ON THE PRINCIPLE OF INDIFFERENCE

A DEFENCE OF THE CLASSICAL THEORY OF PROBABILITY

Draft: 29/09/2012

The classical theory of probability has long been abandoned and is seen by most philosophers as a non-contender—a mere precursor to newer and better theories. In this paper I argue that this is a mistake. The main reasons for its rejection—all related to the notorious principle of indifference—are that it is circular, of limited applicability, inconsistent, and dependent upon unjustified empirical assumptions. I argue that none of these claims is true and that the classical theory remains to be refuted.

The classical theory of probability states that the probability of an outcome is equal to the proportion of favourable outcomes among the total number of equally possible outcomes. For example, the probability of drawing a spade, of which there are 13, from an ordinary deck of 52 cards is on this view 13/52, or 1/4. The classical theory is a *logical* theory, according to which probabilities can be deduced from the statement of the problem at hand. This seems to me to be fundamentally correct. If I take one of the 52 cards from the deck I know that I hold either the two of hearts, the three of hearts, the four of hearts...and so on for all 52 cards. The fact that I hold one of these cards entails—to some degree—that I hold each individual card. We call the degree to which a particular outcome is entailed the *probability* of that outcome.

Or at least this is what I claim. The classical theory is widely rejected and has been for some time. In this paper I will argue that this is unjustified.

The biggest difficulty in assigning probabilities according to the classical method is in saying what the "equally possible" outcomes are. For instance, why not say that the equal alternatives in our above example were *spades* and *non-spades*? On this characterisation the probability of drawing a spade would be 1/2 as there are two equally possible outcomes (*spades* and *non-spades*) and one of these is favourable (i.e. *spades*). Not only does this contradict our earlier answer of 1/4 but it quickly leads to paradox. If we accept that the probability of *spades* is 1/2, then by the same reasoning it seems we should accept that the probabilities of *clubs*, and *diamonds*, and *hearts* are all 1/2 also. This is a variant of the *book paradox*.[1]One might also wonder whether "equally possible" is just a covert was of saying "equally probable," making the classical theory circular. To address these difficulties we can appeal to the *principle of indifference* (or *principle of insufficient reason*). John Maynard Keynes (1921, p. 42), who coined the former term, states the principle (which he goes on to defend) as follows:

> If there is no *known* reason for predicating of our subject one rather than another of several alternatives, then relatively to such knowledge the assertions of each of these alternatives have an *equal* probability. Thus *equal* probabilities must be assigned to each of several arguments, if there is an absence of positive ground for assigning *unequal* ones.

According to Keynes, the principle of indifference cannot properly be applied to *spades* and *non-spades* because there *is* known reason for assigning unequal probabilities to the two.

[1] In the original version the question is about the colour of a book (e.g. Keynes, 1921, p. 43) and the alternatives are *red* and *non-red, blue* and *non-blue,* and so on.

Specifically, we know that *non-spades* can be further divided into the possibilities, *hearts*, *clubs*, and *diamonds,* without there being corresponding subdivisions of *spades* of the same form (Keynes, 1921, p. 59). Importantly for Keynes, the meanings of the terms count as evidence too.

Note also that the principle of indifference addresses the worry of circularity in our formulation of the classical theory by defining equal possibility in terms of equal evidence (Háyek, 2011). According to the principle, if there is equal evidence for several possibilities—that is, they are "equally possible"—then they are equally probable. The move from the former to the latter does not therefore rely on a slide from one term to the other (at least not obviously).

That being said, it is fair to say that the biggest problems facing the classical theory have something or other to do with that the notion of equal possibility. For this reason much of the criticism of the view has focused on the principle of indifference. It is these that will concern us here. The objections which we will discuss can be presented as separate difficulties for the principle—each of which is damaging on its own—but they are not unrelated. Taken together they attempt to show that the classical theory is either *trivial*, in that it is circular or only applicable to a small range of rather artificial cases; or *unjustifiable*, in that it leads to paradox or relies upon tacit empirical assumptions which belie its logical foundations. We have already seen an example of the latter in the book paradox. In Section I we will discuss an example of the former: the objection that the classical theory cannot deal with biased chance equipment. In Section II we will deal with another difficulty: that the notion of form used to solve the book paradox makes learning from experience impossible in situations where it shouldn't be. This problem is compounded by the fact that it seems arbitrary and ad hoc to choose one of many alternative applications of the principle of indifference which do allow for learning from experience. Finally, in Section III we will discuss Bertrand's paradox, a problem which is thought by many to be fatal to any account of logical probability. In that section we will also address the worry of circularity mentioned above in more detail.

Ultimately, these objections can all be answered if we are careful to take into account all the details of the case at hand. The classical theory can be applied to biased chance equipment; it can be applied consistently and in a principled manner; and it can be applied usefully in the absence of tacit empirical assumptions.

I

Setting aside the book paradox, opponents of the classical theory are usually happy to admit that it has useful application to what we might call "classical cases" such as fair coins and dice, ordinary decks of cards, and so on. However, the theory struggles in more complicated cases where it is not clear what the alternatives are to which the principle of indifference should be applied. In the book paradox it is easy to see that *non-spades* is divisible into the alternatives: *hearts*, *clubs*, and *diamonds*. A solution is not so readily apparent for loaded dice, biased coins, and rigged decks of cards. Compare a fair coin to a coin which is biased towards landing heads. It seems that in both cases there are only two possibilities—*heads* or *tails*—yet in the latter case we want to say that *heads* is more probable. What we want to say that *heads* and *tails* are not equally possible in this case but, unlike with the book paradox, it is not clear why this is so. What are the subdivisions of *heads* that make it more probable than *tails* in this case? Most commentators take this to be a serious problem for the classical theory. Antony Eagle (2011, p. 283) even goes so far as to say:

> There are clearly cases of non-uniform chances, like biased coins and weighted dice (and their quantum analogues), where the space of possibilities is the same as in the uniform case and yet the probabilities differ. Since there can be a difference in chances with no difference in the space of possibilities, chance cannot supervene on possibility in the way envisaged by the classical theory.[2]

But this is false.[3] Why should we think that the space of possibilities is the same in these cases? Although a biased coin and a fair coin both have two outcomes, *heads* and *tails*, the way the coin ends up is only *part* of what happens. Once we recognise that the tosses of the coin are part of the chance setup too we can see how there are more ways a biased coin can land heads.

