Dialetheism and the Sorites Paradox

Graham Priest

January 28, 2020

Departments of Philosophy, CUNY Graduate Center, and the University of Melbourne

Abstract: The paper explains and defends a dialetheic account of vagueness, and its solution to the sorites paradox. According to this, statements in the middle of a sorites progression are both true and false. After an explanation of an appropriate paraconsistent logic, detailed models of sorites transitions are provided of. Crucial to any supposed solution the the sorites paradox is how it handles the matter of cut-offs. Much of the paper concerns how a dialetheic solution handles this.

Key Words: Vagueness, sorites paradox, dialetheism, paraconsistent logic, cut-offs, forced-march sorites.

1 Introduction

A Sorites Paradox arises when a predicate, P, is vague in a certain sense. Specifically, it appears to satisfy a certain tolerance condition: for any object, a, if Pa, and b is any object which differs relevantly from a in only a very small amount—maybe an indistinguishable amount—then Pb too. Or, since the amount of difference concerned is symmetric, we could put it this way: if a and b differ relevantly from each other in only a very small way then Pa is true iff Pb is true. 'Drunk', 'adult', 'tall', 'bald', are paradigms of tolerant predicates. Given such a predicate, we can construct a sequence of objects $a_0, a_1, ..., a_n$, such that each member of the sequence differs only minimally in the relevant way from its predecessor, and Pa_0 is clearly true whilst Pa_n

is clearly not true. Applying tolerance down the chain allows us to establish that Pa_n is true, thus:.

$$\dfrac{\dfrac{Pa_0 \quad Pa_0 \equiv Pa_1}{Pa_1} \qquad Pa_1 \equiv Pa_2}{Pa_2}$$

$$\ddots$$

$$\dfrac{Pa_{n-1} \quad Pa_{n-1} \equiv Pa_n}{Pa_n}$$

It is perhaps more normal to formulate the sorites with a conditional (from left to right), rather than a biconditional. But the conditional in the other direction is not contentious. And it is the biconditional which expresses tolerance. Hence, using the biconditional is more accurate.[1]

Like its more famous cousin, the liar paradox, the Sorites Paradox is reputed to have been discovered (or invented) by the Megarian philosopher Eubulides;[2] but though there are occasional references to it in Ancient Greek philosophy, and unlike its more famous cousin, it has had a very low profile historically. There are, as far as I know, no discussions of it in the great Medieval period of logic.[3] Nor, with one or two exceptions,[4] is it an issue in modern logic—until, that is, the 1960s and 1970s, when it suddenly took off.[5] Since then, it has generated an enormous literature. Why it should have shot suddenly from oblivion in this way, I have no idea.

Since then, the literature has provided a large number of suggested solutions. The one that will concern us in this essay is a dialetheic solution. In sorites progressions, the things at the beginning are clearly P; the things at the end are clearly $\neg P$. In the middle there appear to be borderline cases, symmetrically poised between the two. In a dialethic account of the matter, these are both P and $\neg P$.[6] The possibility of this approach was certainly

[1]For a general overview of the Sorites Paradox, see Hyde (2011), Sorensen (2012), Keefe (2000), Keefe and Smith (1997), and, of course, the essays in this volume.

[2]According, for example, to Diogenes Laertius. See Hicks (1925), ii, 108. See also the discussion in Kneale and Kneale (1962), esp. p. 108.

[3]Or, for that matter, in any Asian texts I know.

[4]Notably, Russell (1923).

[5]See, for example, Goguen (1969) and Fine (1975).

[6]There is, of course, another symmetric possibility: that they are neither P nor $\neg P$. I will comment on this possibility briefly at the end of this essay.

mooted before the lift-off of the sorites in modern logic.[7] But it was first put squarely on the table, as far as I am aware, by Hyde (1997). Since then, it has been sympathetically explored by a number of people, including Ripley (2005), (2013), Hyde and Colyvan (2008), Weber (2011), and myself (2010a).

2 Logical Background

Of course, saying that the borderline cases are both true and false is only a first move in the game. To explain how it is applied requires some logical background, especially concerning paraconsistent logic. Paraconsistent logics are logics which can tolerate contradictions, since the principle of inference $A, \neg A \vdash B$ (Explosion) fails.

There are many paraconsistent logics,[8] and most of them can be deployed in a dialetheic solution to the sorites. But by far the simplest, and one which exposes the crucial moves in play in the sorites, is LP.[9]

We take a standard first-order language with connectives $\vee, \wedge, \neg$ (*or*, *and*, *not*), and quantifiers $\forall, \exists$ (*all*, *some*). We may suppose that there are no function symbols, and that all the predicates are monadic. $A \supset B$ is defined in the familiar way, as $\neg A \vee B$, and $A \equiv B$ as $(A \supset B) \wedge (B \supset A)$.

An interpretation for the language is a structure $\langle D, \theta \rangle$. D is the non-empty domain of quantification. θ is the denotation function. That is, for every constant, c, $\theta(c) \in D$; and for every monadic predicate, P, $\theta(P)$ is a pair $\langle X, Y \rangle$, where $X, Y \subseteq D$ such that $X \cup Y = D$.[10] X and Y are the *extension* and *anti-extension* of P. Intuitively, the objects in X are those that make P true and the objects in Y are those that make P false. I will write X and Y as $\theta^+(P)$ and $\theta^-(P)$, respectively.

We define what it is for a sentence (that is, a formula with no free variables) to be true, $\Vdash^+$, and false, $\Vdash^-$, in an interpretation, as follows:

- $\Vdash^+ Pc$ iff $\theta(c) \in \theta^+(P)$
- $\Vdash^- Pc$ iff $\theta(c) \in \theta^-(P)$

[7] For example, by Plekhanov (1941), 114 ff; McGill and Parry (1948); and Jaśkowski (1969).

[8] See Priest (2002).

[9] See, e.g., Priest (2008), ch. 7.

[10] If one drops this restriction, one obtains the logic of First Degree Entailment, FDE. If one adds the constraint that $X \cap Y = \emptyset$, one obtains classical logic. See Priest (2008), ch. 8.

