Paradoxes of Probability

NICHOLAS SHACKEL

Abstract:

Dr N Shackel
James Martin Research Fellow in Theoretical Ethics
Future of Humanity Institute
Faculty of Philosophy & James Martin 21st Century School
University of Oxford
Littlegate House
16/17 St Ebbes Street
Oxford
OX1 1PT
+44 1865 286 279
nicholas.shackel@philosophy.ox.ac.uk

18.12.2007

Introduction

   We call something a paradox if it strikes us as peculiar in a certain way, if it strikes us as something that is not simply nonsense, and yet it poses some difficulty in seeing how it could be sense. When we examine paradoxes more closely we find that for some the peculiarity is relieved and for others it intensifies. Some are peculiar because they jar with how we expect things to go, but the jarring is to do with imprecision and misunderstandings in our thought, failures to appreciate the breadth of possibility consistent with our beliefs. Other paradoxes, however, pose deep problems. Closer examination does not explain them away. Instead, they challenge the coherence of certain conceptual resources and hence challenge the significance of beliefs which deploy those resources. I shall call the former kind weak paradoxes and the latter strong paradoxes. Whether a particular paradox is weak or strong is sometimes a matter of controversy— sometimes it has been realised that what was thought strong is in fact weak, and vice versa—but the distinction between the two kinds is generally thought to be worth drawing.

   The pressure of paradox has often been a spur to intellectual endeavour. Weak paradoxes have on occasion led us to greater clarity and precision in our thought. Strong paradoxes have on occasion led us to radical conceptual innovation, indeed, have been the basis of entire research programmes. Such programmes often bifurcate. On the one hand, various means of evading the paradox are instituted, such as conceptual refinement, restriction or substitution. On the other hand, we continue to think about the paradox and think about what status should be accorded the means that avoids the paradox. One way for a strong paradox to be resolved is for the means of evasion to be shown to be adequate to the issues raised by the paradox. For example, it is at least arguable that the mathematical resources developed by nineteenth century mathematicians are adequate to the conceptual problems in understanding time and space that Zeno's paradoxes raised.

   In this chapter I shall cover both weak and strong probabilistic paradoxes. Before we turn to them I need to mention a point about the nature of probability itself. In philosophy of probability we standardly distinguish subjective or epistemic probability, which is regarded as a feature of persons, from objective probability, which is regarded as a feature of the objective world. Subjective probability may be taken to be a model of the degree to which a person believes something, or a measure of the degree to which they ought to believe something. Objective probability may be taken to be a model of the propensity that the world has to go in a certain way. One way in which they may be held to be related is by Lewis's ((1980)) Principal Principle, which says, roughly, that reasonable subjective probabilities conform to known objective probabilities. By and large, what I say applies to probability as a guide to belief, and so is largely concerned with subjective probability. In some cases the point of taking probability as a guide to belief is a matter of looking at what belief is warranted by the evidence and in others the point is to believe in accordance with the objective probabilities. When a probability in a scenario could be an objective probability, I shall call it a chance.

Weak paradoxes

   Probability is especially rich in weak paradoxes, since (it turns out) we are not good probabilistic thinkers but are rather prone to probabilistic fallacy, and for this reason we can find ourselves surprised by what is probabilistically correct and taken in by what is not. For example, we are prone to confusing the conditional probability of an event E given F with the probability of F given E, and this gives rise to the Xenophobic paradox

(see Clark (2002)), the Prosecutor's fallacy (representing the probability of the evidence given innocence as if it were the probability of innocence given the evidence) and the Medic's fallacy (confusing the reliability of a test with the chance of illness/health given a positive/negative test). We are also prone to ignoring prior probabilities and base rates, confusing probability with representativeness, and subject to framing effects with probabilistic information. Examples of these errors are addressed elsewhere in this book and given the pressure on space I shall mention only a few weak paradoxes before devoting most time to strong paradoxes.

Failure to appreciate how aggregation can lead to misleading proportional information.

Despite the conviction of generations of the innumerate, $\frac{a}{b} + \frac{c}{d}$ is not equal to $\frac{a+b}{c+d}$. A consequence of this is that aggregated proportional information such as percentages can be significantly misleading.

Simpson's paradox

We have a new treatment for a disease and when we compare it with the old treatment it cures 2% more people.[1] Surely that means it is a better treatment? Not necessarily. When we analyse the results by sex we find that the new treatment cures 18% fewer men and 17% fewer women.

Treatment	Men		Women			Cured percentages			
	Ill	Cured	Ill	Cured		Men	Women		Overall
Old	1000	500	100	80		50%	80%		53%
New	250	80	800	500		32%	63%		55%

It would be reasonable to conclude that the aggregated proportions produce an illusion that the new treatment is better. This is certainly surprising, but it is a simple consequence of the mathematics of means and what results when you aggregate or disaggregate results. This kind of problem arises more easily when there is a wide disparity in the numbers of the two groups involved, say 1000 women versus 100 men, but does not require such a disparity. In this example the numbers are comparable, 1100 versus 1050. There is no guarding against this problem and it has nothing to do with sample sizes in general nor with problems to do with base rates. Its implication is that aggregation of groups with relevant differences may be dangerously misleading. Here is a topical example.

It is possible for every department in a university to massively discriminate against group B in admissions yet when looking at the figures for the university as a whole it may look as if the university discriminates against group A. Consider a university which has only two departments and applications and acceptances as laid out in the following table:

	Psychology		Mathematics			Acceptance proportions			
	Applications	Accepted	Applications	Accepted		Psych %	Maths %		Overall
Group A	1000	500	100	80		50%	80%		53%
Group B	250	80	800	500		32%	63%		55%

Group A is massively preferred to Group *B* in both psychology and mathematics, and yet on the overall figures it looks as if Group *B* is being favoured. The conclusion should

---

[1] I say 'cure' on the basis of 2% more recovering. We assume here that the circumstances are such as to allow correlation licensing the inference to causation.

be that aggregation can result in dangerously misleading data, especially when attempting to use statistical proportions as proof of discrimination.

The illusions we have just analysed arise because it is mistakenly assumed that aggregation over a conditioning variable (sex or subject) is irrelevant to determining the significant correlations between the input variable (treatment or group) and output variable (response to treatment or admission status). Historically these are the kinds of illusions into which we have fallen. Might there be cases in which *disaggregation* is similarly misleading? Considered purely mathematically, one may take *any* arbitrary variables as input, output and conditioning. For example, we could take feeling as a conditioning variable in the treatment case.

Treatment	Feel worse		Feel better			Cured percentages			
	Ill	Cured	Ill	Cured		Feel worse	Feel better		Overall
Old	1000	500	100	80		50%	80%		53%
New	250	80	800	500		32%	63%		55%

Here it would seem implausible to conclude that the old treatment is in fact better, and so here disaggregation is misleading.

How then to determine which disaggregations are required and which misleading? On the one hand, its worth noting that constructing examples of Simpson's paradox in which disaggregation is clearly misleading tend to depend on cases in which the conditioning variable is more plausibly seen as part of the output. It being part of the output tends to make the figures look gerrymandered: could there really be 500 people who were both cured *and* felt worse? What we would really like, though, is to have criteria for correct aggregation and disaggregation. Since the general project is to find the true reasons for the variability in the output, conditioning variables are properly disaggregated into those that could be among such reasons. To be such a reason is presumably to be something on which the output depends rather than vice versa. The debate takes off from here on the basis of interpreting dependence in evidential terms or in causal terms.