Consider tosses of a perfectly symmetrical coin. Remembering that our account of probability is a logical one, we must ask what states of affairs are described by this statement. What actions, for instance, count as "tosses"? This is a complex matter so let us simplify things and suppose that we have a coin-tossing machine which launches coins of a particular size and weight by applying upward force to one side of the coin (ensuring that it flips). The machine is capable of launching coins at ten different velocities, each time applying force to the same part of the coin. Accounting for the fact that the machine can be "loaded" with a coin facing either heads-up or tails-up, we end up with a total of twenty possible tosses, or ten

[2] In this passage Eagle is talking specifically about *chance*, which he distinguishes from *evidential probability*. He goes on to say that the classical theory does better as a theory of the latter, but his argument applies equally to evidential probability. It doesn't really matter for our purposes which we take the classical theory to be an account of.

[3] Except perhaps for the "quantum analogues." However, it is not clear that these are even "probabilities" in the same sense of the word.

pairs of “counterpart” tosses.[4] Now, if we load the coin into the machine what are the possible outcomes? Because we are considering a coin which is perfectly symmetrical it follows (on the assumption that the laws of physics are constant) that for every possible toss that lands heads there is a possible toss that lands tails—namely, its counterpart toss.[5] There are therefore ten possibilities of heads and ten possibilities of tails.

On the other hand, a coin which is biased towards landing heads will, by definition, probably land heads more often than it lands tails. Given what we know about the relevant physics (and ruling out magnets and other outside influences) this must be because it is asymmetrical in such a way that on some of the possible tosses on which a symmetrical coin lands tails it lands heads. (Assume that everything else is held constant: starting position of the coin, initial velocity and spin, etc.). In other words, there must be at least one pair of counterpart tosses where both tosses land heads (and no corresponding pair where both tosses land tails). Hence there must be more possible tosses of this biased coin which land heads than there are tosses which land tails. We can now see that (in this case) the proposition *The coin lands heads* stands for more states of affairs than the proposition *The coin lands tails*. Eagle’s claim that the space of possibilities is the same for uniform and non-uniform cases is clearly false. It is true that in each case there are ten velocities at which the coin can be launched but the twenty possible tosses that result are quite different.

It is also clear that we can calculate the probability for the biased coin if we know enough about the possible coin tosses. We cannot apply the principle of indifference to *heads*

[4] In fact there are many more if we account for the precise ways in which the coin might land. For example, draw a mark near the edge of one side of the coin: the coin could land with this mark facing north, east, south, west, or anywhere in between. However, it is easy enough to see that each of the twenty tosses will have the same number of possible variations associated with it so we can safely ignore this.

[5] To see why this is so consider one of the tosses. Let us suppose that on this toss the coin is loaded heads-up and lands heads. The counterpart toss of begins (by definition) with the coin loaded tails-up. However, the velocity of the toss remains the same, and the coin will follow the same trajectory through the air. This means that if it landed heads before it will land tails now. The same is true for all ten pairs of counterpart tosses.

and *tails* because they are not of the same form, but we can apply it to the twenty possible tosses, as we have been given no reason to think that some of these are more likely than others. Each has a probability of 1/20 and the probability that the coin lands heads is simply the proportion of possible tosses that have as their outcome heads.

It appears then that the principle of indifference can be applied consistently to these cases after all if we are careful to attend to the details given. Too much has been made of this objection to the classical theory. It was one reason which led philosophers to reject the classical theory in favour of other theories which were thought to be better equipped to deal with this type of case. But in fact the classical theory provides a *better* account of biased coins and loaded dice than many of these theories because it provides an explanation for the biased behaviour. That explanation—that there are more possible outcomes of one type than another for a biased device—is not available to theories which simply allow the assignment of unequal probabilities.

II

We have seen how important it is for the classical theorist to be able to distinguish equal possibilities from unequal possibilities. Our way of doing so in the case of the book paradox was to maintain that the principle of indifference should be applied only to propositions of the same form. Unfortunately, this response runs into difficulties when we try to account for learning from experience. We will see in this section that application of the principle to propositions of the same form and learning from experience appear to be incompatible. The problem is made worse by the fact that alternative ways of applying the principle which do allow for learning from experience look ad hoc and arbitrary. Neither choice is acceptable for the classical theorist.

A central disagreement amongst proponents of logical probability up until the mid twentieth century was about whether the principle of indifference should be applied to *state descriptions* or to *structure descriptions*.[6] Suppose we have an urn containing some (unknown) distribution of black (B) and white (W) balls. If we draw two balls from the urn, they could either both be black (2B), one could be black and the other white (1B), or none could be black (0B)—that is both could be white. These three possible proportions of black balls—2B, 1B, 0B—are the structure descriptions for our case. On the other hand, we could describe the possibilities according to the order in which the balls are drawn. This gives us the four descriptions, BB, BW, WB, and WW, which take into account the colour and order of the balls drawn. These are said to be the state descriptions for our case. ("BW" denotes a scenario where a black ball is drawn first and a white ball second, whereas "WB" describes an outcome where the reverse order is observed). By definition, the state descriptions for a particular case are the maximally specific (given our interests) descriptions of how things could be.

Obviously, structure descriptions and state descriptions are just different ways of describing the same outcomes. The problem is that applying the principle of indifference to each gives us a different result. If we choose to apply it to the state descriptions the space of possibilities or *sample space* is {BB, BW, WB, WW}, and we find that probability of drawing two black balls (BB) is 1/4. On the other hand, if we apply the principle to the structure descriptions our sample space is {2B, 1B, 0B} and the probability of selecting two black balls (2B) is 1/3.

[6] See Carnap (1955) for more detailed discussion of this distinction between state and structure descriptions (complete with a helpful graphical representation). He calls the former *individual distributions* and the latter *statistical distributions*. His case involves an urn with four balls instead of two but the point is the same.