- $\Vdash^+ \neg A$ iff $\Vdash^- A$
- $\Vdash^- \neg A$ iff $\Vdash^+ A$
- $\Vdash^+ A \wedge B$ iff $\Vdash^+ A$ and $\Vdash^+ B$
- $\Vdash^- A \wedge B$ iff $\Vdash^- A$ or $\Vdash^- B$
- $\Vdash^+ A \vee B$ iff $\Vdash^+ A$ or $\Vdash^+ B$
- $\Vdash^- A \vee B$ iff $\Vdash^- A$ and $\Vdash^- B$

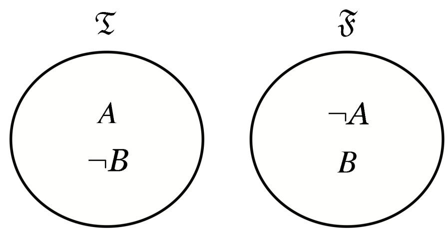
To give the truth/falsity conditions for the quantifiers, we assume that the language has been augmented by a constant, k_d, for each $d \in D$, such that $\theta(k_d) = d$. $A_x(c)$ is A with every free occurrence of the variable x replaced by the constant c.

- $\Vdash^+ \exists x A$ iff for some $d \in D$, $\Vdash^+ A_x(k_d)$
- $\Vdash^- \exists x A$ iff for all $d \in D$, $\Vdash^- A_x(k_d)$
- $\Vdash^+ \forall x A$ iff for all $d \in D$, $\Vdash^+ A_x(k_d)$
- $\Vdash^- \forall x A$ iff for some $d \in D$, $\Vdash^- A_x(k_d)$

An inference is valid if it preserves truth in all interpretations. That is, $\Sigma \models A$ iff for all interpretations, $\Vdash^+ A$ when, for all $B \in \Sigma$, $\Vdash^+ B$.

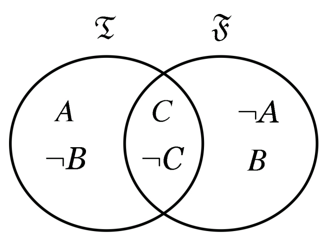
3 The Material Conditional and Biconditional

For those unfamiliar with paraconsistent logic, and in virtue of what is to come, it is worth reflecting on LP a little further. In classical logic, any situation (interpretation) partitions all truth-bearers into two classes, the true ($\mathfrak{T}$) and the false ($\mathfrak{F}$). The two classes are mutually exclusive and exhaustive. A disjunction is in $\mathfrak{T}$ if one or other disjunct is; in $\mathfrak{F}$ if both are; dually for conjunction; and a truth-bearer is in $\mathfrak{T}$ iff its negation is in $\mathfrak{F}$, thus:



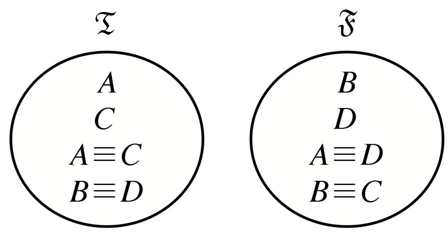
An inference is valid iff there is no interpretation in which all the premises are in $\mathfrak{T}$, but the conclusion is not. Given this set-up, there is no situation where, for any A, both A and $\neg A$ are in $\mathfrak{T}$. *A fortiori*, there is no situation where A and $\neg A$ are in $\mathfrak{T}$, and B is not—whatever B chosen. That is, Explosion is valid.

In LP, everything works *exactly* the same way, except that in some interpretations the $\mathfrak{T}$ and $\mathfrak{F}$ zones may overlap. Given that negation works in the same way, it follows that if C is in the overlap, so is its negation. Thus, we have the following:

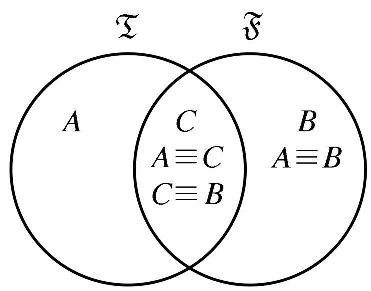


For a situation of this kind, both C and $\neg C$ are in $\mathfrak{T}$ (and in $\mathfrak{F}$ as well; but at least in $\mathfrak{T}$). But B is not in $\mathfrak{T}$. Given exactly the same definition of validity as before, it follows that Explosion is not valid. Note also, that the same diagram shows that material detachment for $\supset$ fails, since C and $\neg C \vee B$ are both in the $\mathfrak{T}$ zone whilst B is not.

Turning to the material biconditional: in classical logic if A and B are both in $\mathfrak{T}$ or both in $\mathfrak{F}$, then $A \equiv B$ is in $\mathfrak{T}$. Whereas if one is in $\mathfrak{T}$, and the other is in $\mathfrak{F}$, $A \equiv B$ is in $\mathfrak{F}$, thus:



In the paraconsistent case everything is the same, except that the $\mathfrak{T}$ and $\mathfrak{F}$ zones may overlap. Thus we have the following picture:



Though a sentence may now be in both zones, it remains the case that if A and B are in the same zone, $A \equiv B$ is in $\mathfrak{T}$; and if one is in $\mathfrak{T}$, and the other is in $\mathfrak{F}$, $A \equiv B$ is in $\mathfrak{F}$. A truth of the form $A \equiv B$ therefore expresses the fact that A and B are in the same zone. Its negation expresses the fact that they are in different zones. In particular, one may check that the following are valid:

- $A, B \vdash A \equiv B$
- $\neg A, \neg B \vdash A \equiv B$
- $A \equiv B \vdash (A \wedge B) \vee (\neg A \wedge \neg B)$
- $A, \neg B \vdash \neg(A \equiv B)$
- $\neg(A \equiv B) \vdash (A \wedge \neg B) \vee (B \wedge \neg A)$

Note that the material biconditional supports detachment no more than does the material conditional. As the above diagram shows, the inference $C, C \equiv B \vdash B$ is not valid.[11]

[11] I note, also, that if the logic is FDE, and so has truth value gaps, a true material conditional, does not indicate membership of the same zone. Thus, suppose that A is both true and false, and B is neither. Then $(\neg A \vee B) \wedge (\neg B \vee A)$ is true (only).

