Brit. J. Phil. Sci. **55** (2004), 137–145, axh107

Dissolving the Wine/Water Paradox

Jeffrey M. Mikkelson

ABSTRACT

Von Mises' wine/water paradox has long served as an argument against the Principle of Indifference. A solution to the paradox is proposed, with a view toward resolving general difficulties in applying the principle.

The Principle of Indifference (PI), an artifact of the classical theory of probability, says that we should assign equal probability to any mutually exclusive and jointly exhaustive set of possible outcomes, iff we have insufficient reason to consider any one of these outcomes more or less likely than any other. Despite its intuitive appeal and its formative role in the field of probability, PI has fallen into serious disrepute among probability theorists. Perhaps the most common and compelling argument against PI is that it leads to irresolvable paradox. Beginning with Bertrand ([1889]), authors have attempted to show that applying PI to certain kinds of examples leads to contradiction. Opinions are divided as to which of these so-called Bertrand paradoxes can be resolved simply by making PI more precise, but critics and defenders of the principle agree that not all of them can be. One particular paradox, von Mises' wine/water paradox, plays a curiously pivotal role in this discussion. Everyone seems to agree that it has no solution. Van Fraassen ([1989]) claims that the paradox signals the 'ultimate defeat' of the Principle of Indifference, nullifying the 'Pyrrhic victory' won by Poincaré ([1912]) and Jaynes ([1973]) over other Bertrand paradoxes. Gillies ([2000]) calls it the one 'fatal' objection to PI, and Oakes ([1986]) uses it to confirm that 'the classical conception of probability cannot withstand more than casual examination'. Even Jaynes ([1973]) and Schlesinger ([1991]), lonely defenders of PI, throw up their hands at this paradox, arguing that it simply stakes out the limitations of the principle. The fact that so many critics rely on this one example to discredit PI is in itself cause for some suspicion. At the very least, it suggests that dissolving the wine/water paradox would score a major victory for PI. I hope to supply this victory.

To begin, a few clarifications. PI, for my purposes, is a constraint on the assignment of *epistemic probability*. In other words, it pertains to our degree of rational belief, given the evidence.[1] Furthermore, PI pertains just in those cases where evidence is lacking in some relevant respect. So, for example, if we toss what appears to be a fair coin, not knowing whether it is in fact biased (or in what way, or to what degree), then we should assign equal probability to each of the two possible outcomes. (It wouldn't be hard, of course, to win money from someone willing to bet on contrary convictions.) On the other hand, experience or context can furnish sufficient reason to ignore PI. In many biological contexts, for example (such as, say, predictions about the phenotype of ravens), a bias toward certain kinds of outcomes (i.e., homogeneous ones) seems justified, given what we know about nature. In such cases, however, PI is absolved from responsibility. It is an explicitly *conditional* rule for assigning equiprobability, and need not militate against legitimate reasons to think that statistical correlations exist in nature. So much is relatively uncontroversial.

Much ado, however, has been made about the fact that PI can be applied in different ways to certain problems, yielding contradictory results. Say, for example, that we toss *two* coins. Borrowing Carnap's ([1950]) terminology, we can remain indifferent over each of the four possible *state descriptions*: {HH, HT, TH, TT}, assigning 1/4 probability to each outcome. Alternatively, we can remain indifferent over each of the three possible *structure descriptions*: {2H, 1H1T, 2T}, assigning 1/3 probability to each. Now, ask a specific question and these two methods will yield conflicting answers. (A simple one: what's the probability that both of the coins will come up heads?) Such examples have kindled a great deal of controversy, even among fellow advocates of PI.[2] In the case just mentioned, however, there is at least one good reason to think that one of the methods is misguided. The second method, which asks us to remain indifferent over each of the three structure descriptions, ignores the 'insufficient reason' part of PI. We *do* have sufficient reason to think that '1H1T' is more likely to obtain than '2H' or '2T', namely, that there are more ways for '1H1T' to obtain (twice as many, in fact).

1 Much confusion about PI stems from the conflation of epistemic probability with *objective chance*. While PI constrains our degree of rational belief, it does not determine objective chance, much less constrain the world itself. Kneale ([1949]) suggests that the authors of the classical theory (James Bernoulli, Laplace, etc.) are responsible for blurring this distinction, but critics of the principle often perpetuate the confusion. Van Fraassen, for example, inveighs in the following way against 'symmetry arguments', of which PI is an example: 'Nothing continent and about the world can be deduced by logic alone. The *a priori* appearance is therefore deceptive' ([1989], p. 233).