A little knowledge of probability theory is a dangerous thing

Monty Hall

When you were young you probably learnt the classical basis for assigning numerical probabilities: to give equal probability to the equally possible. That is why you think the chance of getting heads is ½, the chance of throwing a six is 1/6 and the chance of drawing the ace of spades is 1/52. You are on TV taking part in Monty Hall's game show and have just answered the final question correctly. You are now eligible for the big prize. There are three doors in front of you and Monty tells you that behind one is a car and behind the other two are goats. He invites you to pick a door. As he always does at this point in the show, Monty then opens one of the other doors, shows you a goat behind it and asks you whether you want to change the door you picked. Can you improve your chance of winning by changing? 'Surely not', you think, 'since there are two closed doors, a car behind one of them, it is equally possible for them to be behind either, so the chance that it is behind mine is the same as the chance it is behind his, namely, 50%'.

Wrong answer, but you are in good company. When this was published by Marilyn vos Savant[2] a number of mathematicians insisted that it was the right answer, and for the very same reason. If Monty chose his door at random you would be right, but he doesn't.

The quick way to see why it is wrong is to remember that when you first picked your door there was 1/3 chance it was behind your door and 2/3 chance it was behind Monty's doors. That hasn't changed just because Monty opened one of his two doors, since when you first chose you knew that Monty would open a door and show you a goat. All that has changed is that if you were wrong in the first place the car must now be behind the door Monty didn't open. So there is a 2/3 chance that the car is behind Monty's other door. What confuses us here is that we don't take into account that whenever the car is behind one of his doors Monty doesn't have a free choice of doors to open. He can only open the one with the goat behind it. We fail to realise that we are not equally ignorant about which door it is behind. If that doesn't convince you, think of the case in which there are one hundred doors, you pick one, Monty opens 98 doors showing a goat in each case. Still sure you don't want to swap?

The proof is in conditional probability. You need to know the probability that the car is behind your door given that you see a goat behind one of his, $P(Y|G)$.

$$P(Y|G) = \frac{P(G|Y)P(Y)}{P(G)} = \frac{1 \times \frac{1}{3}}{1} = \frac{1}{3}$$

But if Monty chooses his door to open at random, then instead of the probability of seeing a goat, $P(G)$, being 1, it drops to 2/3, and then $P(Y|G)=1/2$.

The significance of the distinction between numerical identity and qualitative identity

Bertrand's Box

There are three boxes each with two compartments. In one box there are two gold coins, in another two silver coins and in the third one gold and one silver coin. You open one compartment and see a gold coin. What is the chance that the other coin in the box is silver? 'Well', you may think, 'the other coin is either silver or gold and they are equally possible so it must be 1/2'. Alternatively, you may think that since you have seen a gold coin the box is either the box with two gold coins or one gold and one silver, but we don't know which so it must be 1/2. But that is incorrect.

We must distinguish two kinds of identity, qualitative identity and numerical identity. If we say that Jack and John are the same age we mean that there is a property that they have in common, their age. This is qualitative identity, because it is a matter of the identity of a property rather than of an object. But if we say that Jack and John are the same person we don't mean that Jack and John are distinct objects who share the property of personhood (if we meant that we'*d* say that they are *both* persons), but that Jack and John are one and the same person. This is numerical identity and it is a matter of the identity of an object.

Returning to the coins, it is true that so far as qualitative identity goes, there are only two distinguishable options for the other coin, namely, gold or silver. But the possibilities we must distinguish are distinguished in terms of numerical identity. There are in fact

[2] '…rose to fame through her listing in the *Guinness Book of World Records* under "Highest IQ". Since 1986 she has written Ask Marilyn, a Sunday column in *Parade* magazine in which she answers questions from readers on a variety of subjects.' http://en.wikipedia.org/wiki/Marilyn_vos_Savant.

three different gold coins that you might have revealed on opening the compartment. Only one of those coins is paired with a silver coin, the other two are paired with each other. So the chance is 1/3.

Bose-Einstein paradox

Suppose the boxes have not gold and silver coins but two kinds of bosons, calls them yellow and blue. You open a compartment and see a yellow. What is the chance that the other particle in the box is a blue? Our earlier reasoning would imply that the answer is 1/3, but astonishingly, both physical theory and empirical investigation show it to be ½! Why is that? The physicists say bosons (and fermions) are *indistinguishable* particles, by which they seem to mean that they lack numerical identity. If that is the case, the earlier argument we gave based on numerical identity lapses and instead we can reason only on the basis of qualitative identity and distinction. Since the other particle is either yellow or blue the chance is ½.

The idea of particles lacking numerical identity is very difficult to understand. It might be that seeing a yellow is merely acquiring the information that *a* particle is yellow, so the possibilities consistent with that information are both yellow or one yellow and one blue, hence the chance is ½. However, that can't be the whole story. In the case of indistinguishable particles, whilst they are countable, the claim is not the epistemological claim that what we know fails to distinguish them but the metaphysical claim that there is no fact of the matter about whether this particle is the same particle as that particle. That is a deeply puzzling claim, but we shall leave its further investigation to the philosophers of physics!

Strong paradoxes

So far we have looked at weak paradoxes, paradoxes that highlight our weaknesses in understanding probability. We now turn to strong paradoxes, paradoxes that pose challenges to probability itself, either by apparently falsifying principles or axioms of probability which we have independent reasons to think true or by threatening our confidence in the coherence and comprehensiveness of probability theory. We examine them in two areas, probability as a guide to belief and rational decision theory as a guide to action, given by defining choiceworthiness in terms of expected value. In the case of weak paradoxes I was able to suggest the root of the problem. In the case of the strong paradoxes I can only indicate the kind of proposals that have been offered as solutions.

Trouble for belief

Bertrand's chord

Choose a chord of a circle at random. What is the chance that it is longer than a side of the inscribed equilateral triangle?
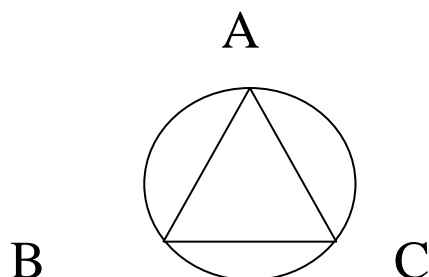
1) Consider chords all the chords that start at A on the circumference. Any chord whose other end is on the circumference between A and *B* or A and *C* will be shorter whilst any chord whose other end is on the circumference between *B* and *C* will be longer. The angle subtended by the set of longer ones is therefore 60° and hence the chance of these being longer is 60/180 = 1/3. By symmetry this applies to all chords, so the chance of being longer is 1/3.

2) Now consider the chords with centres on the radius bisecting BC. They are perpendicular to the radius. Those whose centre is on the same side of BC as the centre are longer and those on the other side are shorter. The distance from the centre to BC equals the distance along the radius from BC to the circumference. Therefore the chance of these chords being longer is ½ and by symmetry this applies to all chords.

3) Now consider all the chords whose centre lies within the circle inscribed in the equilateral triangle. All these chords are longer and the chords whose centre lies outside the inscribed circle are shorter. The area of the inscribed circle is ¼ that of the circumscribing circle, therefore the chance of a longer chord is ¼. [Publisher will need to produce diagrams in accordance with these. I do not have the requisite software since Office draw is imprecise.]

Hence the chance of a longer chord is 1/3 and ½ and ¼. But probabilities are unique, so this is a contradiction.