4 The Dialetheic Analysis

Against this background, a dialetheic analysis of the Sorites Paradox can now be explained very simply: all the premises are true, but material detachment is invalid. As we have seen, a standard sorites argument has premises of the form:

- Pa_0
- $Pa_i \equiv Pa_{i+1}$ (for $0 \leqslant i < n$)

The conclusion is Pa_n.

The state of affairs concerning these statements is given by an interpretation, I, where, for some $j \leqslant k$:

- $\theta(a_i) \in \theta^+(P)$ for $0 \leq i \leq k$
- $\theta(a_i) \in \theta^-(P)$ for for $j \leq i \leq n$

This may be depicted as follows, where $+$ indicates truth, $-$ indicates falsity, and $\pm$ indicates both:

$$\begin{array}{cccccccccc} Pa_0 & \ldots & Pa_{j-1} & Pa_j & \ldots & Pa_k & Pa_{k+1} & \ldots & Pa_n \\ + & \ldots & + & \pm & \ldots & \pm & - & \ldots & - \end{array}$$

It is easy to check that all the premises are true, and the conclusion is not true. ($Pa_i \equiv Pa_{i+1}$ is both true and false if $j-1 \leqslant i \leqslant k$.) And this is possible because detachment for $\equiv$ fails.[12]

Let us write $A!$ for $A \wedge \neg A$. Then the premises of the sorites argument do not, note, tell us which i or is are such that $Pa_i!$. That is, they do not entail $Pa_i!$ for any particular i. But the premises and the negation of the

[12]Hyde himself is more sympathetic to a subvaluationist account of the sorites. The basic idea is the same, except that we subvaluate. Given any LP interpretation of the above kind, let us call a *sharpening* any classical valuation such that for some $j \leqslant m \leqslant k$, Pa_i is just true up to $i = m$, and just false thereafter. A subvaluation makes a formula *sub-true* if it is true on some sharpening, and *sub-false* if it is false on some sharpening. (Something can be sub-true and sub-false.) Every premise of the argument is now sub-true (and maybe sub-false as well), and its conclusion is just sub-false. (So for Hyde, a statement concerning an object in the border area is not, strictly speaking, true and false, but sub-true and sub-false.) Defining validity in terms of the preservation of sub-truth still invalidates material detachment for the conditional and biconditional. However, it also invalidates other things that appear desirable, such as adjunction: $A, B \vdash A \wedge B$.

conclusion do deliver $\bigvee_{0 \leqslant i \leqslant n} Pa_i!$.[13] That is, they entail that a contradiction occurs *somewhere* in the progression.

I note that one might formulate a version of the sorites where the major premises are expressed with a detachable biconditional, $\leftrightarrow$ (and LP can certainly be augmented with such a conditional).[14] But the tolerance of a vague predicate is expressed exactly by the thought that successive members of the progression have the same truth value: both true or both false. (Being true *and* false is not a third truth value. It is the possession of two truth values.) So the material biconditional is the correct connective to use to express tolerance. There is no particular reason to suppose that any stronger connection holds. It would be wrong, then, to express tolerance using a non-material conditional, such as a detachable one.

5 Identity Sorites

There is another sort of sorites argument that is worth noting. This uses not a (bi)conditional, but identity.[15] Suppose that our sorites is a colour sorites, say between red and blue; and let b_i be the term 'the colour of a_i'. Then if the appropriate tolerance obtains (for example if the colour of each a_i is phenomenologically indistinguishable from the colours of its neighbours), we have each of $b_i = b_{i+1}$, for $0 \leqslant i < n$. By $n-1$ applications of the transitivity of identity ($a = b, b = c \vdash a = c$), we have $b_0 = b_n$, thus:

$$\cfrac{\cfrac{\cfrac{b_0 = b_1 \quad b_1 = b_2}{b_0 = b_2} \quad b_2 = b_3}{b_0 = b_3}}{\ddots \quad \cfrac{b_0 = b_{n-1} \quad b_{n-1} = b_n}{b_0 = b_n}}$$

But $b_0 = b_n$ is clearly not true: a_0 is not the same colour as a_n.

[13] Pa_0 and $Pa_0 \equiv Pa_1$ entail $Pa_0! \vee Pa_1$. This, plus $Pa_1 \equiv Pa_2$ entail $Pa_0! \vee Pa_1! \vee Pa_2$, and so on, till $Pa_0! \vee ... \vee Pa_{n-1}! \vee Pa_n$, whence $\neg Pa_n$ delivers the last contradictory disjunct.

[14] See, e.g., Priest (2006), ch. 6.

[15] See Priest (2010b).

Sorites arguments are of a piece, and should have the same kind of solution (the Principle of Uniform Solution: same sort of paradox, same sort of solution).[16] The dialetheic solution to the previous kind of sorites can be extended to this kind very simply. We may define $x = y$ in second-order logic, in the standard Leibnizian way, as $\forall X(Xx \equiv Xy)$, where the second-order variables range over an appropriate set of properties. Note that what the Leibnizian definition requires is that for every property, P, Px and Py have the same truth value.[17] This is exactly what the material biconditional expresses. But if we are in a paraconsistent context, the $\equiv$ is that of a paraconsistent logic. Assuming that second-order quantifiers work as do first-order quantifiers, except with a domain of properties instead of objects, it follows that $=$ is not transitive.[18] That is, $a = b, b = c \nvdash a = c$. Thus suppose, for the sake of illustration, that there is only one property, P, and consider an interpretation where $\theta(a)$ is in the extension of P, but not in its anti-extension; $\theta(c)$ is in the anti-extension of P, but not its extension; and $\theta(b)$ is in both. Then $Pa \equiv Pb$ and $Pb \equiv Pc$ are both true, but $Pa \equiv Pc$ is not. Given that P is the only property, it follows that $a = b$ and $b = c$ are true, but $a = c$ is not.