2 Keynes ([1921]), for example, favors state descriptions (which he calls 'constitutions') over structure descriptions (which he calls 'ratios'), while Carnap ([1950]) favors structure descriptions. The dispute points up a serious problem for Bayesian conditionalization—how to update probabilities in light of evidence—and is quite fascinating, but is somewhat orthogonal to my paper.

A more damaging challenge to PI, in the eyes of both critics and defenders of the principle, is the Bertrand paradox. This type of paradox has a form similar to the two-coin example, but appears (on the surface, at least) far less manageable. We begin by applying PI in a plausible way to a given problem, generating a uniform distribution of probability over its sample space. So far so good. But then we discover that by applying PI in a different but equally plausible way to the same problem, we generate a completely different distribution. It then becomes fairly easy to educe conflicting answers to specific questions. Hence the paradox. Examples include Bertrand's own chord paradox, Buffon's needle problem, and the cube paradox. Rather than rehearse these examples and their proposed solutions, I'd like to dive straight into the wine/water paradox, for which, oddly enough, no solution has been proposed:

> There is a certain quantity of liquid. All that we know about the liquid is that it is composed entirely of wine and water, and that the ratio of wine to water (x) is between 1/3 and 3. So $1/3 \leq x \leq 3$. Now, what is the probability that the ratio of wine to water is less than or equal to 2 (i.e., that $x \leq 2$)?

There are two standard ways to answer this question. (1) Remain indifferent over the possible ratios of wine to water (x). In other words, treat all x-values as equiprobable, assigning a uniform probability distribution between 1/3 and 3. We then answer the problem by calculating the proportion of x-values that are less than or equal to 2, within the space of possible x-values. The answer is straightforward:

$$P(x \leq 2) = (2 - 1/3)/(3 - 1/3) = 5/8$$

(2) Remain indifferent over the possible ratios of water to wine (y). Since $1/3 \leq x \leq 3$, it follows that $1/3 \leq y \leq 3$. Moreover, to say that $x \leq 2$ is equivalent to saying that $y \geq 1/2$. (Ratios have a neat inverse relation.) So, treating all y-values (rather than x-values) as equiprobable, we assign a uniform probability distribution between 1/3 and 3. We then answer the problem by calculating the proportion of y-values that are greater than or equal to 1/2, within the space of possible y-values. The answer is again straightforward:

$$P(y \geq 1/2) = (3 - 1/2)/(3 - 1/3) = 15/16$$

So we have two conflicting answers to the question at hand—5/8 and 15/16—both of which result from a seemingly plausible application of PI.

Now, the wine/water paradox causes so much trouble for PI because there is no apparent way to choose between the two alternatives. Everything that can be said for the first method can also be said for the second, and vice versa: they are perfect counterparts. In fact, this antinomy so impresses probability theorists that many treat it as an exemplar, showing why we should 'regard it as clearly settled now that probability is not uniquely assignable on the basis of a Principle of Indifference' (van Fraassen [1989], p. 292). Like some other notable antinomies, however, this one deserves a second look.

Imagine that the quantity of liquid described above is poured into a graduated cylinder (Figure 1), and that the wine doesn't mix with the water, but sinks to the bottom. Depending on the actual quantities of wine and water in the liquid, the line separating them will fall somewhere between 1/4 and 3/4 of the total quantity (since $1/3 \leq x \leq 3$). Call this total range 'Ω'. We know that $x \leq 2$, so the line separating wine from water will fall at or below 2/3 of the total quantity. Call this range 'P'. We can see from Figure 1 that P occupies exactly 5/6 of the total range Ω. So $P(x \leq 2) = 5/6$.

The first thing to notice about this solution is that it's different in kind from the other two. Rather than remaining indifferent over the possible *ratios* of wine and water, we remain indifferent over their possible *quantities*. For reasons that will become clear, we can call this method the *symmetry method*. Ignoring, for the moment, the deeper reasons to prefer the symmetry method, there is a lot on the surface to recommend it. First, the solution doesn't depend on how the question is asked. Remaining indifferent over ratios leads to paradox because the two ways of asking the question about ratios (i.e., about x-values or about y-values) yield two different probability distributions, and therefore two conflicting answers. No matter how we ask the question about quantities, we get the same answer. In other words, the solution is invariant with respect to frame of reference.[3] Others—notably Poincaré and Jaynes—have invoked invariance while attempting to dismantle Bertrand paradoxes, arguing that an acceptable solution must be invariant with respect to certain transformations, frame of reference, and so on. To the dismay of PI's advocates, however, invariance does not distinguish between the two standard answers to the wine/water paradox. Fortunately, invariance does distinguish *an* answer—namely, the one I'm proposing.