To cut a long story very short, given a range of possible outcomes, mathematical probability theory *alone* does not give numerical probabilities for those possibilities.[3] What probability theory will do, given numerical probabilities for what can be regarded as, in some sense, the atomic possibilities, is tell you what the numerical probabilities for all the compound possibilities are. The right basis on which to assign probabilities to the atomic possibilities is a controversial issue in the philosophy of probability. One position in that controversy is called the Principle of Indifference,[4] which says that possibilities of which we have equal ignorance have equal probability. For example, given a shuffled pack of cards I am equally ignorant with respect to the 52 possibilities for the top card and so should assign the probability of 1/52 to each possibility. Supposing that I know the top card is red, then I am now no longer equally ignorant over all the possibilities, but am equally ignorant over the red cards, so I assign probability of 0 to the top card being black and 1/26 to each of the red cards.

Bertrand designed his paradox as a refutation of the applicability of probability to infinite sets of possibilities. In each of the calculations above we made implicit use of the Principle of Indifference. For example, in the second case we take it that we are equally ignorant with respect to the distance of the centre of the chord from the centre of the triangle and so apply a uniform distribution to that random variable. His argument is roughly that numerical probabilities can only be got by use of the Principle of Indifference, but there is no unique way to apply that principle to infinite sets, therefore

---

[3] Part of the long story cut short is the distinction that philosophy of probability makes between the mathematical *theory* of probability and *probability*. The mathematical theory is a part of measure theory and is not, as such, about probability, properly so-called, until we have interpreted it as a model of degrees of belief or propensities. Compare the distinction between mechanics as a piece of mathematics and interpreted in terms of particles, motions and forces. Bertrand was one of the originators of measure theory, and one of the points he hoped to make with this paradox was that the mathematical probability theory in its full generality lacks interpretation as probability properly so-called.

[4] So named by Keynes (1921/1963). J Bernouilli and Leibniz called an essentially similar principle the Principle of Insufficient Reason.

probability does not apply to infinite sets. It has recently turned out that there is reason to reject the third calculation (see Shackel (Forthcoming)) but that is of no help. The problem is that the three cases are merely examples of the infinitely many ways there are of applying the Principle of Indifference to calculating the probability of a longer chord.

This paradox has sometimes been thought to be resolved (e.g. Jaynes (1973); Marinoff (1994)). Certainly there are some empirical cases in which a particular way of calculating the probability of a longer chord both fits the features of the case and gets the right empirical answer. But the problem posed by Bertrand is quite general. Arguably, (see Shackel (Forthcoming)) the current main contenders for resolution do not work and there are good reasons for thinking that it is irresolvable. We may not wish to join Bertrand in his finitism, but whilst his paradox is unresolved it threatens our confidence in the coherence of applying probability to infinite sets.

Sleeping Beauty

It is Sunday night and you, sleeping beauty, go to sleep knowing the following. We will toss a fair coin. If it lands heads we will wake you briefly on Monday and put you back to sleep with a drug which will erase your memory of that waking and you won't wake till Wednesday. If it lands tails we will wake you briefly on both Monday and Tuesday, putting you back to sleep with the same drug. Before you go to sleep on Sunday you think the probability of the coin landing heads, $P(H)$, is ½. We wake you on Monday. You don't know what day it is. What now is the probability that the coin landed heads?

1. It must be ½, since it was a fair coin, that was your opinion of Sunday night, and you have learnt nothing new (since you knew when you went to sleep that you would wake at least once not knowing which day it was).

2. It must be 1/3. This is either a Monday waking following a head (HM) or a Monday waking following a tail (TM) or a Tuesday waking following a tail (TT). These possibilities are indistinguishable to you so equiprobable. Furthermore, by the law of large numbers, were this experiment repeated many times the proportion of wakings when the coin fell heads tends to 1/3.

What we need to know here is a conditional probability, namely, the probability that the coin is heads given that you woke, $P(H|W)$. We can prove that this probability is the same as the probability that it fell heads and today is Monday, $P(HM)$. Since we know that this waking is either HM or TM or TT and that these three events are mutually exclusive and jointly exhaustive, we know

$$P(\text{HM or TM or TT}) = P(\text{HM}) + P(\text{TM}) + P(\text{TT}) = 1$$

So if we can determine some relations between $P(\text{HM})$, $P(\text{TM})$, $P(\text{TT})$ we can determine $P(\text{HM})$. Elga ((2000)) argues that the Principle of Indifference gives us that $P(\text{TM}|\text{TM or TT}) = P(\text{TT}|\text{TM or TT})$ and the latter implies that $P(\text{TM}) = P(\text{TT})$. We also know that

$$P(H|M) = P(\text{HM})/(P(\text{HM}) + P(\text{TM}))$$

$$P(\text{T}|M) = P(\text{TM})/(P(\text{HM}) + P(\text{TM})) = 1 - P(H|M)$$

So far, this is compatible with either answer. Elga now argues that $P(H|M) = P(T|M)$, and this together with the last two equations means that $P(\text{HM}) = P(\text{TM})$. But since $P(\text{TM}) = P(\text{TT})$ and the three together add up to one, $P(\text{HM}) = 1/3$.

The basis of Elga's argument that $P(H|M) = P(T|M)$ is that it doesn't really matter when we toss the coin. We could just as well toss it after the Monday wakening. In that case you

would agree that $P(H|M)= P(H) = ½$ and hence $P(T|M) = 1- P(H|M) = ½$.[5] So knowing that it is Monday increases the probability of heads by 1/6 ($P(H|M)-P(H|W)= P(H|M)-P(H$M)=1/2-1/3=1/6).

Lewis ((2001)) rejects this argument on the ground that knowing that you are awake in one of three indistinguishable wakenings is not *relevant* evidence to the question of heads, and so $P(H|W)$ must equal $P(H)= ½$.Hence, although he agrees that knowing it is Monday increases the probability of heads by 1/6, so agrees that $P(H|M)= P(H|$W) +1/6, he thinks that $P(H|M) = 2/3$ and therefore is not equal to $P(T|M)$.

So we have two plausible thoughts which seem compatible yet which result in a contradiction. On the one hand, it seems that your ignorance means it doesn't matter when the coin is tossed and hence knowing it is Monday and the coin is yet to be tossed makes $P(H|M)= P(H)=1/2$ plausible. On the other hand knowing that you are awake in one of three indistinguishable wakenings doesn't seem relevant evidence for how the coin landed. The first implies that $P(H|W)$ is 1/3 and the second that it is ½.

Lewis discusses Elga's application of Lewis's Principal Principle to future chances, saying that applications to future chance events must satisfy a proviso that doesn't apply in the case to which Elga applies it, the case where you know it is Monday and the coin has yet to be tossed. Interestingly enough, Lewis's position here might be thought to be at odds with his very own Principal Principle. One notion of objective probability is limiting frequency, and on that basis the objective probability of a head given you woke is 1/3. According to his principle, that means that your subjective probability ought also to be 1/3. And yet Lewis is saying the subjective probability is ½.

This paradox bears an analogy to Monty Hall and perhaps also to Bertrand's Box (compare the Tuesday wakings to two goats or the two different gold coins in the same box). What makes it importantly different is that in those paradoxes there are not plausible arguments to be given on both sides; rather, we can explain why one of the two claims about probabilities is erroneous. In sleeping beauty, both sides have put forward plausible arguments and neither side has shown the other side's arguments to be decisively flawed. Although most discussion so far has inclined towards Elga's position, Lewis has a substantial band of defenders. The paradox is still a developing controversy and recently authors have even put forward arguments for probabilities between 1/3 and 1/2. Further recent literature: Dorr (2002); Horgan (2004); White (2006).