Since the transitivity of identity fails, the sorites argument is broken. Thus, suppose, to consider the same illustration, that P is 'is a shade of red' (where this is the only predicate). Then the situation is as follows:

$$\begin{array}{cccccccc} Pb_0 & \ldots & Pb_{j-1} & Pb_j & \ldots & Pb_k & Pb_{k+1} & \ldots & P_n \\ + & \ldots & + & \pm & \ldots & \pm & - & \ldots & - \end{array}$$

If $I_{i,j}$ is $b_i = b_j$, we then have:

$$\begin{array}{ccccccccc} I_{0,1} & \ldots & I_{j-2,j-1} & I_{j-1,j} & \ldots & I_{k,k+1} & I_{k+1,k+2} & \ldots & I_{n-1,n} \\ + & \ldots & + & \pm & \ldots & \pm & + & \ldots & + \end{array}$$

And the values of $I_{0,i}$ look like this:

[16] See Priest (1995), Part 3.

[17] I shall speak indifferently of predicates and properties, the use/mention elision circumventing tiresome prolixity (which the reader may provide for themselves).

[18] See Priest (2014), ch. 2. The book shows that non-transitive identity has a lot more going for it than what is at issue here.

$I_{0,0}$	$\dots$	$I_{0,j-2}$	$I_{0,j-1}$	$\dots$	$I_{0,k}$	$I_{0,k+1}$	$\dots$	$I_{0,n}$
$+$	$\dots$	$+$	$\pm$	$\dots$	$\pm$	$-$	$\dots$	$-$

While we are on the topic of other versions of the Sorites Paradox, I note that the present solution deals equally with some prominent versions of these. In one, we do not have a collection of major premises, but a single quantified one: $\forall x(Px \supset Px')$, where the variables range over the $a_i s$ and x' is the object next to x. So the argument now is: $Pa_0, \forall x(Px \supset Px') \vdash Pa_n$. Material detachment fails in exactly the same way. This version has a contraposed form: $Pa_0, \neg Pa_n \vdash \exists x(Px \wedge \neg Px')$. The version is valid, but the conclusion does not express the existence of a unique cut-off point. $Pa_i \wedge \neg Pa_{i+1}$ is true for all $j-1 \leq i \leq k$. More on this matter in a second.

6 Cut-Offs

So much for the basic ideas of a dialetheic account of the Sorites Paradox. We are far from done yet, though. Let us see why, by turning to the major objection to the above account.

Come back to our original sorites. Suppose that one subscribed to classical logic: statements are either true or false, but not both. Then there would be some l such that:

(*) $\forall i \leq l\, Pa_i$ and $\forall i > l\, \neg Pa_i$

l is a precise cut-off point. That there should be a cut-off point of this kind seems completely wrong. It appears to be in the very nature of vague progressions that there is no such distinguished point. That is, indeed, what drives the Sorites Paradox.

A dialetheic account of the sorites has a simple solution to that problem. As just noted, (*) is true for all $j-1 \leq l \leq k$. There is, as required, no unique cut-off point.[19]

But, it may fairly be said, the solution has just moved the problem. Before, we had a problem with the cut-off between truth and falsity. Now, we have the same problem with the cut-off between being true (only)[20] and being true and false. This seems just as bad. Indeed, the problem is even

[19]See Weber (2010) for further discussion.

[20]A frequent objection to a dialetheic solution to the paradoxes of self-reference is that the dialetheist cannot express the claim that something is true only. This is completely

worse. We now have *two* cut-off points: there is another between being true and false, and being false (only) as well.[21]

Neither is this simply an artifact of the model of Section 4. The very basis of a dialetheic approach to the sorites is that the objects, a, in a borderline area between those that are P and those that are $\neg P$ are characterised by the conjunction of these extremes, $Pa \land \neg Pa$. Now, consider the right-hand borderline in the dialetheic case. (Details with the left-hand one are similar.) If there were objects in a borderline between being P and $\neg P$, and being $\neg P$, they would therefore be characterised by the conjunction $(Pa \land \neg Pa) \land \neg Pa$. But this is logically equivalent to $Pa \land \neg Pa$. (To be borderline between being P and being borderline P is just to be borderline P.) That is, there is no borderline transition between the two categories: the sequence goes straight from one to the other.

One might think that this mislocates the issue. The borderline in question is not that between being P and $\neg P$, and being $\neg P$, but between P being true and P not being true; that is, between those as such that $T\langle Pa\rangle$ and those such that $\neg T\langle Pa\rangle$ (where T is the truth predicate, and angle brackets indicate a name-forming device). If T is a crisp predicate, there will be a precise cut-off point. But if T is a vague predicate, there will be as such that $T\langle Pa\rangle \land \neg T\langle Pa\rangle$. Of course, if negation commutes with truth, this is just $T\langle Pa\rangle \land T\langle \neg Pa\rangle$. Given the T-schema, this is equivalent to $Pa \land \neg Pa$. Hence there is no mislocation: there is no separate location. There are good reasons to suppose that negation does not commute with truth.[22] However, even in this case, we are no better off. For there will still be a last a such that $T\langle Pa\rangle$ (even if it is the case that $T\langle Pa\rangle \land \neg T\langle Pa\rangle$) and after which it is not.[23] Again, we have a counter-intuitive cut-off.

---

incorrect: that A is true and not false is expressed in the obvious way: $T\langle A\rangle \land \neg F\langle A\rangle$. What it cannot do is force this claim to be consistent. That is quite another matter; and it is not at all obvious that this is a requirement that should be met. Indeed, it is not even clear that the requirement *can* be met—even by a classical logician. See Priest (2006), 20.4.

[21] In discussions of the paradoxes of self-reference, there is a phenomenon of the revenge paradox, in which the theoretical machinery of a solution is used to reformulate a new paradox of the same kind—or really just to rephrase the old paradox. The situation we now face with the Sorites Paradox can be thought of as a revenge problem of exactly the same kind. The parity between the two sorts of revenge phenomena is discussed in Priest (2010a).

[22] See Priest (2006), 4.7.