A second reason to prefer the symmetry method is that remaining indifferent over quantities rather than ratios prevents us from making unjustified assumptions about the composition of wine and water in the cylinder. To remain indifferent over x-values or y-values is to make a substantial assumption about the liquid's composition. For example, to assign a uniform probability distribution to x-values between 1/3 and 3 is to say that it is *likely* that there is more wine than water—in fact, $P(x \geq 1) = 3/4$. To assign a uniform probability distribution to y-values between 1/3 and 3 is to assume the opposite—that it is *likely* that there

[3] The solution is also invariant with respect to *scale*, in the following sense: Say we use the first ratio method to answer the wine/water problem, choosing an arbitrary total quantity—say, 1 liter. The first ratio method asks us to remain indifferent over x-values, so $P(x \leq 2) = 5/8$. We can take this to mean that the probability that there is no more than 2/3 liter of wine (given that there is between 1/4 and 3/4 liter of wine) is 5/8. Now, choose a different total quantity—say, 2 liters—and ask: What is the probability that there is no more than 2/3 liter of wine, given that there is between 1/4 and 3/4 liter? Using the first ratio method once again, the answer changes to 25/32. If we use the symmetry method, the answer is in each case 5/6.

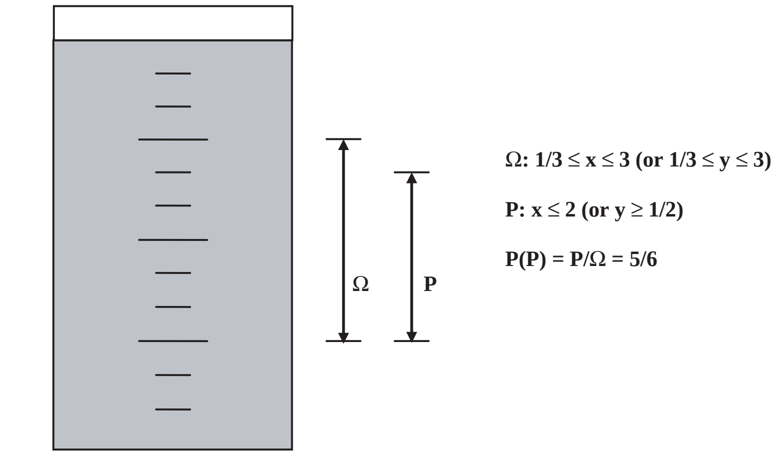

Figure 1.

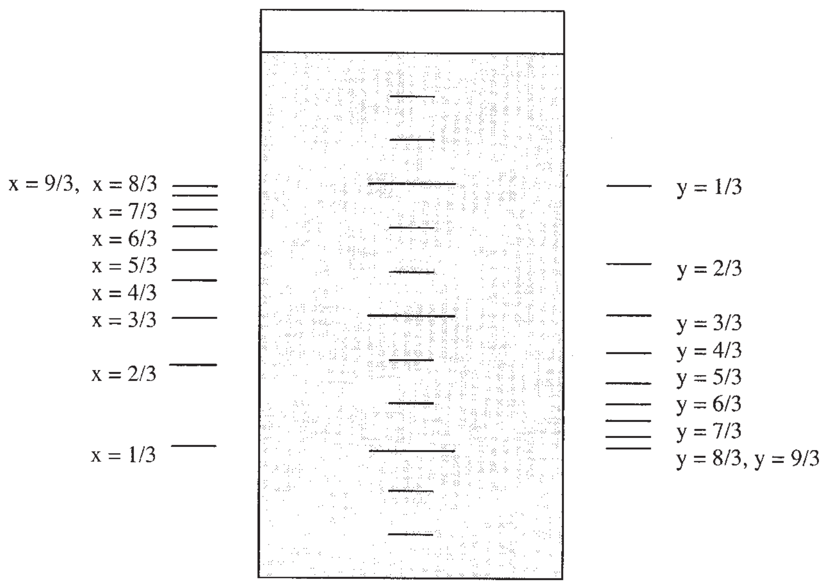

Figure 2.