Which is the correct method? Keynes (1921, p. 57) argued in favour of assigning equal probabilities to state descriptions.[7] Specifically, he argued that the structure descriptions are not of the same form in the same way that *spades* and *non-spades* are not. Specifically, the structure description 1B stands for two outcomes, BW and WB, whereas 2B and 0B each stand for only one (BB and WW respectively). In opposition to this Carnap (1955) argued that the initial probabilities should be divided evenly between structure descriptions, giving a probability of 1/3 for drawing two black balls (2B) from the urn (instead of the 1/4 given by Keynes' method). To find the probability of BW one then applies the principle of indifference to the state descriptions: in this case 1/3 * 1/2 = 1/6. He points out that any good theory of "inductive thinking" should allow for learning from experience. For example, suppose that we have drawn 99 black balls from the urn. Surely we want to say that the probability of drawing a black ball on the 100th draw is higher than the probability of drawing a white ball. Carnap argued that Keynes' method—unlike his own—does not meet this requirement and is therefore inadequate. Recall that on Keynes' view the equally possible outcomes are BB, BW, WB, and WW. As there are four of them the probability of each is 1/4. This is show in Table 1 below.[8]

[7] Keynes calls the view that the principle of indifference should be applied to state descriptions the *constitution hypothesis*. He calls the structure description view the *ratio hypothesis*.

[8] The table is adapted from Carnap (1955) in Eagle (2011, p. 322) as is Table 2 below.

Table 1

Draw 1	Draw 2	
B	B	Pr = 1/4
B	W	Pr = 1/4
W	B	Pr = 1/4
W	W	Pr = 1/4

We can use the table to find the probabilities of various outcomes. The probability of selecting a black ball (B) on the second draw—which we shall denote "Pr(B_2)"—is found by calculating the proportion of "B"s in the second column. Half of the possible draws are black so Pr(B_2) = 1/2. We call this the *prior* probability. What if we have already drawn a black ball? Does this raise the chance of drawing another one? Here we are after the probability of drawing a black ball on the second try given that we have already drawn one. This is called the *conditional* or *posterior* probability. We can write it as "Pr($B_2|B_1$)". There are two outcomes that have a "B" under "Draw 1" (i.e. BB and BW); out of these 1/2 also have a "B" under "Draw 2" (i.e. BB). This tells us that Pr($B_2|B_1$) = 1/2, which is the same as Pr(B_2); so no learning from experience can occur. Knowing the outcome of the first draw doesn't change our knowledge of the second. Even if we had chosen an example in which 99 black balls had been drawn, Keynes' method would have yielded the same probability: 1/2.

On the other hand, applying the principle of indifference to the structure descriptions, 2B, 1B, and 0B, as Carnap recommends can accommodate learning from experience. On his view each of these has a probability of 1/3, with 1B equally divisible into BW and WB. Thus BW and WB both have probability $1/3 * 1/2 = 1/6$ and we have:

Table 2

Draw 1	Draw 2	
B	B	Pr = 1/3
B	W	Pr = 1/6
W	B	Pr = 1/6
W	W	Pr = 1/3

Transforming this so that we have a uniform distribution of probabilities to which we can apply the classical method (and which is more easily comparable to Table 1) we get:

Table 3

Draw 1	Draw 2	
B	B	Pr = 1/6
B	B	Pr = 1/6
B	W	Pr = 1/6
W	B	Pr = 1/6
W	W	Pr = 1/6
W	W	Pr = 1/6

The prior probability of getting a black ball on the second draw is $Pr(B_2) = 1/2$: the same as before, and as one would expect given no knowledge of the ratio of black to white balls in the urn. What of the probability of drawing a second black ball? Looking at those rows which contain a “B” in the first column, we find that 2/3 of these also contain a “B” in the second column. Thus $Pr(B_2|B_1) = 2/3$, and we have learning from experience![9][10]

These results cast considerable doubt on Keynes’ solution, and therefore pose a problem for our view. As Kyburg (1970, pp. 35–36) argues, the use of the principle of indifference in this case is “perfectly analogous” to the use we put it to in solving the book paradox. In that case we said that the alternatives *spades* and *non-spades* were not equal because there are more ways a card can be a non-spade than a spade. But then, for the same reason, it seems that we should deny that the structure descriptions 2H, 1H, and 0H, are equally possible. Making things worse, there are also problems associated with Carnap’s method. First, it is ad hoc to choose to apply the principle of indifference to structure descriptions just because we want to be able to allow for learning from experience (Gillies, 2000, p. 46). Second, Carnap’s solution is arbitrary: if the goal is to accommodate learning from experience, Carnap’s method is only one of infinitely many ways to do this, a fact that he himself proved (Carnap, 1952). (To his credit, Carnap (1955) is quick to admit this. He also admits that there are other methods besides the one he chose that have desirable features). Remember that what we are after is a logical principle not one that is merely useful.

A further difficulty for Carnap’s method which seems to have gone unnoticed is that it has the opposite problem to Keynes’ method. That is, it gives us learning from experience even when we don’t want it! Consider an urn which is known to contain an equal proportion

[9] The same result can be attained from Table 2 by using the Bayesian formula: $Pr(H|E) = Pr(H \& E) / Pr(E)$; or in this case: $Pr(B_2|B_1) = Pr(B_2 \& B_1) / Pr(B_1)$.

$Pr(B_1) = 1/2$

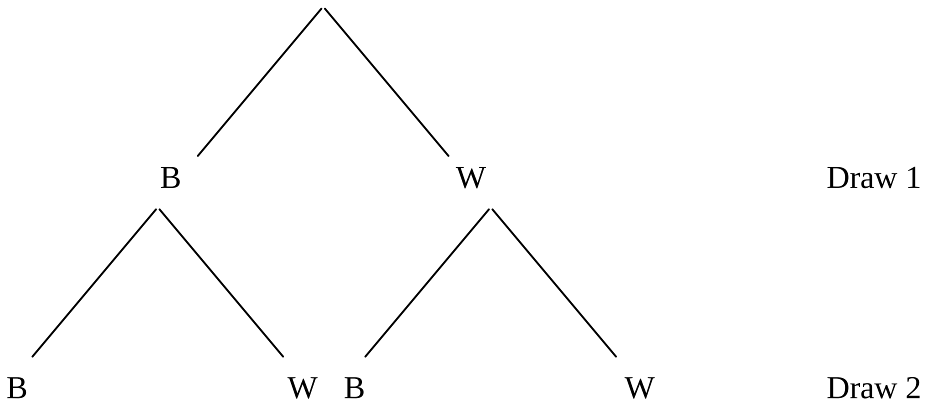
$Pr(B_2 \& B_1) = Pr(BB) = 1/3$

$$Pr(B_2|B_1) = Pr(B_2 \& B_1) / Pr(B_1)$$
$$= {}^1/_3 \,/\, {}^1/_2$$
$$= 2/3$$

[10] Laplace’s *rule of succession,* which states that the probability of an outcome which has occurred m times occurring again is $(m + 1)/(m + 2)$, also gives the answer 2/3 (see Zabell, 2005).