[23] There is an important issue here about the logic in which the semantics is given, and

7 The Forced March Sorites

Things seem bad. The only way to make them better is to make them worse. The hard fact of the matter is that *whatever solution one endorses*, one is stuck with a precise cut-off point of some kind. One way to see this is to consider the "forced march sorites".[24]

Consider our example sorites. Let Q_i be the question 'Is it the case that Pa_i?' If asked this question, there is some appropriate range of answers. What these are, exactly, does not matter. They might be 'yes', 'no', 'I don't know', 'yes, probably', 'er...' or anything else. All that we need to assume is that an appropriate answer is justified by the objective state of affairs; specifically, by the nature of Pa_i. (The justification here is semantic, not epistemic. The answerer is personified simply to make the situation graphic.) Now, suppose I ask you the sequence of questions: Q_0, Q_1, ... Given any question, there may be more than one appropriate answer. For example, you might say 'yes'; you might say 'same answer as last time' (having said 'yes' last time). All I insist is that once you answer in a certain way you stick to that until that answer is no longer appropriate. Suppose that in answer to the question Q_0, you answer A. This may also be justified in answer to Q_1, Q_2, and so on. But because of the finitude of the situation, there must come a first i where this is no longer the case, or it would be justified in answer to Q_n, which it is not.[25] Nor is the logic one takes to apply to the situation relevant, be is classical, or to have truth value gaps, or to have truth value gluts. How things of the form $\neg Pa_i$ behave have not come into the matter. Thus, for some i, Pa_i justifies this answer; Pa_{i+1} does not. The objective situation therefore changes between Pa_i and Pa_{i+1} in such a way. We have a precise cut-off.

Let us consider a couple of replies. Here is one. The existence of a cut-off point seems odd because of the apparently arbitrary nature of its location.

---

so of the sense of 'not' here. However, important as this issue is, it is not relevant here. The claim holds whether the negation is Boolean or paraconsistent.

[24] The term was coined, as far as I know, by Horgan (1994), section 4, though the form of argument is essentially the same as the original Eubulidean version. The version I give here is slightly different from, and, it seems to me, tougher than, the version Horgan gives. For the original formulation of the argument see Wiliamson (1994), ch.1, and Keefe and Smith (1997), ch. 2. What follows in this section comes, essentially, from Priest (2003).

[25] There are, in fact, continuous versions of the sorties, in which there is no such finitude. See Weber and Colyvan (2010). But as that paper makes clear, there are still precise cut-off points: the appropriate least upper bound or greatest lower bound.

Suppose that the correct answer in the forced march sorites changed at *every* question. The arbitrariness, and so the oddness, would then disappear. But how could it change at every step? One possibility is that to answer the question I simply *show* you the object at issue—which is changing from stage to stage. Another is that an answer is of the form 'It is true to degree r'—as in fuzzy logic—where r is a different real number every time.

The first response raises the question of what, exactly, a language is. If you ask me, for example, what colour something is, and I simply show it to you, is this response part of a *language* game? Perhaps so, but even if it is, the point is irrelevant. The sorites problem is generated by a verbal language, with vague predicates, questions, and answers. We need a solution that applies to *that* language.

One response to the second suggestion is similar. Even though, in this, the response to the question is by saying, not showing, a language which can refer to every real number—an uncountable number of entities—is not a language we could speak: we seek a solution for *our* language. But I think that there are greater problems with this response. However one conceptualises degrees of truth, there are sorites progressions where truth value does not change all the time. Thus, even if you were changed by replacing one molecule of your body with a molecule of scrambled egg, you would still be as you as you could be. You change more than that every morning after breakfast. Similarly, dying takes time, and so is a vague notion. But when your ashes are scattered to the four winds, and thereafter—if not before—you are as dead as dead can be. And if a correct answer to the relevant question does not change at every point, we have a cut-off.

A second reply is to the effect that the answerer may "refuse to play the game". Of course, if they do this for subjective reasons, such as the desire to be obstreperous, this is beside the point. They might, however, do so for a principled reason, namely that the rules of the "game" are impossible to comply with. They lead the answerer, at some point, into a situation where they cannot conform.[26] Now, it would certainly appear that it is possible to comply with the rules at the start: the first few answers present no problem. But then we may simply ask the person to play the game as long as it is possible. If the answer changes before this, the point is made. If, however, they stop at some point before this, it must be because the situation is such as to require them both to give and not give the same answer as before. This

[26]Arguably, one finds a view of this nature in Dummett (1975).

was not the case at the question before, so the semantic situation has changed at this point. The only other possibility is for the answerer to say that the game is unplayable right at the start. But this can only be because there is no appropriate answer they can give even in the first case—and presumably, therefore, in all subsequent cases. This is not only implausible; it means that even in the most determinate case there is no answer that can be given. We are led to complete and unacceptable semantic nihilism.

What we see, then, is that any solution to the sorites will be forced to accept a counter-intuitive cut-off of some kind. It cannot disappear it. Of course, how we should theorise this cut-off is another matter. Different theorists do this in different ways. A cut-off may be theorised as a change from truth to falsity; a change from truth to neither truth nor falsity, or to both truth and falsity; a change from being 100% true to less than 100% true; a change from maximal degree of assertibility to less than maximal degree; and so on. Never mind the details. What the forced march sorites demonstrates is that any solution must face the existence of a cut-off.

This is, in fact, the *real* sorites problem. Anything in a proposed solution before this matter is addressed is just preliminary. A solution to the sorites must accept the existence of some cut-off point or other, and must explain, given the machinery it endorses, why we find the existence of such a thing counter-intuitive. That is all it can do; and it is on the strength of this explanation that it must be judged.

8 The Dialetheic Solution

So what is the dialetheic explanation? It is simple. A precise cut-off point is counter-intuitive because whatever i we choose, $Pa_{i-1} \equiv Pa_i$ and $Pa_i \equiv Pa_{i+1}$. In other words, Pa_i has the same truth value as each of its neighbours. This is what makes us think that there can be no cut-off. It might be pointed out that, on the present account, some of the negations of the biconditionals are true too; but this is beside the point. Tolerance is the *obvious* feature of vague predicates. We are moved by what is obvious.

We can put the point in another way, by looking at the truth values themselves. Let b_i be 'the (truth) value of Pa_i'. There are only two relevant properties: T and F, being true and being false.[27] What is the value of Tb_i? It is natural to suppose that statements of the form Tb_i are themselves vague.