Doomsday argument

As far as we can tell, even if the life of the universe is infinite, there is a finite (if very large) amount of time before all life will become impossible, and that means a finite amount of time for life to continue. There are only finitely many humans in existence. Therefore the total number of humans there will ever be is finite. Is the end of humanity near or far?

We estimate that there have been 60 billion humans so far and there are millions of years in which humans might well flourish. Consider two hypotheses:

Few: the total number of humans will be 100 billion.

Many: the total number of humans will be a million billion

There is nothing special about you and so you should consider yourself a typical human. But if Many is true then you are a very untypical human. Relatively speaking, to be

---

[5] This amounts to an application of Lewis's Principal Principle to future chances: 'credences about future chance events should equal the known chances' (Lewis (2001):175).

roughly the 60 billionth human is to be very early in the whole history of mankind if Many is true. So Many is probably false. (Cf. Leslie (1996))

That seems a bit quick. Can it be right to conclude thus only on the basis of your numerical place in the birth order of humans? Well, consider an analogous argument which seems correct. Suppose you had two vases in front of you, one containing a million numbered balls and one containing only 10 numbered balls. You pick a ball out at random and it has the number 7 on it. It is very unlikely you would have got such a low number from the first vase, so it looks like you picked from the vase containing only 10 balls.

We can firm up the Doomsday argument reasoning with some probability calculus. To keep the maths simple we'll assume that Few and Many are the only possibilities. Prior to taking into account your birth order, but given only the information about the millions of years in which humans might flourish, you might reasonably estimate $P(F) = 5\%$ and $P(M)=95\%$. Now we consider the Evidence: that you exist and are roughly the 60 billionth human. Gott ((1993)) proposes the Copernican anthropic principle: that you should take yourself to be a random sample from the set of all intelligent observers (which so far as we know means a random sample from all humans) and that it is equally likely for you to be any one of those observers. Applying that principle means the conditional probabilities of you being the 60 billionth human are $P(E|F) = 1$ in 100 billion, and $P(E|M)= 1$ in a million billion. Then

$$P(F \mid E) = \frac{P(E \mid F)P(F)}{P(E)} = \frac{10^{-11} \times 0.05}{P(E)} = \frac{5 \times 10^{-13}}{P(E)}$$

$$P(M \mid E) = \frac{P(E \mid M)P(M)}{P(E)} = \frac{10^{-15} \times 0.95}{P(E)} = \frac{9.5 \times 10^{-16}}{P(E)}$$

$$\frac{P(F \mid E)}{P(M \mid E)} = \frac{5 \times 10^{-13}}{9.5 \times 10^{-16}} \approx \frac{10^3}{2}$$

So the probability of Few given the evidence of your place in the birth order is roughly 500 times the probability of Many. If you thought I was unreasonably optimistic in setting $P(M)$ at 95% and think Few and Many should start as equally likely then the probability of Few given your place in the birth order is 10,000 times as likely!

There are many versions of the Doomsday argument in addition to these. For example, Gott's ((1993)) uses his Copernican principle to work out the probability that the total number of humans born will be less than 20 times the number already born is greater than 95%.

The Doomsday argument has received much attention and there are numerous conflicting attempts at refuting it. One interesting line proposes that your existence as an observer makes probable there being many observers in the world history, and this increased likelihood undermines the Doomsday argument (see Dieks (1992)). What is interesting about this line is that reasoning on the basis of one's own existence is used in two different ways, and it draws attention to the very feature which many people find fishy about the argument.

The Doomsday argument makes use of anthropic reasoning, reasoning which takes as a premiss one's own existence as an intelligent reasoner capable of making observations. There are many other uses of such reasoning, for example, what are called the fine tuning

arguments[6] make frequent use of the premiss that if a theory implies that the existence of such reasoners is very unlikely then that would seem to count against the theory.[7]

The status of anthropic reasoning is controversial. As Bostrom has shown, a serious problem for anthropic reasoning is its vulnerability to what he calls observation selection effects, for example:

> How big is the smallest fish in the pond? You catch one hundred fishes, all of which are greater than six inches. Does this evidence support the hypothesis that no fish in the pond is much less than six inches long? Not if your net can't catch smaller fish. (Bostrom (2002):1)

Bostrom proposes that we need a comprehensive theory of observation selection effects if we are to use anthropic reasoning without falling foul of various subtle fallacies. In his view the Doomsday argument is a central case in illuminating the difficulties here, and this is the explanation for the very extensive disagreements about how best to formulate it and what might be right or wrong about it. It is one of the cases which he thinks should drive us to a principle of anthropic reasoning which he calls the Strong Self-Sampling Assumption

> (SSSA) Every observer at every moment should reason as if their present observer-moment were randomly sampled from the set of all observer-moments. (Bostrom (2002):162)

The Doomsday argument is a good example of the nagging power of a strong paradox. The premisses seem reasonable and the steps in the argument appeal to principles which in other areas we think unobjectionable. By the rational principle that requires us to follow where an argument leads we ought to accept the conclusion. Yet the argument takes us far further than we think reasonable. When we try to settle what has gone wrong, we cannot do so in a satisfactory manner. We can dispute the premisses or the steps in the argument, but the weaknesses we find are not severe enough to resolve the matter.

A good review of the literature on the Doomsday argument is Bostrom (1998). The Doomsday argument's wider significance as a paradox is grounded in the general controversy about the status of probabilistic anthropic reasoning, most recently, its uses by proponents of intelligent design such as Dembski (1998).

Trouble for action

St Petersburg

In gambling a fair price for a bet is regarded as the expected return on the bet. So if you stand to win £4 on the cut of a card and bet on hearts, the fair price is $1/4 \times 4 = £1$. A bet cheaper than its fair price is a good bet. Casinos make very large amounts of money from taking all good bets that are only a few percent cheaper than their fair price. But let's be cautious. Perhaps as individuals we shouldn't take all good bets, but only all very good bets, say all bets that are at least 50% cheaper than their fair price.

A coin is going to be tossed until it lands head. If it takes one toss you will be paid £2, two tosses £4, and in general, n tosses will pay £2^n. What is the fair price?

The expected return is the sum of the products of the probability of the number of heads with the winnings on that number. The probability of the first head being on the first toss is ½, on the second it is the probability of getting first a tail and then a head, which is

---

[6] Fine tuning is there being only very narrow ranges within which various physical constants must lie if the possibility of life is to be permitted by the laws of nature.
[7] But see Sober (2002) for a rejection of the validity of such reasoning.

½×½, and in general the probability of the first head being on the nth toss is $\frac{1}{2}^{n}$. The expected return is an infinite sum.

Expected return = $1/2\times2+1/4\times4 +\ldots+\frac{1}{2}^{n}.\times 2^{n} + \ldots$ and so on

$= 1 + 1 + \ldots +1+\ldots$ and so on.

We say that this sum tends to infinity because for any finite number, add up enough terms of this infinite sum and we can exceed it.

Of course, there is the practicality of there being an upper bound on the amount of goods on the world and therefore an upper bound on the amount of money. But consider that the universe is unbounded and that were we part of an intergalactic civilization, there might really be an infinite amount of goods in the universe. Or just consider the matter as a theoretical problem. In principle, the expected return on this game is infinite and consequently, any finite price is massively cheaper than the fair price. So on the principle of taking all very good bets, you should play the game for any finite amount. Your entire savings is a finite amount so you should take the bet if offered it at that price. But that's mad, isn't it?