is more water than wine ($P(y \geq 1) = 3/4$). To see what's going on here, it might help to plot some x-values and y-values on the graduated cylinder (Figure 2):

As Figure 2 illustrates, remaining indifferent over ratios generates probability distributions that are heavily biased. This point can be made into an embarrassing *reductio* for the ratio method. Say we increase the range of possible x-values so that it's bounded by zero on the low end and by a very large number, say googolplex, on the high end (pretending, for the moment, that wine and water are infinitely divisible). If we now remain indifferent over x-values, we can predict with *certainty* that the mixture is composed all but entirely of wine. Talk about apriorism! The way to avoid assuming such systematic bias is to remain indifferent over quantities rather than ratios.

A third reason to prefer the symmetry method emerges by comparing it with an alternative method for resolving the wine/water paradox. Given the parity of reasons to accept the two ratio methods, we might be tempted to simply split the difference between them. Why not average the two probability distributions (i.e., remain indifferent between them), and thereby determine a unique solution to the wine/water problem without ignoring either of the two standard answers? Call this method *super-indifference*. The simplest way to carry this out would be to average the two standard answers, which we might call the *naïve method* of super-indifference. We might even expect the naïve method to yield the same answer as the symmetry method, given the parity of the two ratio methods. Unfortunately, the numbers don't cooperate. The average of 5/8 and 15/16 is 25/32, not 5/6. On reflection, however, this is exactly what we should expect. Each of the two ratio methods distributes probability in a way that favors one of the two liquids. If we average the two methods, then we end up with a distribution that favors homogeneous outcomes over heterogeneous ones. In other words, outcomes with lots of wine or lots of water are more probable than outcomes with a roughly balanced mixture. (This point can once again be made into a *reductio*, since remaining super-indifferent over x-values and y-values ranging from zero to a very high number would imply that the liquid is composed either all but entirely of wine or all but entirely of water.) If we want to remain properly super-indifferent, we need to average not only the two ratio methods, but also those methods that favor, to a corresponding degree, heterogeneous outcomes over homogeneous ones. This would yield an answer of 5/6. But why stop there? The number and kind of potential biases is infinite, and we have no reason to believe that one sort of bias is more likely to obtain than another. To be truly super-indifferent, we need to average all of them. Fortunately, the symmetry of the wine/water problem obviates this logistical nightmare. By remaining indifferent over the uniquely invariant measure, quantity, we automatically split the difference between all potential biases. While super-indifference and the symmetry method may be logically equivalent in some cases, the symmetry method is considerably more practicable.

To review: invariance (with respect to frame of reference and scale), parsimony (with respect to assumptions), and simplicity (with respect to calculations) all lend support to the symmetry method. Now, apart from these 'superficial' considerations, are there 'deep' reasons to prefer the approach?

Am I proposing a bold metaphysical thesis by claiming that PI should apply to quantities in the wine/water example, rather than ratios? Well, in my view, pragmatic considerations alone make it clear that we should prefer the symmetry method in the wine/water example. The temptation is strong, however, to ground this solution in a more general view regarding the proper application of PI. I will conclude with some tentative remarks in this direction.

Implicit in my solution to the wine/water problem is a kind of modest realism, which says that there are better and worse ways to partition the space of possibility, even in a state of ignorance. One favorable approach is to construct a model that maps onto the world in a structure-preserving way. Working out the relevant symmetries is then relatively easy. It's also easy to see how we can be misled by models that formalize features of the world in a way that doesn't preserve structure. For example, using the ratio method in the wine/water example distorts the space of possibility in a way that reflects the idiosyncratic behavior of a mathematical entity, rather than the geometric symmetries of the problem. The result is confusion. In general, it seems that Bertrand paradoxes stem from a dubious proposition: that all candidates for the application of PI are created equal. A simple example shows this to be a mistake:

> A standard-looking die is thrown. What is the probability that the square of the number that appears on the die is less than 18.5?[4]

If we remain indifferent over the possible values of the square of the number that appears on the die (as the wording of the question seems to suggest), then the answer is 1/2 (since $1 \leq x \leq 36$). This implies, however, that the probability of rolling 1, 2, 3 or 4 is also 1/2—a dicey proposition, to say the least. Unless we wish to embrace an extreme sort of catholicity, deeming any method of partitioning a given sample space as good as any other, we are forced to conclude that something is seriously wrong with the proposition that all candidates are created equal. Paradoxes of indifference, it seems, are cheap (but illusory) when we apply PI indiscriminately.