[27] Other properties either apply to all truth values, or none, and so need not concern us.

(If the status of 'a is red' is moot, so is that of ' "a is red" is true'.) If truth commutes with negation then $Pa_i \wedge \neg Pa_i$ iff $T\langle Pa_i \wedge \neg Pa_i\rangle$ iff $T\langle Pa_i\rangle \wedge T\langle \neg Pa_i\rangle$ iff $T\langle Pa_i\rangle \wedge \neg T\langle Pa_i\rangle$ iff $Tb_i \wedge \neg Tb_i$ so the borderline of P lines up with that of T. But even if not, because of the vagueness of T, then between the area where Tb_i is true only and the area where it is false only, there must be a region where Tb_i is both true and false. Thus, for some $0 < x \leqslant y < n$:

$$\begin{array}{cccccccc} Tb_0 & \dots & Tb_{x-1} & Tb_x & \dots & Tb_y & Tb_{y+1} & \dots & Tb_n \\ + & \dots & + & \pm & \dots & \pm & - & - & - \end{array}$$

If we write E_i for the biconditional $Tb_i \equiv Tb_{i+1}$, we then have:

$$\begin{array}{cccccccccc} E_0 & \dots & E_{x-2} & E_{x-1} & E_x & \dots & E_y & E_{y+1} & \dots & E_n \\ + & \dots & + & \pm & \pm & \dots & \pm & + & + & + \end{array}$$

Every biconditional is (at least) true. Symmetrically, the same is true for the falsity predicate, $F\langle Pa\rangle$ $(= T\langle \neg Pa\rangle)$. Every biconditional of the form $Fb_i \equiv Fb_{i+1}$ is also (at least) true. Since this is so for both the (relevant) properties of the b_i, then $b_i = b_{i+1}$, for all i. That is, the truth value of each statement in the sorites progression is the same as those of its neighbours. No wonder a cut-off is counterintuitive!

In any given sorites, there remains the challenge of saying where, exactly, the cut-off between the borderline and non-borderline cases is. To find out, one simply has to take the forced march test. Run down the sequence until you can no longer give the answer 'yes'. That is where it is. Or roughly, anyway: you are not the idealised answerer of Section 7—and neither is anybody else. Individuals have too many subjective factors operating on them. Since meaning is not subjective, but socially embedded, a more accurate guide to the cut-off is to take multiple speakers and circumstances, and aggregate out the answers in some way. This will provide a more robust determination. Note that this is not to say that what is so is what "an average person" believes: it merely reflects the fact that words are our words, and mean what we use them to mean. Can one find out this meaning by empirical considerations? Of course: this is what empirical linguistics is all abouit.

9 The Epistemic Solution

I have located the heart of the problem with the Sorites Paradox as how to explain why the existence of precise cut-off points seems so counter-intuitive, and given a dialetheic explanation. This is not the place to discuss all other possible explanations. However, a major alternative to a dialethic solution to the Sorites Paradox is an epistemic one, based on classical logic. Let me comment on that.

An epistemic solution to the Sorites Paradox is advocated, among others, by Sorensen and Williamson.[28] This solution takes apparently vague predicates to be precise. (In a sense, then, there are no vague predicates.) In a soritical progression there is a precise cut-off where the sentences turn from true to false or vice versa. The explanation for why one finds the existence of such a cut-off counter-intuitive has, then, to be provided in epistemic terms.[29] The most articulated explanation of the matter is given by Williamson, and depends on the fact that we cannot know where the cut-off is. (In particular, then, it cannot be determined by any empirical investigations.) Before we even get to the explanation, let us consider this claim.

Why can we not know where the cut-off point is? This is due to what Williamson calls the 'margin of error principle', which he states as follows: 'A' is true in all cases similar to cases in which 'It is known that A' is true. Why endorse this principle? Because if a and b are effectively the same in terms of the evidence they deliver, but Pa is true and Pb is false, then I cannot know that Pa. My evidential state is not such as to make my belief reliable, so it is not knowledge.[30] Now, suppose that the cut-off point is at Pa_i. That is, Pa_i is true, and Pa_{i+1} is not. Suppose one knows where the cut-off point is. Then one knows that Pa_i is true. But since a_i and a_{i+1} evidentially indistinguishable, Pa_{i+1} by the margin or error principle—which, *ex hypothesi*, it is not.

A major worry here is that the very phenomenon that explains why we cannot know where the cut-off point is undercuts its very existence. The meanings of vague predicates are not determined by some omniscient being in some logically perfect way. Vague predicates are part of *our* language. As a result, their meaning must answer in the last instance to the use that *we* make of them. It is therefore difficult to see how there could be a semantic cut-off at a point that is *in principle* inaccessible to agents with our cognitive apparatus. To suppose that such exists would appear to be a form of semantic

[28]See Sorensen (1988), esp. pp. 189-216, Williamson (1994), esp. chs. 7, 8.

[29]I note that, though Williamson endorses classical logic, the possibility of an epistemic explanation of the counter-intuitive nature of the cut-off point is, in principle, available for any standard semantics for vagueness.

[30]See Williamson (1994), 8.3.

mysticism.[31]

But set these matters aside. What is Williamson's explanation of the fact that the existence of a cut-off point is counter-intuitive? He argues that we find it so because we cannot imagine one; and we cannot imagine one, because we cannot know where it is.[32] Now, 'because' is transitive. If x is in the causal chain leading up to y, and y is in the causal chain leading up to z, then x is in the causal chain leading up to z. (If Johnny cannot go to the movies because his parents refused to give him his pocket money, and they refused to give him the money because he did not clean up his room, then he cannot go to the moves because he failed to clean up his room.) Hence, for Williamson we find the existence of a cut-off counter-intutive because we cannot know where it is. However, the mere unknowability of something does not explain why its existence is counter-intuitive. There are many things that we cannot know and whose existence we do not find puzzling at all. For example, there is a well-known model of the physical cosmos according to which the universe goes through alternating periods of expansion and contraction. In particular, the singularity at the big bang was just the end of the last period of contraction and the beginning of the current period of expansion. Suppose this is right. Then there must be many facts about what happened in the phase of the universe prior to the big bang—for example, whether there was sentient life. Yet all information about this period has been wiped out for us—lost in the epistemic black hole that was the big bang. Yet we do not find the existence of determinate facts before the big bang counter-intuitive. Indeed, we seem to have no problem accepting the thought that there are such things, though they be cognitively inaccessible to us, and ever will be so. That we cannot know the existence of something does not, therefore, explain why we find its existence counter-intuitive.

