|\mathrm{R_t})$ where $\mathrm{R_t}$ is a complete set of (nomically and/or causally) relevant conditions instantiated in the world at time t. $\mathrm{R_t}$ consists of all those conditions that are relevant to the occurrence of the events under consideration, but does not amount to a complete state of the universe at t. This is where Fetzer's version of the single-case propensity theory differs from that of Miller and the later Popper. $P(A|R_t) = p$ means that there is a propensity of degree p for the conditions R_t to produce the event A at some time later than t. Here, as in the previous case, propensity is thought of as a generalized cause.

Let us as before raise the question of how the two event conditional probabilities $P(D_v|M_u)$ and $P(M_u|D_v)$ are to be interpreted in this theory. In fact $P(D_v|M_u)$ can be interpreted quite straightforwardly as $P(D_v|M_u\&R_t)$. The set of relevant conditions at t is R_t. At the later time u, however, the frisbee which will eventually be selected is produced by machine 1. This is the event M_u. The occurrence of this event is part of the set of relevant conditions for D_v at u. Thus we have $P(D_v|M_u) = P(D_v|M_u\&R_t) = P(A|R_u)$. So this event conditional probability can be interpreted as a fundamental conditional probability at a different time, and therefore has the causal influence of a fundamental conditional probability. The inverse event conditional probability $P(M_u|D_v)$ cannot however be interpreted in this way. If we tried to extend $D_v\&R_t$ to R_v, then $P(M_u|R_v)$ would no longer make sense as a generalized cause propensity since v is later than u and the direction of causality is wrong. Of course such event conditional probabilities could still be introduced in a formal sense by a definition analogous to the one given earlier.

Humphreys' paradox is thus resolved within Fetzer's theory by saying that some, but not all, event conditional probabilities are propensities (in the sense of generalised causes). This is how Fetzer puts it ([1982], p. 195):

> [. . .] by virtue of their 'causal directedness', propensities cannot be properly formalized either as 'absolute' or as 'conditional' probabilities satisfying inverse as well as direct probability relations.

and again ([1991], pp. 297–8):

> [. . .] that propensities are not probabilities (in the sense of satisfying standard axioms, such as Bayes's theorem) by virtue of their causal directedness was not generally recognized before the publication of Humphreys (1985).

On Fetzer's account then propensities do not satisfy the standard Kolmogorov axioms. Working with Nute, however, Fetzer developed an alternative set of axioms for propensities. This system which he calls 'a probabilistic causal calculus' is presented in his ([1981], pp. 59–67). It has the feature that (Fetzer [1981], p. 284): '[. . .] *p* may bring about *q* with the strength *n* (where *p* occurs prior to or simultaneous with *q*), whether or not *q* brings about *p* with *any* strength *m*.'

Fetzer's position might seem to be that propensities are not probabilities, but he objects to this formulation on the grounds that the Fetzer–Nute probabilistic causal calculus has many axioms which are definitely probabilistic in character. It might therefore be more accurate to describe the Fetzer–Nute calculus as a non-standard probability theory. As Fetzer himself says ([1981], p. 285):

> Perhaps this means that the propensity construction has to be classified as a *non-standard* conception of probability, which does not preclude its importance *even as an interpretation of probability*! Non-Euclidean geometry first emerged as a *non-standard* conception of geometry, but its significance is none the less for that. Perhaps, therefore, the propensity construction of probability stands to standard accounts of probability just as non-Euclidean constructions of geometry stood to standard accounts of geometry before the advent of special and of general relativity.

Since 'non-standard' has connotations of 'non-standard analysis' it might be better to speak of the probabilistic causal calculus as a non-Kolmogorovian probability theory by analogy with non-Euclidean geometry.

The Fetzer–Nute suggestion of a non-Kolmogorovian probability theory is a bold and revolutionary one, but its revolutionary character will naturally create problems in its achieving general acceptance. There is an enormous body of theorems based on the Kolmogorov axioms. The mathematical community is unlikely to give up this formidable structure and substitute another for it unless there are very considerable gains in so doing. This is one reason for preferring a propensity theory (such as the long-run propensity theory described in this paper) which retains the standard Kolmogorov axioms.

This concludes my account of how the three propensity theories deal with Humphreys' paradox. What is interesting is that all three propensity theories do solve the paradox, but in quite different ways. Naturally I prefer the solution offered by the long-run propensity theory, but the situation is quite intricate and there are positive and negative features of each solution.