of black and white balls. Supposing now that each ball that is drawn is put back into the urn afterwards it is clear that no possible selections from the urn should affect our judgements as to the probability of drawing a black ball.[11] The case is analogous to a fair coin. If we are certain that the coin is fair, then no matter how many heads or tails we toss in a row, the probability of the next remains 1/2. The tendency to think otherwise is known as the "gambler's fallacy". Possible outcomes of two selections from the urn can thus be represented as in Figure 1.

Figure 1



The diagram is fairly self-explanatory. The first division of the tree (moving from the apex of the triangle downwards) represents the possibilities associated with the first ball selected. The bottom two divisions show the possibilities for the second draw given that the first ball drawn was black (left fork) or white (right fork). Each path beginning at the apex and ending at the base represents a possible state of affairs over two tosses. Reading these off the diagram we see that the possible states of affairs are BB, BW, WB, WW—the state descriptions. It is hard to see how one could justify the ascription of equal possibility to the structure descriptions

[11] Assuming we know that there is no way to systematically select balls of one type.

2B, 1B, and 0B, in this case. Not only is it quite obvious that the state descriptions are equal alternatives, but we don't even want learning from experience to be an option in this case. However, the application of Carnap's method gives the result, 2/3, as before.

Given the flaws in both methods it is not hard to see why commentators have seen this as being seriously problematic for the classical theory (and for logical accounts of probability in general). Carnap's argument alone seems like good reason to forgo the identification of equal possibility with state descriptions, but his solution is no better. Luckily, the notion of equal possibility we relied upon in our solution to the book paradox, and the application of the principle of indifference to state descriptions that follows from its endorsement, is perfectly compatible with learning from experience and with Carnap's method (in our original case). To see how let us return our attention to the outcomes represented in Figure 1. Is it not strange that these are exactly equivalent to those given by Keynes and Carnap for our first case? Remember that our first case involved an urn with an unknown distribution of black and white balls (let us call it a "mystery urn"), whereas the second involved an urn which contained an even ratio of balls (let us call it a "fair urn" in analogy with a fair coin). On the face of it we would expect the sample space associated with each case to be quite different, because the goal is to model our epistemic state in each circumstance. In the case of the fair urn our ignorance extends only to colour of the ball that will be drawn. Our ignorance in the case of the mystery urn is far greater, however, because not only do we not know the colour of the ball to be drawn, but we do not know what proportion of balls of each colour the urn contains either.

As it turns out, we can account for these differences when we model the problem. To do so it will be helpful to simplify our example significantly (as we did with the coin tosses in Section I). We will also need to specify more details of the case than is common in the literature. This does not count against our approach in any way. On the contrary; most authors within the literature confuse matters by failing to be clear whether they take the urn to hold a finite or infinite number of balls or whether balls are imagined to be replaced after selection or not (see Zabell, 2005, pp. 67–68 for a brief discussion). We will see that these factors matter, especially for Carnap's approach. For now let us stipulate that we are dealing with an urn known to contain two balls, either black or white. A black ball has been drawn from the urn and subsequently replaced. We want to know the probability of drawing a second black ball. At this stage we can deduce that the urn may contain either two black balls, one black ball and one white, or two white balls. There thus appear to be three possible urns. We will (unimaginatively) call these urns "Urn 1", "Urn 2", and "Urn 3", respectively. For each of these urns we can then model the possible sequences of draws. This is shown in Table 4 below. (It is immediately clear from this that there are two parts to the problem not one and that both Keynes and Carnap overlooked this). To our satisfaction we find that Urn 2—and only Urn 2—has the same possibilities associated with it as the fair urn in Figure 1.[12] The sample space in each case is, as hypothesised, quite different.

Table 4

Urn 1	Urn 2	Urn 3
BB	BB	WW
BB	BW	WW
BB	WB	WW
BB	WW	WW

[12] The only difference between the cases is that we did not specify how many balls the urn in Figure 1 contained. It can be shown that only the distribution of balls matters to the probability, and not the number itself (see footnote 14).

Reading off the table we find that six of the sequences of draws begin with a black ball. On five of these occasions a second black ball will be selected, so $Pr(B_2|B_1) = 5/6$. The prior probability of drawing a black—Pr(B)—on either the first or second draw is 1/2 (as it should be). The posterior probability is higher than this so we have modelled learning from experience. And we have apparently done so by applying the principle of indifference to state descriptions!

Two things need to be shown before we can deem this result favourable to our view. It needs to be shown that (i) the three urns are in fact equally possible as we assumed; and (ii) the sequences in the table are the state descriptions of the case. (We will see why each of these is important as we deal with it). At first glance it may seem unnecessary to argue for (i); however one might point out that Urn 2, which contains one black ball (1B), can be described in two ways—BW and WB—whereas Urn 1 and Urn 2 can each only be described in one (BB and WW respectively). That is, 2B, 1B, and 0B are not of the same form. This is the same point that Keynes made for *draws* from the urn. We can counter this objection by pointing out that BW and WB are descriptions of the same urn—order is simply irrelevant here, whereas it is not in the case of draws. Of course, we can make order relevant by stipulating how "BW" and "WB" correspond to different arrangements of balls within an urn. For instance, we can imagine that our urn is a cylindrical tube slightly larger in diameter than the balls; we then stipulate that "BW" describes an urn which contains a black ball on top and a white on bottom, and vice versa for "WB" as in Figure 2.

