THE SOLUTION TO THE SORITES PARADOX[1]
by Andrew Boucher

The Sorites Paradox is one of the older paradoxes around, being traceable at least back to the ancient Greeks from whom its name is derived. Like most paradoxes, it still stimulates discussion and disagreement today. As such I would like to put in my own two cents.

Here is one formulation of the Sorites Paradox. Consider a heap of sand on the beach. If you take away one grain of sand, then surely it still remains a heap. So if one keeps removing one grain at a time, one always still has a heap. Continuing in this way, eventually one ends up with none at all, which mathematically should also therefore be a heap. But clearly nothing is not a heap, contradiction. Symbolically, let H(i) mean “An accumulation of i grains of sand is a heap.” Then H(0) is false, while H(10 000 000) is true. The Sorites Paradox turns on the seemingly plausible assertion that, if H(n+1) is true, then also H(n) is. For if this conditional is true, then one may infer H(9 999 999), and then H(9 999 998), and so on, all the way down to H(0), contradicting the falsity of H(0).

The problem seems to lie in the vagueness of the word ‘heap’. A great quantity of sand forms a heap, one grain of sand does not, but somewhere in between there is a region where the matter seems vague. Unfortunately, past this observation, there is less agreement, because the introduction of vagueness does not, at least at first sight, resolve the underlying problem, since it takes none of the cogency away from the inference of H(n) from H(n+1).

That is, if the problem is that there is no sharp boundary between being a heap and being a non-heap, then the introduction of one more boundary does nothing to eliminate the problem of sharpness. If one cannot say where “heap” ends and “non-heap” begins, then one can no more say where “heap” ends and “vaguely a heap” begins, or where “vaguely a heap” ends and “non-heap” begins. If “heap” is vague, then so is “vaguely a heap.” And if it is difficult to imagine how H(n+1) can be true and H(n) false, then it is equally difficult to imagine how H(n+1) can be true and H(n) vaguely true. Yet some things do seem to be - unvaguely, flat out - a heap, while some things seem to be clearly not-heaps, and so sometimes it is just true, without any vagueness, that something is a heap, or not a heap.

A more modern formulation of the Sorites Paradox is with colours, so let us consider this construction. There are 1000 bars. The first is clearly red, the last (1000th) is clearly not red, and they are ordered in decreasing redness. It is supposed that one cannot detect a difference in color between the nth and (n+1)th bar. (If this is not true, then one can simply increase the number of bars in the thought experiment. 1000 is used because it seems large enough to do the trick, but is small enough to write down quickly.) So letting R(i) mean “the ith bar is red”, we have R(1) and not-R(1000). Since one cannot distinguish between the colours of the nth and (n+1)th bar, it seems we should be able to infer R(n+1) from R(n). But as before this produces a contradiction.

Again, we might like to say the problem is the vagueness of ‘red,’ and that it is difficult to pin down the n where the nth bar is red but not the (n+1)th. Fair enough. But

[1] Laurence Goldstein, in “Stephen Clark, the Laws of Logic and the Sorites”, in Philosophy, Vol. 84, No. 327 (Jan. 2009), pp. 135-143, enunciated many if not most of the arguments in this paper much earlier.

then equally it is difficult to pin down the n where it is difficult to pin down where the nth bar is red but not the (n+1)th.

The reason why I like the formulation in terms of colours is that, scientifically anyway, there *is* in fact a sharp boundary for colours. Red (I am consulting wikipedia, which shows pretty much the extent of my knowledge) has wavelength 620-750 nm, while orange has wavelength 590-620. Leaving aside the uncomfortable fact that, according to this description, at 620 exactly something is both red and orange, we can say that, even though at 619.9999 nm and 620.0001 nm colours are probably indistinguishable, one is orange (and so not red) and the other red. So, scientifically speaking anyway, the Sorites Paradox is resolved because one cannot always infer R(n+1) from R(n), and this conditional fails to hold when n is the colour at 620.0001.

But, you may say, and correctly at that, our intuitive notion of red is not the scientific one. Fair enough. But then it is fair as well to point out that philosophers may attribute a bit too much to our intuitive notion of red. For instance, suppose we actually asked people, for each of the thousand bars, to answer Yes or No to the question "Is this red?" If the bars are not shown in order (and perhaps even if they are), for many individuals there are probably i,j,k with $i < j < k$ such that the ith and kth bar receive a "Yes" answer but the jth bar receives a "No". If the same individual undergoes the same experiment twice, then probably there exists i (and probably several) such that the ith bar receives different replies in the first and second experiments. In any case, except for the presumably small minority which sees the first bar as non-red or the 1000th bar as red, for each instance of the experiment, there will always be an n such that R(n) but not R(n+1). There may be more than one such n, but there will at least be one. The n may differ according to the individual and instance (if the same individual undergoes more than once the same experiment), but again there will be at least one.

It is correct that in this experiment, the individual may feel unable to reply Yes or No, but instead prefers to say, I Can't Tell. This does not change any of the observations of the previous paragraph since every I Can't Tell can simply be considered as a non-Yes, and thus counted as a No.

Now maybe one might object to this notion of experiment I have proposed by saying that an individual can misapply what he means by 'red', and say 'Yes" when he should be saying 'No', or *vice versa*. While I don't want to dismiss the possibility completely out of hand - people do admit to misusing words, after all, every now and again - I don't think there really can be any better indication of what a person means by the word 'red' than asking whether or not a particular thing looks red at a particular moment. Only it must be noted, first, the same person may truthfully the next moment call a thing non-red when he called it previously red or *vice versa*. This is not inconsistent, but simply an indication either the context has changed (a brighter light, for instance) or, perhaps, his meaning of 'red' has changed. And secondly, if we decide to avoid the problem of changing meaning by coronating the meaning of 'red' as the meaning we give at one particularly chosen moment in our experiment, then that still does not determine how 'red' should be applied in every other example of the experiment.

One might thinking of changing "is red" to "looks red." For if two samples are indistinguishable, and one looks red, then the other must surely look red as well. In fact, this does not necessarily follow. One can see two samples and consider them indistinguishable, then be shown one sample and declare it red, and then at some later time be shown the other sample and say it looks non-red. Indeed the *same* sample might look red at one time or to one individual, but not at another or to another. So "looks red" is not completely specified, and one needs to spell out the instant of looking red and the individual doing the looking, just as one should be specifying the instant and individual for any comparison of indistinguishability. Now first, the same individual probably cannot be thinking of the indistinguishability of two samples and the redness of the samples at the same time, so "x looks red to z at t1" and "x and y look indistinguishable to z at t2" can only be true if t1 is not equal to t2. But let that pass, and suppose the contrary, so that the same individual can determine the indistinguishability and redness of samples n-1 and n at the same time, and the indistinguishability and redness of samples n and n+1 at the same time. But these "same times" are *different* times, and one would find by experiment, where one shows successive samples (n,n+1) in some order, for n = 0,…,999, to the same individual, that there is some n such that the individual will differ about the redness of sample n at the two times. That is, at some point in the experiment, for some n, an individual will face samples n-1 and n and aver they are indistinguishable and both red, and at some other point the individual will face samples n and n+1 and assert they are indistinguishable and both non-red. Such logically *must* be the case, should the first sample be clearly red, the last sample clearly non-red, and successive samples indistinguishable. This n may differ for each experiment conducted, and there may even be more than one n per experiment, but there *must* be at least one such n per experiment.

The problem of the Sorites Paradox is not one of vagueness, but one of meaning. Reflecting on meaning correctly, resolves the paradox.