As we saw, Williamson derives his connection between counter-intuitiveness and unknowability of the cut-off from two other claims:

- it is counter-intuitive because we cannot imagine it
- we cannot imagine it because we cannot know it.

The preceding considerations do not tell us which of these statements is false. This may depend on what, exactly, Williamson intends in saying that one

[31] As Crispin Wright puts it in his detailed critique of epistemicism (1995). See also Horgan (1994), section 5.

[32] Williamson (1997), p. 218 ff.

can imagine something. However, at least *prima facie*, both of these claims are in trouble.

For the first: I cannot imagine what the suffering of being burnt at the stake feels like. Yet I do not find its existence counter-intuitive. Nor can I imagine the experience of what it feels like to drown; but I do not find it counter-intuitive to suppose that there is such an experience. For the second, I cannot know whether there are planets outside my light-cone; but I have no problem imagining such. Similarly, I have no way of knowing (this side of death) what an afterlife is like. But I can certainly imagine one.

Williamson's epistemic explanation of why we find the existence of a cut-off point counter-intuitive, does not, then, work.

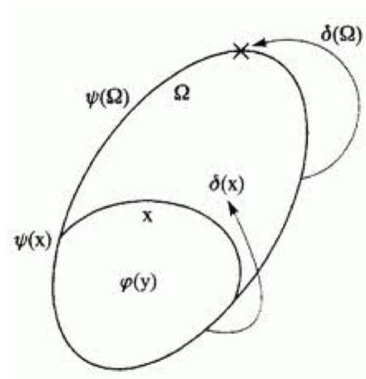
10 Inclosure Paradoxes

Before I finish, let me comment on one further, and important, aspect of the Sorites Paradox: its connection with the paradoxes of self-reference, such as the liar paradox and Russell's paradox.

There is a general structure that underlies the paradoxes of self-reference: they all fit the *inclosure schema*.[33] The schema arises when there is an operator, δ, and a totality, Ω (of the form $\{x : \varphi(x\}$, for some φ), which appear to satisfy the following conditions.[34] Whenever δ is applied to any subset, x, of Ω, of a certain kind—that is, one which satisfies some condition ψ—it delivers an object that is still in Ω (Closure) though not in x (Transcendence). If Ω itself satisfies ψ, a contradiction is forthcoming. For applying δ to Ω itself will then produce an object that is both within and without Ω, so that $\delta(\Omega) \in \Omega$ and $\delta(\Omega) \notin \Omega$. We may depict the situation as follows ($\times$ marks the contradictory spot—somewhere that is both within and without Ω):

[33] The inclosure schema was first proposed in Priest (1994). Since then, the major controversy over it has been whether Curry's paradox fits, or ought to fit, the schema. For the most recent round in the controversy, see the appendix of Priest (2017).

[34] I note that the conditions do not *actually* have to be true, just *prima facie* so. (See the second edition of Priest (1995), 17.2.) The inclosure schema is a diagnostic tool, not an argument for dialetheism.



Thus, consider Russell's paradox, for example. Ω is the set of all sets; $\delta(x) = \{y \in x : y \notin y\}$; and $\psi(x)$ is the vacuous condition, $x = x$. Or consider the Liar paradox. Ω is the set of all truths, $\psi(x)$ is 'x has a name', and $\delta(x)$ is a sentence, σ, of the the form $\langle\sigma\rangle \notin \dot{x}$ (where angle brackets are a name-forming operator, and $\dot{x}$ is a name of x). In each of these cases it is not difficult to show that the inclosure conditions appear to be satisfied.[35]

Now, the Sorites Paradox is an inclosure paradox, too.[36] Given our sorites sequence, Ω is the set of all a_is such that Pa_i. ψ is the vacuous condition, $x = x$. If $x \subseteq \Omega$ there is a maximum j such that $a_j \in x$. $\delta(x)$ is a_{j+1}. $a_{j+1} \notin x$, by construction (it's the first thing that is not P); and $a_{j+1} \in \Omega$ since Pa_j and a_{j+1} is next to a_j (that's just tolerance) The contradiction is that the first thing that is not P is P.

The fact that the Sorites Paradox is of a piece with the paradoxes of self-reference tells us that they should have the same kind of solution—the Principle of Uniform Solution. This does not mandate a dialetheic solution for sorites/self-referential paradoxes. However, elsewhere I have argued for a dialetheic solution to the paradoxes of self-reference.[37] And if such be correct, then a dialetheic solution to the Sorites Paradox is exactly what we should expect.

[35]Thus, in the case of Russell's paradox, suppose that $x \subseteq \Omega$. If $\delta(x) \in \delta(x)$ then $\delta(x) \notin \delta(x)$. So $\delta(x) \notin \delta(x)$. Hence $\delta(x) \notin x$, or it would be the case that $\delta(x) \in \delta(x)$. Hence we have Transcendence. Closure is true by Definition. In the case of the Liar paradox, suppose that $x \subseteq \Omega$. If $\langle\sigma\rangle \in \dot{x}$, then σ is true, so $\langle\sigma\rangle \notin \dot{x}$. Hence $\langle\sigma\rangle \notin \dot{x}$. So we have Transcendence. But since $\langle\sigma\rangle \notin \dot{x}$, σ is true. So we have Closure. For full details, see Priest (1995), Part 3.

[36]See Priest (2010a).

[37]E.g., Priest (2006), Part 1.