Nicholas Bernoulli posed this problem in 1713 and his cousin Daniel offered a solution. The starting point is two thoughts: that what matters about wealth, its value, is its usefulness to us; and that an extra £1 to a millionaire is not as useful as an extra £1 to a tramp. In general, the usefulness of your first £100 is greater than of your second £100 and so on. We define the measurement of usefulness to be utility. What we need in order to properly assess the value of an amount of money is to know the utility of that amount of money. In general, what we need to know is the utility function of money. It is perhaps misleading to speak of *the* utility function, since we can make inferences about a person's utility function from his behaviour and we find that people have differing utility functions which reflect their differing attitudes to risk. But for the sake of this argument, we will consider only utility functions that respect the point about decreasing usefulness, which technically put, amounts to the marginal utility of wealth decreasing as wealth increases. What this means is that the gradient of the utility function decreases, and looks something like this:
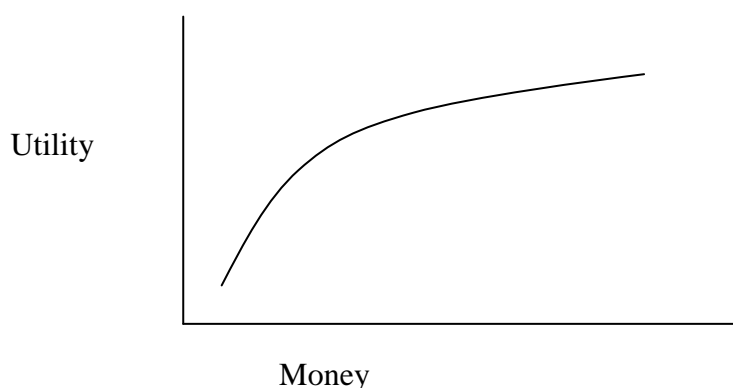
A logarithmic utility function looks roughly like this. Suppose, for the sake of illustration, we take our utility function to be U(money) = the logarithm to base two of money.[8] This would give the utility of £2 to be 1 utile,[9] the utility of £4 to be 2 utiles, and in general the utility of £2^{n} to be n utiles. When we now work out the expected return not in term of money but in terms of utility we get

---

[8] $U(m) = \log_2 m$.
[9] We define utiles to be the units of utility.

Expected utility $= 1/2 \times U(2) + 1/4 \times U(4) + \ldots + \tfrac{1}{2}^n \times U(2^n) + \ldots$ and so on

$$= 1/2 \times 1 + 1/4 \times 2 + \ldots + \tfrac{1}{2}^n \times n \ + \ldots \text{ and so on}$$

$$= 2.$$

We say that this sum equals 2 because (roughly) adding up more and more terms of this infinite sum gets us closer and closer to 2.

Now 2 utiles is probably too cheap, but that is because we chose a utility function that would make the mathematics easy to illustrate the point. The critical point is that the decreasing utility of wealth means that the value of the bet is finite. On another utility function[10] the bet is worth 2884 utiles which is £2895.

Whether the solution succeeds can be questioned. The notion of utility is certainly correct, but then you might be offered the opportunity of playing St Petersburg in terms of utiles rather than money, when once again it seems that you should bet your entire savings on the game. This doesn't seem rational. If that is right, then the claim that expected utility determines choiceworthiness in the case of betting is weakened. But what other basis for the rational choiceworthiness of *bets* can there be? In this way the St Petersburg Paradox continues to discomfort us, eroding our confidence in the applicability of probability theory exactly where it would seem to be unquestionably applicable in guiding action. Hence it remains a strong paradox. For more extensive discussion see Jeffrey (1990):150 ff..

Allais' paradox

You are offered a choice of

A: certainty of £1,000 or

B: 8% chance at £5,000, 91% chance at £1,000, 1% chance of nothing.

You are offered a choice of

C: 9% chance of £1,000, 91% chance of nothing

or D: 8% chance at £5,000, 92% chance of nothing.

Most people prefer A to *B* and prefer *D* to C. But choosing A over *B* and *D* over *C* is not consistent with determining choiceworthiness by expected utility. Let U be our utility of money function, with U(0)=0. Preferring A over *B* on the basis of expected utilities gives

EU(A)>EU(*B*)

i.e. U(1,000)> .08*U(5,000) +.91*U(1,000)

i.e. 0.09*U(1,000)> .08*U(5,000)

i.e. EU(*C*)> EU(*D*)

So if choiceworthiness is given by choosing to maximise expected utility, then if you choose A over *B* you ought to choose *C* over D.

Allais ((1953)) produced his paradox in order to embarrass the Independence Axiom of von Neumann-Morgenstern ((1944)) decision theory: for any lotteries x, y, z,[11] and for any $p \in [0,1]$, you prefer x to y iff you prefer $px + (1-p)z$ to $py + (1-p)z$. The Independence Axiom implies that preference doesn't change if you supplement both sides of a choice

---

[10] $U(m) = m^{0.9995}$

[11] A lottery is a probability function on a set of outcomes. So the lottery B is $\{P(0) = 0.01, P(1,000) = 0.91, P(5,000) = 0.08\}$.

with the same further benefit. Setting out our scenario by decomposing A makes this evident:

A		B		C		D	
Payoff	Chance	Payoff	Chance	Payoff	Chance	Payoff	Chance
1000	0.09	5000	0.08	1000	0.09	5000	0.08
1000	0.91	1000	0.91				

Examining this table makes it clear that the choice between A and *B* is the choice between *C* and *D* supplemented with the same opportunity in each case, namely, a 91% chance of winning £1000. So independence implies that we will choose A over *B* iff we choose *C* over D.[12]

Savage's ((1972)) sure thing principle implies the same result: if choosing *x* or *y* produces the same result in circumstances consistent with *Q* then the choice between them should depend only on the consequence of circumstances consistent with ¬Q. So we suppose here that the 91% chance of £1000 in choices A and *B* are circumstances consistent with Q. Then only the other consequences should determine which way we choose between A and B. But the other consequences are the same consequences we are choosing between when choosing between *C* and D. Hence the sure thing principle implies that we will choose A over *B* iff we choose *C* over D.

We all feel a pressure to choose A over *B* and *D* over *C*, and it is rumoured that even Savage chose this way when first presented with this paradox. If it were clear that this tendency is irrational then the Allais paradox would amount to an illusion of choice under uncertainty. But it is controversial whether it is irrational. On the one hand, considered over many decisions for moderate amounts as above, choosing in accordance with A over *B* does significantly worse, on average £310 worse per decision. That sounds a bad policy and is arguably irrational. On the other hand, if we make it a single decision for an amount that is life changing, such as A being certainty of £100 million, one might think that it was a bad policy to risk getting nothing by choosing *B* over A. Furthermore, this is not evaded by the decreasing marginal utility of money, since one can pose the whole problem in utilities instead. The question is rather, given certainty of a great benefit, is it worth taking a small risk of having nothing for the sake of a greater, perhaps even enormous, benefit? The independence axioms and Savage's sure thing principle can commit us to saying yes, but is that really right? It depends on what our attitudes to risk should be, and whether facts such as magnitude of reward, and frequency of opportunity to risk something for the reward, influence what those attitudes should be.