Acknowledgements

My main thanks are due to David Miller and Jim Fetzer with whom I had long discussions about the questions dealt with in the paper. My interest in these questions was reawakened by reading David Miller's ([1994]), and by a memorable dinner with him and Jim Fetzer when the latter was visiting London. Discussions continued at subsequent meetings and via Email. Earlier versions of the paper were read at Warwick University in 1997, and at the annual conference of the British Society for the Philosophy of Science at Nottingham in 1999. On the latter occasion the paper was part of a session with Peter Clark whose comments on propensity I found most helpful and interesting. Versions of the paper were also read by Max Albert, David Corfield, Colin Howson, Jochen Runde, and Jon Williamson, who all made helpful comments.

Department of Philosophy,
King's College London,
Strand,
London WC2R 2LS, UK
donald.gillies@kcl.ac.uk

References

Ayer, A. J. [1963]: 'Two Notes on Probability', in *The Concept of a Person and Other Essays*, London: Macmillan, pp. 188–208.

Earman, J. and Salmon, W. C. [1992]: 'The Confirmation of Scientific Hypotheses', in M. H. Salmon *et al.*, *Introduction to Philosophy of Science*, Englewood Cliffs, NJ: Prentice Hall, Ch. 2, pp. 42–103.

Fetzer, J. H. [1981]: *Scientific Knowledge: Causation, Explanation, and Corroboration*, *Boston Studies in the Philosophy of Science*, *Vol. 69*, Dordrecht: D. Reidel.

Fetzer, J. H. [1982]: 'Probabilistic Explanations', *PSA, Vol. 2*, pp. 194–207.

Fetzer, J. H. [1988]: 'Probabilistic Metaphysics', in J. H. Fetzer (*ed.*), *Probability and Causality*, Dordrecht: D. Reidel, pp. 109–32.

Fetzer, J. H. [1991]: 'Critical Notice: Philip Kitcher and Wesley C. Salmon (*eds*), *Scientific Explanation*; and Wesley C. Salmon, *Four Decades of Scientific Explanation*', *Philosophy of Science*, **58**, pp. 288–306.

Fetzer, J. H. [1993]: 'Peirce and Propensities', in Edward C. Moore (*ed.*), *Charles S. Peirce and the Philosophy of Science*, Tuscaloosa and London: University of Alabama Press, pp. 60–71.

Gillies, D. A. [1973]: *An Objective Theory of Probability*, London: Methuen.

Howson, C. and Urbach, P. [1989]: *Scientific Reasoning: The Bayesian Approach*, Lasalle, IL: Open Court.

Humphreys, P. [1985]: 'Why Propensities Cannot Be Probabilities', *Philosophical Review*, **94**, pp. 557–70.

Keynes, J. M. [1921]: *A Treatise on Probability*, London: Macmillan [1963].

Kolmogorov, A. N. [1933]: *Foundations of the Theory of Probability*, second English edn., New York: Chelsea [1956].

McCurdy, C. S. I. [1996]: 'Humphreys's Paradox and the Interpretation of Inverse Conditional Propensities', *Synthese*, **108**, pp. 105–25.

Miller, D. W. [1994]: *Critical Rationalism: A Restatement and Defence*, Chicago and Lasalle, IL: Open Court.

Miller, D. W. [1996]: 'Propensities and Indeterminism', in A. O'Hear (*ed.*), *Karl Popper: Philosophy and Problems*, Cambridge: Cambridge University Press, pp. 121–47.

Milne, P. [1986]: 'Can There Be a Realist Single-Case Interpretation of Probability?', *Erkenntnis*, **25**, pp. 129–32.

Peirce, C. S. [1910]: 'Notes on the Doctrine of Chances', reprinted in *Essays in the Philosophy of Science*, *The American Heritage Series*, Indianapolis and New York: Bobbs–Merrill [1957], pp. 74–84.

Popper, K. R. [1934]: *The Logic of Scientific Discovery*, sixth rev. impression of the 1959 English translation, London: Hutchinson [1972].

Popper, K. R. [1957]: 'The Propensity Interpretation of the Calculus of Probability, and the Quantum Theory', in S. Körner (*ed.*), *Observation and Interpretation, Proceedings of the Ninth Symposium of the Colston Research Society*, University of Bristol, pp. 65–70, 88–9.

Popper, K. R. [1959]: 'The Propensity Interpretation of Probability', *British Journal for the Philosophy of Science*, **10**, pp. 25–42.

Popper, K. R. [1983]: *Realism and the Aim of Science*, London: Hutchinson.

Popper, K. R. [1990]: *A World of Propensities*, Bristol: Thoemmes.

Runde, J. [1996]: 'On Popper, Probabilities and Propensities', *Review of Social Economy*, **LIV**, **4**, pp. 465–85.

Salmon, W. C. [1979]: 'Propensities: a Discussion Review', *Erkenntnis*, **14**, pp. 183–216.

von Mises, R. [1928]: *Probability, Statistics and Truth*, 2nd rev. English edn., London: George Allen & Unwin [1961].