Figure 2

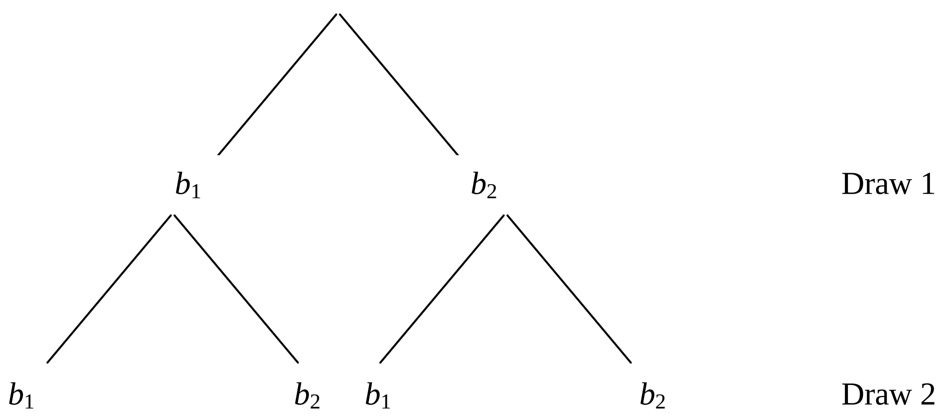
This makes no difference. The same two divisions are available for both 2B (Urn 1) and 0B (Urn 2). Disregard for a moment the colour of the balls and instead individuate them by name. Call one "b_1" and the other "b_2". We then get:

Figure 3

But now it is clear that the colour of the balls is irrelevant to our question. The possible arrangements of the balls shown in Figure 3 are available whatever the respective colours of b_1 and b_2. If both balls are black there are nevertheless two ways they can be arranged. (This may seem quite obvious but it is necessary to demonstrate it because the same does not hold for draws from the urn. Using numerals to individuate balls will also allow us to derive a general answer to questions of the type we are considering).

The second aspect of Table 4 which required proof was (ii) that the twelve sequences are in fact the state descriptions of the case; for perhaps they are not and we should not count each "BB" in the table as a separate possibility. (There would thus be only one BB sequence, one WW sequence, one BW sequence, and one WB sequence. We have already seen that applying the principle of indifference to these alternatives cannot lead to learning from experience). We can demonstrate that (ii) is true by considering a single urn containing two balls: b_1 and b_2 as before. Given no knowledge of the selection method these are also the equal possibilities on the first draw. Remembering that we are sampling with replacement, the alternatives for the second draw are the same. This is shown in Figure 4.

Figure 4



The possible sequences of balls are therefore b_1b_1, b_1b_2, b_2b_1, b_2b_2. We can see that this is true regardless of the colours of the individual balls, as these have been left unspecified. Suppose that b_1 and b_2 are both black (or that our urn is Urn 1 in Table 4). Reading off the figure we see that there is not one possible sequence that meets the description "BB", but four.

Extending this result to the mystery urn scenario we can derive the results shown in Table 5. (Imagine three of the diagrams in Figure 4 side by side. To avoid confusion, the balls in Urn 1 have been labelled "b_1" and "b_2"; the balls in Urn 2, "b_3" and "b_4"; and those in Urn 3, "b_5" and "b_6".).

Table 5

Urn 1	Urn 2	Urn 3
b_1b_1	b_3b_3	b_5b_5
b_1b_2	b_3b_4	b_5b_6
b_2b_1	b_4b_3	b_6b_5
b_2b_2	b_4b_4	b_6b_6

Table 5 is structurally equivalent to Table 4. Take Urn 2, for example. We know it contains one black and one white ball. We have seen that the order of these is irrelevant so we can arbitrarily stipulate that b_3 is black and b_4 white. From this we derive the values shown in column two of Table 4: namely, BB, BW, WB, and WW.

What Table 5 demonstrates is that the five BB sequences in Table 4 are distinct state descriptions. This means that BB is not itself a state description in the mystery urn case as Carnap and Keynes supposed! (On the other hand, it is easy to see that BB *is* a state description for a fair urn: see Urn 2, Table 4). Applying the principle of indifference to the genuine state descriptions (shown in Table 4) gives us the desired learning from experience. The reason why previous authors failed to provide a satisfactory solution was because they misapplied the principle of indifference. The equal possibilities are not in this case BB, BW, WB, and WW or 2B, 1B, and 0B. To think this is to overlook the fact that there are two parts to the problem: the possible ways the urn could be, *and* the possible selections from the urn. Keynes and Carnap considered only the latter, but there is a big difference between these two cases. (Just how big can be seen by comparing Table 1 and Table 4).

It is worth showing one last thing before we conclude our discussion in this section. The following are the state descriptions for two selections *without* replacement from a mystery urn containing two balls (remembering that we have been considering sampling *with* replacement thus far):

Table 6

Urn 1	Urn 2	Urn 3
b_1b_2	b_3b_4	b_5b_6
b_2b_1	b_4b_3	b_6b_5

Substituting the appropriate values for b_1, b_2, b_3, and b_4 we get:

Table 7

Urn 1	Urn 2	Urn 3
BB	BW	WW
BB	WB	WW

Table 7 is structurally equivalent to Table 3 which represented the sample space from which Carnap derived his solution of 2/3! This is a satisfying result. But note that in this case the structure descriptions 2B, 1B, and 0B are equal possibilities *because* they stand for the same number of state descriptions. It is only in virtue of this that Carnap's method is successful. If he had in mind sampling without replacement from an urn containing a finite number of balls his solution is correct, but only incidentally so.[13] Application of the principle of indifference to structure descriptions will sometimes give the right answer. Application of the principle to state descriptions will always do so.

This brings us to the end of the present section. We have seen that despite initial appearances the principle of indifference can be applied consistently to state descriptions—or propositions of the same form—in a way that allows for learning from experience. Critics and proponents alike have failed to identify the correct state descriptions because they considered only part of the problem at hand. It is no surprise that Keynes' solution makes impossible learning from experience—he mistakenly models a known fair urn! It is easy to confuse the two types of case—mystery urn and fair urn—because both have the same prior probability of drawing a black (or white) ball: namely, 1/2. But it should now be apparent that they are far from the same and should not be treated as such. There is a clear and important difference in what is known in each case and any good theory of logical probability must take this into account.