Allais himself argued that his paradox shows that choiceworthiness is not expected utility but is rather a function of both expected utility and the variance of utility. Variance of outcome is sometimes regarded as a measure of risk. So we might understand him as seeking to make some allowance for risk in assessing choiceworthiness. But the example of the Allais paradox raises the suspicion that analogous problems can be posed to any attempt at characterising choiceworthiness in these formal terms. Prima facie, for any formal specification one can always gerrymander an example in which the certainty of a big enough bird in the hand intuitively outweighs the risk of letting it go for a chance at the many in the bush. The wider significance of Allais's paradox may be that it leads us to develop arguments for the proposition that choiceworthiness is not scale-free, and its scale dependence cannot be represented by the formal apparatus of standard decision theory—a

---

[12] This way of presenting the paradox is sometimes known as the common consequence effect. See Kahneman and Tversky (1979)

proposition which is paradoxical for standard decision theory. For extensive discussion see the edited collection of papers Allais and Hagen (1979).

Newcomb's problem

There is an opaque and a transparent box in front of you. You can see £1000 in the transparent box. You know that if a reliable predictor has predicted that you will open only the opaque box he will have put £1,000,000 in that box. Otherwise he will leave it empty. You can either open both boxes or just the opaque box and you get to keep whatever is in the boxes you open. What should you do?

What is the expected value of opening one box and opening two boxes? Suppose the probability that the predictor is right is p. Then the expected value of opening both boxes

E(two boxes)= $1000p + 1001000(1\text{-}p)$=$1001000\text{-}1000000p$ (=251000 if $p = $ ¾).

The expected value of opening just the opaque box

E(one box) = $1000000p + (1\text{-}p)\times0 = 1000000p$ (=750000 if $p = $ ¾)

Comparing the two

E(two boxes)<E(one box) iff $1000/1999$<p

So as long as the chance that the predictor is right is a little bit more than ½, then maximising expected value means you should open only the opaque box.

Now consider what is called the dominance principle, which says that if one action is better than another in each of the possible circumstances that might obtain, then you ought to choose that action. Here, whether or not the predictor has put £1,000,000 in the opaque box, you will be £1000 better off taking two boxes than one. So the dominance principle says open both boxes.

This paradox has been very fruitful in the development of decision theory. The challenge it poses is twofold: on the one hand to explain which answer is correct and on the other to explain what is wrong with the other answer.

Causal decision theorists generally think that two boxing is the right answer. When deciding what to do you should focus on the causal powers of your actions, not their evidential aspect. For example, suppose that both smoking and cancer were caused by a gene. In that case, the correlation of smoking with cancer would not be because smoking caused cancer. Rather, the correlation would arise because smoking is evidence that you have the cancer causing gene. But because of the correlation, smokers would still have a higher probability of getting cancer than non-smokers. Consequently, calculating the expected utility of smoking would make it look like it was a bad thing to do despite the fact it had no causal impact on your getting cancer or not.

Likewise, say causal decision theorists, what is relevant about your choice in Newcomb's problem is the expected benefit of an act as cause of benefits, not what evidence your choice is for what the predictor did. In this case, nothing you do now can change what the predictor has already done. Consequently two boxing is the right answer.

What is wrong with the calculations of expected value of each act (two boxing or one boxing) is that it uses probabilities conditional on the act. When your act has evidential significance in addition to its causal significance, then the conditional probability will be different from the absolute probability in part because of that evidential significance. So in such cases calculating the expected value of an act using conditional probabilities rather than absolute probabilities will amount to tainting what you want — numerical information about expected causal benefit of that act — with quantities which arise out of the irrelevant evidential significance of that act. Hence the proposal of causal decision

theory is to reform standard decision theory by the use of absolute rather than conditional probabilities in calculating expected values, but absolute probabilities based on dependency hypotheses about causal efficacy.[13]

Evidential decision theorists disagree but in two different ways. There are evidential decision theorists who are two-boxers and who propose their own adjustment to standard decision theory to avoid the expected value calculation recommending one-boxing. There are others who think that standard decision theory is right. They may criticise the dominance argument on the ground that dominance reasoning assumes the truth of the general principle of acting so as to maximise value. Consequently the dominance principle is a *subsidiary* principle to the general principle. In Newcomb's problem, the application of the dominance principle results in transgressing the general principle, since dominance reasoners don't get rich but one-boxers do (remember, the predictor is *reliable*). Consequently, because it is subsidiary, the dominance principle must give way to the more general principle.

For an edited collection on the paradox with extensive bibliography see Campbell and Snowden (1985). For a lucid exposition of causal decision theory covering his own and other's see Lewis (1981a), and for an interesting argument that prisoner's dilemma is a kind of Newcomb's problem see Lewis (1981b). For an evidential decision theory compatible with two-boxing see Price (1986); Jeffrey (1990). For some recent argument in favour of one-boxing see Blackburn (2000):189.

Two-envelope paradox

There are two envelopes, one of which has twice the amount of money in it as the other. You take one at random and I take the other. I ask you whether you'd like to swap. Eager to apply your new found knowledge of probability theory, you decide that the way to decide is to work out the expected value of my envelope. If it is higher than yours you'll decide to swap, if lower, not, and if the same you won't care. So, let the amount in your own envelope be x; then the amount in mine is either $\frac{1}{2}x$ or $2x$, and they are equally likely. So the expected value of my envelope is

$$\text{Expected value} = \frac{1}{2} \times \frac{1}{2}x + \frac{1}{2} \times 2x = 5/4\, x > x$$

So the expected value of my envelope is greater than the value of your envelope so you should swap. But just before you do, you decide to check the expected value of your own envelope. So you reason that the amount in my envelope is y, and then proceed as before, finding that the expected value of your envelope is $5/4\, y$, so you conclude that expected value of your envelope is greater than mine. So my envelope is worth more than yours and my envelope is worth less than yours. That can't be right!

It is important to be clear of the precise nature of the problem here. It is not that there is any problem in knowing what to do. Quite obviously, you should be indifferent between your and my envelope. The problem is that an apparently correct application of rational decision theory gets the wrong answer, and worse still, gives two contradictory answers.

It has turned out that this paradox has hidden depths. First, it is devious in its exploitation of our tendency to erroneous understandings of probability theory. Second, correction of the errors eliminates the paradox as first presented but leads us on to versions for which the paradox remains. To understand it we will have to make full use of the technical vocabulary of probability theory.

---

[13] 'Dependency hypotheses' is Lewis's ((1981a)) term. Other explanations may be given in terms of counterfactual conditionals, e.g. see Gibbard and Harper (1978).

First, we must distinguish cases in which there is an upper bound on the amount of money in the envelopes from those in which there is not. We call the former finite cases and the latter infinite cases. In all cases the calculation you applied is simply incorrect if the amount in your envelope is the minimum sum that could be in an envelope (since there is no possibility of having half that amount) and in finite cases it is also incorrect if it is the maximum (since there is no possibility of having twice that amount).

In analysing a stochastic situation, when we say let A be the amount of money in my envelope and B the amount in yours, what we have done is specified two random variables, A and B. What you wanted to know was the expectation of each envelope, that is, $E(A)$ and $E(B)$. When you calculated what you called the expected value of my envelope what you actually calculated was the *conditional* expectation of A given B and compared it with the conditional expectation of B given B. You took the statement '$E(A|B) > E(B|B)$' either as if it were the statement '$E(A) > E(B)$', or as if it implied that statement—and likewise for the statement '$E(B|A) > E(A|A)$'. So the first point to note is that as it stands, you either mistook conditional expectations for absolute expectations, or assumed, perhaps without reason, that a statement about conditional expectations implied a statement about absolute expectations.