This leaves open the question of what distinguishes a good candidate for the application of PI from a poor one. Admittedly, a truly all-purpose criterion is difficult to imagine, but the wine/water problem does demonstrate what I believe to be a promising distinction. The problem shows that the difference between a good candidate for PI and a poor one can boil down to a difference between *primary facts* and *derivative facts*. A primary fact about the mixture of wine and water is that it has a certain composition (or constitution, to use Keynes' term). This fact is determined by a unique state of affairs that obtains when there exists a certain quantity of wine and a certain quantity of water in the liquid. A derivative fact about the mixture is that it satisfies a certain ratio.

[4] This is actually quite similar to one of Bertrand's examples, which evolved into the cube paradox.

A ratio, after all, is just the quotient of two quantities; it's a mathematical convention that we use to represent a relationship between two quantities. The fact that the mixture satisfies a certain ratio derives from, and is determined by, the fact that the mixture contains certain quantities. So quantities have priority; they represent primary facts, while ratios represent derivative facts. Another way to appreciate this point is to observe that ratios supervene on quantities, not vice versa. The supervenience relation holds in two ways. Supervenience$_1$: quantities *determine* ratios (that is, no difference in ratios without a difference in quantities). Supervenience$_2$: ratios obtain *by virtue of the fact* that certain quantities obtain. More generally, derivative facts are determined by and obtain by virtue of primary facts.

Now, why are primary facts better candidates for the application of PI than derivative ones? Is there *a priori* reason to assume reductionism when assigning equiprobability? Clearly, this is not always the case. Oftentimes we have good reason to apply PI to derivative facts. (Calling these facts derivative is perhaps misleading in such cases.) For example, if we know that there's a particular mechanism at work in the wine/water problem, we might have sufficient reason not to remain indifferent over quantities.[5] Also, as I mentioned earlier, some contexts (e.g., biological ones) seem to warrant holistic assumptions as often as reductionistic ones. Even so, it does seem possible to motivate a *ceteris paribus* preference for primary facts. When we lack good reason to focus on derivative facts, primary facts furnish a baseline measure for the assignment of equiprobability, evenly dividing the range of potential biases. Arguably, this puts us in an ideal position to conditionalize on further evidence, and minimizes our distance from the truth. Implicit in this view, once again, is a modest form of realism, which says that our ability to perceive symmetry (and asymmetry) in nature should inform the assignment of probability—although it should by no means blind us to her oftentimes bewildering complexity and caprice.

Acknowledgements

Thanks to Roger White, Greg Mikkelson, Paul Horwich, Peter Graham, Masahiro Yamada and two anonymous referees for comments on earlier drafts.

Department of Philosophy
New York University
New York, NY 10003
jeff.mikkelson@earthlink.net

[5] Say, for example, that a vending machine produces mixtures of wine and water by randomly selecting wine/water ratios between 1/3 and 3. We should remain indifferent over the wine/water ratios. If we don't know whether the machine randomizes over wine/water ratios or water/wine ratios, we can even employ super-indifference, averaging the two distributions. Our epistemic situation is, of course, very different in these cases than in the standard wine/water example.

References

Bertrand, J. [1889]: *Calcul des probabilités*, Paris: Gauthier-Villars.

Carnap, R. [1950]: *Logical Foundations of Probability*, 2nd edition, Chicago: University of Chicago Press.

Gillies, D. [2000]: *Philosophical Theories of Probability*, London/ New York: Routledge.

Jaynes, E. T. [1973]: 'The Well-Posed Problem', *Foundations of Physics*, **3**, pp. 477–92.

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

Kneale, W. [1949]: *Probability and Induction*, Oxford: Clarendon Press.

Oakes, M. [1986]: *Statistical Inference: A Commentary for the Social and Behavioral Sciences*, New York: Wiley.

Poincaré, H. [1912]: *Calcul des probabilités*, Paris: Gauthier-Villars.

Schlesinger, G. N. [1991]: *The Sweep of Probability*, Notre Dame, IN: University of Notre Dame Press.

van Fraassen, B. C. [1989]: *Laws and Symmetry*, Oxford: Clarendon Press.