[13] The finite version of the *rule of succession*, which states that the probability of an outcome which has occurred m times occurring again is $(m + 1)/(m + 2)$ functions as a proof that the number of balls in an urn is irrelevant to the posterior probability of an outcome when sampling without replacement (Zabell, 2005, p. 40). (Because there is no place for the number of balls to even enter into the formula). Consequently it seems that Carnap's answer will be correct for urns containing any number of balls. Our discussion also suggests that the rule of succession is an extension of the classical theory (given that we have shown that the probabilities in these cases supervene on state descriptions) and not an ad hoc addition made to accommodate learning from experience as is often thought (e.g. Háyek, 2011).

III

We now come to what many believe to be the most serious objection to the principle of indifference: *Bertrand's paradox* and its close relatives. The difficulties raised in sections I and II against the classical theory might be seen as reasons to adopt more sophisticated logical theories of probability. Bertrand's paradox and the problems it raises are thought by many to block this move, showing the "impossibility of the ideal of logical probability" (van Fraassen, 1989, p. 293). In fact we will see in this section that Bertrand's paradox has less to do with the principle of indifference than is typically thought. We will also address some misunderstandings about the role of principle of indifference and two allegations of circularity.

In what follows we will not discuss Bertrand's original (1889) random chord paradox which is the typical referent of the term "Bertrand's paradox" (although it is one of a family of paradoxes that he presents). Instead we will focus on a variant discussed by Bas van Fraassen (1989, pp. 303–304).[14] Van Fraassen's version of the paradox—the *perfect cube factory*—is mathematically simpler, more concise, and (to my mind) more compelling. He writes:

> A precision tool factory produces iron cubes with edge length ≤ 2 cm. What is the probability that a cube has length ≤1 cm, given that it was produced by that factory? (p. 303).

[14] For a particularly clear overview (and discussion) of the original paradox see Tissier (1984). It should be noted that there are small differences between the original and the variant we discuss. However, what we will say here does not depend on any such difference. If anything, the original version of the paradox is more easily solved.

At first glance the question is easily answered. The principle of indifference apparently dictates that we assign the intervals (0, 1] cm and (1, 2] cm equal probability.[15] So the probability of the factory producing a cube between 0 and 1 cm in length is 1/2. However, posing the same question in terms of side *area* gives a different answer. The factory produces cubes between 0 and 4 cm^2, and the question is now: "What is the probability of getting a cube ≤ 1 cm^2?" Applying the principle of indifference in this case gives the probability 1/4. But these are just two different ways of describing the very same problem so we have a contradiction!

Why the problem arises is quite clear. A uniform distribution over side length amounts to a non-uniform distribution over side area and vice versa. In fact, any non-linear transformation of the problem will potentially result in a different answer. The problem is multiplied by noting that there are infinitely many transformations of this type.

Bertrand's paradox raises several difficulties for the principle of indifference, which are similar to those that arose in Section II. The paradox suggests that the principle of indifference is inconsistent or ambiguous, for it leads to contradictory results (van Fraassen, p. 315). Furthermore, even if the principle were made consistent it would be a matter of luck whether the resulting solution would be the right one (van Fraassen p. 315). For instance, suppose we disambiguated the principle in a way that validated the result Pr(≤1 cm) = 1/4. Why should we expect *that* to be the right answer? Without further information about the factory it seems this result would be unjustified.

The general consensus outside of philosophy—and the view which will be defended here—is that the problem is not well-posed (e.g. Tissier, 1984). On this view the paradox arises because the descriptions of the problem in terms of side length and side area are not in fact equivalent as we assumed. (This is not to say that a cube with side length 2 cm is

[15] Parentheses indicate an *open interval* in which the endpoints are not included. Square brackets indicate a *closed interval* in which the endpoints are included. The interval (0, 1] cm includes 1 cm but excludes 0 cm.

different to a cube of side area 4 cm^2). The point is perhaps easiest to demonstrate if we stipulate a further detail of the case. Let us suppose that inside the factory is a dial with numbers on it. The direction in which the needle on the dial points determines the size of the cube produced. Now, imagine that values around the dial are limited to rational numbers—that is, numbers which can be expressed as the fraction of two integers. (We shall dispense with this assumption shortly). If the values on the dial are in square centimetres (between 0 and 4 cm^2), as in the second description of the factory, and the dial is turned to "2 cm^2", the factory will produce a cube of precisely that side area. But note that a cube of that size has a side length which is not a rational number (namely: $\sqrt{2}$). It is thus impossible—according to the first description in terms of side length— for the factory to produce this cube because there is no such number on the dial! These are not just different descriptions of the same scenario.

If we drop the proviso regarding rational numbers it is no longer the case that the two factories produce cubes of different sizes in the way explicated above. Now the side lengths of the cubes correspond to the real numbers which (unlike the rational numbers) have no gaps. Nevertheless, in each case the *measure* of the space of possibilities is different.[16] That is, in the first case the interval (0, 1] cm takes up half of the dial whereas in the second, the interval (0, 1] cm^2 take up only a quarter of the dial. It is quite clear that we obtain different answers because of this and that the paradox relies upon the false assumption that the statement of the problem unambiguously describes a single factory.[17] Note, also that this has nothing to do with the principle of indifference per se. The problem can be stated in such a

[16] Length, area, and volume are good examples of measures. See Shackel (2007) for an example of how one might adapt the principle of indifference to continuous cases. I pass over this issue here as I don't believe Bertrand's paradox arises specifically because it is a continuous rather than discrete case.

[17] That this is so is even clearer in the original "random chord" version of the paradox. Tissier (1984) reports that computer simulations confirm that each solution to the paradox corresponds to a different distribution. Each solution is therefore the correct solution to a different specific problem.

way that it is made explicit that each cube less than or equal to 1 cm in length is equally *probable*, but the paradox still arises.

Commentators in the philosophical literature, following van Fraassen (1989, Chapter 12) tend to see this as an illegitimate response for those whose concern is to defend the principle of indifference (e.g. Eagle, 2011; Gillies, 2000; Shackel, 2007).[18] Van Fraassen's (1989, p. 305) objection is this:

> But that response asserts that in the absence of further information we have no way to determine the initial probabilities. In other words, this response rejects the Principle of Indifference altogether. After all, if we were told as part of the problem which parameter should receive a uniform distribution, no such Principle would be needed. It was exactly the function of the Principle to turn an incompletely described physical problem into a definite problem in the probability calculus.