The second point to note is that you have been beguiled into mistaking random variables for an expectation. The term 'conditional expectation' is ambiguous between being an expectation properly so called and being a random variable. If I calculate the conditional expectation of A given that $B = 5$, $E(A|B=5)$, then I will have calculated a true expectation. But if I calculate such an expectation just given the random variable B, then the conditional expectation of A given B, $E(A|B)$, is itself a random variable, and to get a true expectation we must calculate the expectation of *this* random variable, namely $E(E(A|B))$, and a standard theorem of probability theory shows this to be equal to $E(A)$. This technique of calculating an expectation via a conditional expectation is a standard and valuable technique of problem solving, frequently applied when there is no means of calculating $E(A)$ directly.

When we calculate $E(E(A|B))$ in the finite cases the paradox vanishes entirely, so we need only consider the infinite cases. There is a substantial taxonomy of infinite cases which is too extensive to properly explain here. So I will now simply mention some of the results that are available (Clark and Shackel (2000, (2003)).

1) In the infinite cases, because there cannot be a uniform probability function over an infinite set, it is not possible that for all amounts in your envelope, the conditional probability of the other envelope being half yours is ½ and the conditional probability of the other envelope being twice yours is ½.

2) It is possible for $E(B|A) > E(A|A)$ and yet for $E(B) = E(A)$ and so the inference from '$E(B|A) > E(A|A)$' to '$E(A) > E(B)$' is invalid.

3) If $E(A)$ is finite then no paradoxical cases arise. If $E(A)$ is not finite then two kinds of paradoxical cases arise. There are cases in which the expected gain on swapping envelopes ($E(E(A|B)-B)$) is infinite (so setting the paradox off again). However, in infinite cases all the expectations are sums of infinitely many terms and the mathematics of such sums must be respected. Saying that a sum is infinite is just short hand for saying it is unbounded, i.e. for any finite number however large one can add up finitely many of the infinitely many terms and exceed that number. We (Clark and Shackel (2000, (2003)) therefore call such cases 'unbounded paradoxical'. Because of the just explained precise meaning of 'having an infinite sum', it is controversial whether having an infinite sum is a way of having a well defined value, and some people have rejected the paradox on that ground (e.g. see Chalmers (1996, (2002)).

4) Whether those who reject unbounded paradoxical cases on those grounds are right or not, there are infinite cases for which the expected gain on swapping is finite, and hence that rejection cannot solve the paradox in general. The latter cases we call 'best paradoxical', best because the expected gain on swapping being finite is uncontroversially a way of the value of swapping being well defined, and yet from the set up of the scenario, there should not be an argument for swapping over sticking.

5) If best paradoxical cases are to be solved then some explanation must be given for why we can rule out calculating the expected gain on swapping by $E(E(A|B)\text{-}B)$ when using the technique of calculating an expectation via a conditional expectation.

6) In our published work on this paradox we advance the proposal that applying the latter technique must be done in such a way as to respect the causal features of the situation. Applying this constraint to the two-envelope case rules out using $E(E(A|B)\text{-}B)$ because of the symmetry of the causal features, but permits using $E(E(B\text{-}A|A\text{+}B))$ for the same reason, and the latter calculation gives zero expected gain on swapping. Hence when rational decision theory is formulated in a way which respects the causal features of the situation it can get the right answer. It is controversial whether this proposal is a solution (see Meacham and Weisberg (2003) and reply Clark and Shackel (2003)).

One line of attack is based on a thought which strikes many people as appealing on first hearing the paradox: that the paradoxical outcome is foisted on us by a subtle equivocation on 'x', and so it can be solved by specifying constraints that rule out such errors. I do not think this line can succeed and suspect that it is in part based on failure to understand the nature of random variables and the points made above about mistaking a conditional expectation for an expectation. For publications in this line see Jackson et al (1994); Chihara (1995); Horgan (2000); Schwitzgebel and Dever (2004).

There is a variant of the paradox in which you open your envelope and then decide to swap, and a further variant based on the argument that since you know that you would want to swap if you opened the envelope you should swap anyway. We say these thoughts are simply ways to beguile you into calculating the wrong expectation again (Clark and Shackel (2000):429), but again, our solution is controversial. Smullyan ((1993):189-92) puts forward an interesting non-probabilistic variant: that you will either gain x or lose $x/2$, so you will gain more than lose on swapping (or sticking when so reasoning based on the other envelope), and a good discussion of this variant is Chase (2002).

For further literature, see bibliographies of mentioned literature. Wikipedia has a reasonable online bibliography at http://en.wikipedia.org/wiki/Two_envelope_problem.

Pasadena paradox

A new paradox based on the St Petersburg game was published by Nover and Hajek (2004), which they called the Pasadena Paradox. The mathematical details of it are complex and so, whilst I am going to give the full story, I am not going to explain it in full mathematical generality. Similar to the St Petersburg game, we toss a coin until the first head appears. The outcome of the game is given according to the instructions on a stack of cards:

Top card. If the first head is on the first toss we pay you £2

Next card. If the first head is on the second toss you pay us £2

Next card. If the first head is on the third toss we pay you £8/3

Next card. If the first head is on the fourth toss you pay us £4

……

nth card. If the first head is on the nth toss the payment is $(-1)^{n-1}2^n/n$

and so on

Should you play the game?

Working out the expected value gives what is called an alternating series, which is an infinite sum in which the sign of the terms alternates.

E(value of game) = 1-1/2+1/3-1/4…= ln 2 (the natural logarithm of 2) $\approx$ 0.69.

This is positive so you should play. Before we play the cards are knocked over, and when restored it turns out that their order is now a positive card for you (instruction for pay off if first head on first toss), followed by the next five negative cards (instruction for pay off if first head on toss number 2 or 4 or 6 or 8 or 10), followed by the next positive card (instruction for pay off if first head on third toss), and then the next five negative cards, and so on. When we now calculate the expected value in this way we get

E(value of game) = 1 +-1/2-1/4-1/6-1/8-1/10 + 1/3 –1/12-1/14-1/16-1/18-1/20….and so on

=ln 2+ ½ ln1/5 $\approx$ -0.11

This is negative so you shouldn't play. But hang on a minute, all we've done is rearrange the instruction cards, so now we have shown both that you should and shouldn't play.

This paradox has not been much discussed in the literature yet. It exploits a well known feature of alternating series,[14] which is that if a series is convergent (has a finite sum) but not absolutely convergent (the sum of the absolute values of their terms does not converge) then for any real number its terms can be rearranged to give a series which sums to that number, and also for any of the three ways a series can diverge, its terms can be rearranged to give a series that is divergent in that way. It may appear that the two alternating series given above are the same infinite sum just because one is the rearrangement of the other. But that is not the case. The identity of an infinite sum is defined not just by what its terms are but also by the order of those terms.[15] In effect, the case we are considering contains a proof of that fact. Sums must have unique answers (since if they don't we can prove that all numbers are the same number) and hence if two sums have different answers they must be distinct sums (e.g. since 1+1 = 2 and 2+3 = 5, the sum 1+1 is distinct from the sum 2+3). The two infinite sums above have different answers therefore the sums must be distinct.

If we are willing to accept the infinite set of instruction cards and that rearrangements of sequences of that set of cards don't change the game that is being played, then apparently the game is well specified, and yet decision theory gives contradictory advice. We might deny the existence of the infinite set of cards, and so reject the game; but they are merely heuristic devices. If we accept the abstract nature of ordinary mathematics and of language there doesn't seem to be any problem with the existence of an infinite set of instructions, nor any problem with rearrangements of sequences of those instructions. The significant question is whether rearrangements of those instructions still constitute the same game. On the one hand, it is not obvious that it doesn't. On the other, given the contradictions into which we easily fall when considering infinite sets, one might insist

---

[14] Called the rearrangement theorem.
[15] In this way an infinite sum is very different from a finite sum, which depend only on their terms and not at all on their order.

that our proper understanding of such sets is constituted by the conceptual resources of the mathematics of such sets. Hence in proposing and thinking about decision cases involving infinite sets our proposals and thoughts must respect the content of those conceptual resources. In that case, since the mathematics of infinite sets insists that sequences with the same members in different orders are *necessarily* distinct, it is not enough to appeal to the heuristic of shuffling cards to ground the claim that the same cards in a different order constitute the same game. On the contrary, some reason must be given for why, *despite* the sequences of cards being regarded as necessarily distinct sequences, the game is the same.

An answer that has some force is this. The sequence of cards is not necessary for specifying the game. All that is required to fully specify the game is that all the possible outcomes of the game be specified and the payoff for each outcome be specified. As described, for each natural number it is specified what the payoff is if the number of throws to the first head is that number. But every possible outcome is correlated with a natural number and hence the payoffs for any particular outcome *are* specified. Hence it is merely the set of instructions that determines the identity of the game, not their order. If this answer is correct, then the value of the Pasadena game is indeterminate. In a recent discussion Colyvan (2006) proposes that the Pasadena game is ill-posed just because it has no expected value, and Hajek and Nover (2006) rebut that proposal.

References

Allais, M. 1953. La Psychologie De L'homme Rationnel Devant Le Risque: Critique Des Postulats Et Axiomes De L'école Américaine. *Econometrica*, 21 pp. 503.46.

Allais, M. & Hagen, O. 1979. *Expected Utility Hypotheses and the Allais Paradox : Contemporary Discussions of Decisions under Uncertainty with Allais' Rejoinder*. Dordrecht ; Boston: D. Reidel Pub. Co.

Blackburn, S. 2000. *Ruling Passions*. Oxford: Clarendon Press.

Bostrom, N. 1998. The Doomsday Argument: A Literature Review. Online at http://anthropic-principle.com/preprints/lit/index.html

Bostrom, N. 2002. *Anthropic Bias : Observation Selection Effects in Science and Philosophy*. New York ; London: Routledge.

Campbell, R. & Snowden, L. 1985. *Paradoxes of Rationality and Cooperation : Prisoner's Dilemma and Newcomb's Problem*. Vancouver: University of British Columbia Press.

Chalmers, D. J. 1996. The Two-Envelope Paradox: A Complete Analysis? : Online at http://consc.net/papers/envelope.html

Chalmers, D. J. 2002. The St. Petersburg Two-Envelope Paradox. *Analysis*, 62 (2), pp. 155-6.

Chase, J. 2002. The Non-Probabilistic Two Envelope Paradox. *Analysis*, 62 (2), pp. 157-9.

Chihara, C. S. 1995. The Mystery of Julius, a Paradox in Decision Theory. *Philosophical Studies*, 80 pp. 1-16.

Clark, M. 2002. *Paradoxes from a to Z*. London: Routledge.

Clark, M. & Shackel, N. 2000. The Two-Envelope Paradox. *Mind*, 109 (435), pp. 415-42.

Clark, M. & Shackel, N. 2003. Decision Theory, Symmetry and Causal Structure: Reply to Meacham and Weisberg. *Mind*, 112 (448), pp. 691-701.

Colyvan, M. 2006. No Expectations. *Mind*, 115 pp. 695-702. Numb 459.

Dembski, W. A. 1998. *The Design Inference : Eliminating Chance through Small Probabilities*. Cambridge: Cambridge University Press.

Dieks, D. 1992. Doomsday - Or: The Dangers of Statistics. *Philosophical Quarterly*, 42 (78.84),

Dorr, C. 2002. Sleeping Beauty: In Defence of Elga. *Analysis*, 62 (4), pp. 292-5.

Elga, A. 2000. Self-Locating Belief and the Sleeping Beauty Problem. *Analysis*, 60 (2), pp. 143-6.

Gibbard, A. & Harper, W. L. 1978. Counterfactuals and Two Kinds of Expected Utility. In *Foundations and Applications of Decision Theory*. Ed. Hooker, L., and McLennen (eds.) Dordrecht: Reidell.

Gott, J. R. 1993. Implications of the Copernican Principle for Our Future Prospects. *Nature*, 363 pp. 315. Issue 6427.

Hajek, A. & Nover, H. 2006. Perplexing Expectations. *Mind*, 115 pp. 703-20. Numb 459.

Horgan, T. 2000. The Two-Envelope Paradox, Nonstandard Expected Utility, and the Intensionality of Probability. *Nous*, 34 (4), pp. 578-603.

Horgan, T. 2004. Sleeping Beauty Awakened: New Odds at the Dawn of the New Day. *Analysis*, 64 (1), pp. 10-21.

Jackson, F., Menzies, P. & Oppy, G. 1994. The Two Envelope 'Paradox'. *Analysis*, 57 pp. 34-41.

Jaynes, E. T. 1973. The Well Posed Problem. *Foundations of Physics*, 4 (3), pp. 477-92.

Jeffrey, R. C. 1990. *The Logic of Decision*. 2nd ed. Chicago: University of Chicago Press.

Kahneman, D. & Tversky, A. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47 pp. 263-91.

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

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

Lewis, D. 1980. A Subjectivist's Guide to Objective Probability. In *Studies in Inductive Logic and Probability*. Vol. 2. University of California Press. Reprinted in — Lewis, D. *Philosophical Papers Vol 2*.

Lewis, D. 1981a. Causal Decision Theory. *Australasian Journal of Philosophy*, 59 pp. 5-30. 1986. *Philosophical Papers Vol 2*.

Lewis, D. 1981b. Prisoner's Dilemma Is a Newcomb Problem. *Philosophy and Public Affairs*, 8 pp. 235-40. 1986. *Philosophical Papers Vol. 2*.

Lewis, D. 2001. Sleeping Beauty: Reply to Elga. *Analysis*, 61 (3), pp. 171-5.

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

Meacham, C. J. G. & Weisberg, J. 2003. Clark and Shackel on the Two-Envelope Paradox. *Mind*, 112 (448), pp. 685-90.

Nover, H. & Hajek, A. 2004. Vexing Expectations. *Mind*, 113 (450), pp. 237-49.

Price, H. 1986. Against Causal Decision Theory. *Synthese*, 67 pp. 195-212.

Savage, L. J. 1972. *The Foundations of Statistics*. 2nd rev. ed. New York: Dover.

Schwitzgebel, E. & Dever, J. 2004. Using Variables within the Expectation Formula. Online at http://www.faculty.ucr.edu/~eschwitz/SchwitzPapers/TwoEnvelope040216.pdf

Shackel, N. Forthcoming. Bertrand's Paradox and the Principle of Indifference. *Philosophy of Science*,

Smullyan, R. M. 1993. *Satan, Cantor and Infinity : And Other Mind Boggling Puzzles*. Oxford: Oxford University Press.

Sober, E. 2002. Intelligent Design and Probability Reasoning. *International Journal for Philosophy of Religion*, 52 (2), pp. 65 - 80. Online at http://philosophy.wisc.edu/sober/ID&PRword.PDF

Von Neumann, J. & Morgenstern, O. 1944. *Theory of Games and Economic Behavior*. Princeton: Princeton university press.

White, R. 2006. The Generalized Sleeping Beauty Problem: A Challenge for Thirders. *Analysis*, 66 (290), pp. 114-9.