There are in fact several points here, all of which are false. The first is that endorsing the "ill-posed" response (as we shall call it) amounts to the claim that there is no probability of getting a cube less than or equal to 1 cm in side length. In fact all we have said is that the paradox relies on a conflation of several distinct problems. We have not claimed that there is no answer to Bertrand's question. Bertrand's *paradox* arises because we interpret different statements of the problem as different descriptions of a single factory. It is resolved by noting that this was a mistake. Bertrand's *question*, on the other hand, is simply, "What is the probability of the factory producing a cube of side length ≤ 1 cm?"[19] This question would

[18] But see Marinoff (1994) for a defence of the "ill-posed" solution. However, Marinoff seems to have overlooked van Fraassen's criticisms (although he cites him approvingly).

[19] I follow Shackel (2007) in my choice of terminology here.

appear to be answerable in the same way the question "What is the probability of drawing a black ball from an urn containing black and white balls in an unknown proportion?" was. (We will return to this shortly).

Van Fraassen's second claim is that being given more information about the factory would do away with the need for the principle of indifference in the first place. This is simply not true. Imagine that we were told as part of the problem that the dial inside the factory has square centimetres as units (ranging from 0 to 4 cm^2). Is the principle of indifference really not needed in this case? Van Fraassen seems to think that being given this information amounts to being told the probabilities. But there is no mention of probabilities in this new problem, and without a particular theory of probability we are none the wiser as to what they are. Only when we apply the principle of indifference to the interval (0, 4] cm^2 do we find that the desired probability in this case is 1/4. Furthermore, if we have been told the probabilities in this case surely we are also told the probabilities in classical cases of dice, card games, and coin tosses. But we aren't. If this is the claim van Fraassen is making he must think the classical theory is circular, and that equal possibilities just are equal probabilities. We have already seen that this is false. It is a substantive claim of the classical theory that probabilities supervene on possibilities. Rival theories of probability deny this. Some philosophers, for instance, have held the view that if the factory has never produced a cube *there is no probability* of a cube of length $\leq$ 1 cm being produced. That this claim makes sense—even if it appears to be false—demonstrates that there is no vicious circularity here. (There is of course circularity if the classical theory is true. But all definitions are circular in this way).

This being said, a more serious charge of circularity needs to be addressed at this point. The principle of indifference allows the classical theorist to claim that equal possibility and equal probability are distinct notions. According to the classical theory, two possibilities are equal if there is no known reason to favour one over the other; that is, if there is equal evidence for each. Háyek (2011) argues that this may only push the problem of circularity back because the notion of equal evidence appealed to here seems itself to be probabilistic. Specifically, it seems that the only way to spell out what "equal evidence" means is by reference to conditional probabilities. That is, there is equal evidence for two outcomes O_1 and O_2 if and only if $Pr(O_1|E) = Pr(O_2|E)$ (where "|E" is to be read, "given the evidence"). Certainly, Keynes' own treatment seems to fall foul of this objection. He claims that a piece of evidence E is irrelevant to an outcome O (roughly) if $Pr(O) = Pr(O|E)$ (Keynes, 1921, p. 55). We can, however, resist this line of reasoning by observing that the notion of equal evidence we have made use of throughout the paper is not probabilistic in this way. What made it the case that Urn 1, Urn 2, and Urn 3 in Section II were equally possible was the fact that there were an equal number of possible states of affairs associated with each outcome. To resist Háyek's line of criticism we must insist that evidence is always relative to number of possibilities in this way. Given what has been said throughout the paper this is very plausible. In fact, the idea that evidence can only be understood in relation to probabilities makes it utterly mysterious why a given piece of evidence *is* evidence at all.

This brings us to the final point made by van Fraassen in the passage above. He claims that we should reject the "ill-posed" response to Bertrand's paradox because "It was exactly the function of the Principle to turn an incompletely described physical problem into a definite problem in the probability calculus" (van Fraassen, p. 305). This is representative a misinterpretation of the principle of indifference which appears to be widespread. The role of the principle of indifference is *not* to turn an incompletely described problem into a definite one in the sense he has in mind. As critics of the principle rightly point out, this would require unjustified assumptions about contingent empirical matters. The proper role of the principle of indifference is simply to tell us what equal possibilities are (namely: possibilities for which we have equal evidence). The classical method does not depend on any empirical assumptions. The goal in any given case is to accurately model our epistemic state with regard to the possibilities. In Section II we saw that failure to do this leads to problems. Keynes' solution to the mystery urn case in that section seems to be what van Fraassen has in mind when he talks of an incompletely described problem turned into a definite one. But we saw that Keynes' solution was inadequate precisely because of this. In effect his solution assumed that the urn was "fair", but this kind of assumption not only goes against the spirit of the principle, which essentially tells us to assume nothing, but straightforwardly contradicts it. For there is no known reason to think the urn is Urn 1, Urn 2, or Urn 3; and hence by the principle of indifference we should assign these equal probability, *not* discount Urn 1 and Urn 3. The view that the role of the principle of indifference is to rule out two of these urns is mistaken. The correct use of the principle and the classical method, given an incompletely described problem, *is to model it as such*.

There is one last problem we must address before our defence of the principle of indifference is complete. Earlier, we made a distinction between Bertrand's *paradox* and Betrand's *question*. We concluded that the "ill-posed" solution is perfectly sufficient as an answer to the former but that an answer to the latter should be possible. Shackel (2007) argues that Bertrand's question, which is "What is the probability of getting a cube ≤ 1 cm in length?" is *supposed* to be ambiguous in the way the ill-posed response claims it is. For this reason, the thinks that our response fails. He is wrong about this. As we saw, the formulation of the paradox clearly involves the assumption that each description of the problem to which the principle of indifference can be applied is a description of the very same factory. Nevertheless, Shackel is right that Bertrand's question can be reinterpreted as a general question in this way. Here the question should be understood as, "What is the probability of a cube of side length ≤ 1 cm *given that the factory produces cubes according to side length, or* *side area, or side volume, or...*?" It seems we cannot refuse to answer this question on the grounds that it is ill-posed, because it is essentially no different to the mystery urn case in Section II. But Shackel thinks we cannot answer to it. He argues that Bertrand's paradox recurs for this "general" question and that this leads to an infinite regress. His argument is that the original paradox arises because there is no unique or natural measure relative to which we can apply the principle of indifference so we must be indifferent between all the possible measures (side length, side area, volume, and so on). But there will be multiple measures of *these* measures, none of which will be *the* unique measure, so we should be indifferent between these too. And so on, ad infinitum (Shackel, 2007, p. 173).

The first thing to note in response is that "finite" problems don't seem to lead to infinite regress. Consider a restricted version of the problem: "What is the probability of a cube of length ≤ 1 cm, given that the factory produces cubes by side length, or side area, or volume?" The principle of indifference tells us that each of the three ways the factory could produce cubes is equally possible (and thus equally probable) because there is no known reason to expect the factory to produce cubes in one way rather than another. This is exactly analogous to the mystery urn case and there is no regress problem here. Shackel's argument presupposes that it is possible that there be more equal possibilities associated with one method than the others. But to say that there could be more equal possibilities associated with one method is—on our view—to say that there could be known reason to expect the factory to produce cubes one way rather than another. However, all the evidence we have is given by the statement of the problem and the statement of the problem contains no such reason. (The case is, after all, entirely hypothetical). What is possible is governed by the information given in the statement of the problem. To be sure, we can ask whether it is possible that the problem contains information that we have missed. But it makes no sense to ask whether it is possible *given the statement of the problem* that there is evidence for one method over the others contained in the statement of the problem. It either does or it doesn't.

Can we apply this to the infinite version of the case? If what we have said is correct it appears that we should be able to do so; for the finite problem, and also the book paradox, are closely related to it and appear to have determinable solutions. What we want is a measure which preserves the features of the finite case. We will not attempt to find such a measure here (although a non-standard approach involving infinitesimals looks promising). We can note, however, that if there is no such measure and our answer to the finite version of the case does not carry over to the infinite version this will be because it is indeterminate whether our evidence for a cube within the interval (0, 1] cm in length is equal to our evidence for a cube within the interval (1, 2] cm. (And not just because of some epistemic deficiency on our part). If our problem is like this it is deeply ill-posed. Perhaps it is even nonsensical in the same way it is nonsensical to ask whether "*x* is red" is true or false, but this is neither here nor there. As we saw, the role of the principle of indifference is to identify the equal possibilities associated with a given problem. If these are not determined by the statement of the problem the principle simply cannot be applied.[20] Of course this would mean that no answer can be given to Bertrand's question, but who said that there must be one?[21][22]

---

[20] Luckily such a situation will not arise for real-world cases if meaning is partly dependent on the world as seems plausible. For imagine that we did not know the meanings of *spades* and *non-spades*: nonetheless they would still refer to the same cards, and the probability of *spades* would still be 1/4.

[21] Certainly Keynes did not. His response to Bertrand's paradox is to point out that it is ill-posed, as we have done.

[22] Presumably the thought is that there is no known reason to favour *any* outcome so we will end up with an irresolvable version of the book paradox (with any two intervals having probability 1/2). In this case we cannot fail to apply the principle of indifference because we have an abundance of equal possibilities. (The problem is just that these contradict each other). However, if there is no known reason to favour the interval (0, 1] cm over (1, 2] cm then there *is* known reason to favour (0, 1] cm^2 over each of (1, 2] cm^2, (2, 3] cm^2, and (3, 4] cm^2; and vice versa. If Shackel is right and there is no unique measure in this case, then it is simply indeterminate whether the available evidence for (0, 1] cm and (1, 2] cm is equal or not. But if it is indeterminate then there

Bertrand's paradox was supposed to give us reason to reject the principle of indifference, yet it does no such thing. Nor does Shackel's version of it. Either there is a unique probability of the factory producing a cube ≤ 1 cm in side length, or the problem is not one to which the principle can be applied. Neither of these possibilities is a problem for the classical theory.

CONCLUSION

In this paper I have defended the classical theory of probability against many of the objections that have led to its abandonment. The principle of indifference on which it relies has been misunderstood and misused. A proper understanding of it appears to avoid the aforementioned objections. There are of course other objections which we have not discussed. Nor has anywhere near a comprehensive account of the classical theory been given. Therefore, there are issues that remain to be addressed, and I don't claim to have shown definitively that the use of the principle of indifference is justified or that the classical theory is correct. However, I do believe that I have shown that the objections discussed throughout the paper are far from decisive. The classical theory is not circular or restricted to artificial cases. The principle of indifference is not inconsistent. Nor does its application require tacit and unjustified empirical assumptions. If there *are* reasons to reject the classical theory those discussed in this paper are not among them.

MICHAEL DUNCAN

are no equal possibilities on the basis of which to calculate the contradictory probabilities which are needed to generate the paradox.

REFERENCES

Bertrand, Joseph (1889) *Calcul des Probabilitiés*. Paris: Gauthier-Villars.

Carnap, Rudolph (1952) *The Continuum of Inductive Methods*. Chicago IL: University of Chicago Press.

Carnap, Rudolph (1955) "Statistical and Inductive Probability" in Antony Eagle (ed.) *Philosophy of Probability: Contemporary Readings*, 2011. New York: Routledge.

Eagle, Antony (2011) in Antony Eagle (ed.) *Philosophy of Probability: Contemporary Readings*. New York: Routledge.

Gillies, Donald (2000) *Philosophical Theories of Probability*. New York: Routledge.

Háyek, Alan (2011) "Interpretations of Probability" in Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy (Summer 2012 Edition)*, URL = <http://plato.stanford.edu/archives/sum2012/entries/probability-interpret/>.

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

Kyburg, Henry E. (1970) *Probability and Inductive Logic*. New York: Macmillan.

Marinoff, Louis (1994) "A Resolution of Bertrand's Paradox", *Philosophy of Science*, 61, pp. 1–24.

Shackel, Nicholas (2007) "Bertrand's Paradox and the Principle of Indifference", *Philosophy of Science*, 74, pp. 150–175.

Tissier, P. E. (1984) "Bertrand's Paradox", *The Mathematical Gazette*, 68(443), pp. 15–19.

van Fraassen, Bas C. (1989) *Laws and Symmetry*. Oxford: Clarendon Press.

Zabell, S. L. (2005) *Symmetry and its Discontents*. New York: Cambridge University Press.