The fact that the Sorites Paradox is an inclosure paradox, together with the Principle of Uniform Solution, provide another argument against an epistemic solution to the sorites. There is no obvious way in which epistemicism can be deployed to account for the paradoxes of self-reference.[38]

Solutions which invoke truth value gaps, by contrast, are often mooted in the case of both the Sorites Paradox and the paradoxes of self-reference. I find such solutions to the paradoxes of self-reference unconvincing: they cannot happily handle so called "extended paradoxes".[39] More to the point here, I see no obvious way of deploying the gap-stratagem to explain why it is that we find the existence of a precise cut-off in a sorites sequence counter-intuitive. In particular, logics with truth value gaps standardly endorse *modus ponens* (and the substitutivity of identicals). They are therefore required to say that some of the major premises of the argument are false or, at least, untrue. This does not help an explanation of why we find the existence of a cut-off counter-intuitive one iota. It is simply a denial of tolerance (which a dialetheic approach endorses). Any symmetric duality between a gap solution and a glut solution breaks down at this point—if not others.

11 Conclusion

In this essay I have explained some of the core details of a dialetheic account of the Sorites Paradox, together with its attractions. I have not attempted a systematic comparison with other possible solutions, nor a systematic evaluation of the strengths and weaknesses of these. That would take much longer—indeed, in some ways it is what this book is about. Nor do I suppose that what I have said will end any debate. After hundreds of years of hibernation, the genie has finally come out of its bottle—and it is not going to go back in quietly. I think it is fair to say, though, that a dialetheic solution

[38]Horwich (199a), pp. 41-42, attempts an epistemic account of the Liar, claiming that the Liar sentence sentence is either true or false, though one cannot know which. This, however, is not his solution to the paradox, which is to give up certain instances of the T-schema (a completely *ad hoc* move, given Horwich's views on truth). And, epistemicism provides no explanation of which instances of the T-schema are not true, independent of the fact that they produce contradiction. For a fuller discussion, see Armour-Garb (2004). And even Horwich has not attempted an epistemic solution to the set-theoretic paradoxes.

[39]For a discussion of such paradoxes, see Beall (2008). It has certainly been argued by some that dialetheic approaches are also subject to the same problem, but I find the arguments unpersuasive—indeed, frequently confused. See Priest (2006), 20.3.

to the Sorites Paradox has so far been given less airplay than most other solutions. No doubt this is because dialetheism itself has been taken to be beyond the pale. That attitude is, I think, slowly changing. Even if few people currently accept dialetheism, it occupies a position in logical space which those not possessing blinkers[40] must engage with. If this essay generates a greater engagement than has hitherto happened in the context of the Sorites Paradox, it will have served its purpose.[41]

References

[1] Armour-Garb, B. (2004), 'Minimalism, the Generalization Problem and the Liar', *Synthese* 139: 491-52.

[2] Beall, J. C. (ed) (2008), *Revenge of the Liar: New Essays on the Paradox*, Oxford: Oxford University Press.

[3] Dummett, M. (1975), 'Wang's Paradox', *Synthese* 30: 301-24.

[4] Fine, K. (1975), 'Vagueness, Truth and Logic', *Synthese* 30: 265–300.

[5] Goguen, J. A. (1969), 'The Logic of Inexact Concepts', *Synthese* 19: 325–373.

[6] Hicks, R. D. (tr) (1925), *Diogenes Laertius: Lives of Eminent Philosophers*, Cambridge, MA: Harvard University Press.

[7] Horgan, T. (1994), 'Robust Vagueness and the Forced-March Sorites', *Philosophical Perspectives* 8: 159-88.

[8] Horwich, P. (1998), *Truth*, 2nd edn, Oxford: Oxford University Press.

[9] Hyde, D. (1997), 'From Heaps and Gaps to Heaps of Gluts', *Mind* 106: 641-660.

[10] Hyde, D. (2011), 'Sorites Paradox', in E. Zalta (ed), *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/sorites-paradox/.

[40] That is, blinders, if you speak North American English.

[41] I am grateful to the editors of this volume, Sergei Oms and Elia Zardini, for their comments on the first draft of this essay which certainly improved it.

[11] Hyde, D., and Colyvan, M. (2008), 'Paraconsistent Vagueness: Why Not?', *Australasian Journal of Logic* 6: 107–21.

[12] Jaśkowski, S. (1969), 'Propositional Calculus for Contradictory Deductive Systems', *Studia Logica*, 24: 143-57. Originally published in 1948 in Polish in *Studia Scientarium Torunensis*, Sec. A II, pp. 55-77.

[13] Keefe, R. (2000), *Theories of Vagueness*, Cambridge: Cambridge University Press.

[14] Keefe, R., and Smith, P. (eds) (1997), *Vagueness: a Reader*, Cambridge, MA: MIT Press.

[15] Kneale, W., and Kneale, M. (1962), *The Development of Logic*, Oxford: Clarendon Press.

[16] McGill, V. J., and Parry, W. T. (1948), 'The Unity of Opposites: a Dialectical Principle', *Science and Society* 12: 418-44.

[17] Plekhanov, G. (1941), *Fundamental Problems of Marxism*, ed D. Ryazanov, London: Lawrence and Wishart.

[18] Priest. G. (1994), 'The Structure of the Paradoxes of Self-Reference', *Mind* 103: 25-34.

[19] Priest, G. (1995), *Beyond the Limits of Thought*, Cambridge: Cambridge University Press; 2nd edn, Oxford: Oxford University Press, 2002.

[20] Priest, G. (2002), 'Paraconsistent Logic', pp. 287-393, Vol. 6, of D. Gabbay and F. Guenthner (eds), *Handbook of Philosophical Logic*, 2nd edn, Dordrecht: Kluwer Academic Publishers.

[21] Priest, G. (2003), 'A Site for Sorites', pp. 9-23 of J. C. Beall (ed), *Liars and Heaps: New Essays on Paradox*, Oxford: Oxford University Press.

[22] Priest, G. (2006), *In Contradiction*, 2nd edn, Oxford: Oxford University Press.

[23] Priest, G. (2008), *Introduction to Non-Classical Logic: From If to Is*, Cambridge: Cambridge University Press.

[24] Priest, G. (2010a), 'Inclosures, Vagueness, and Self-Reference', *Notre Dame Journal of Formal Logic* 51: 69-84.

[25] Priest, G. (2010b), 'Non-Transitive Identity', ch. 23 of R. Dietz and S. Moruzzi (eds), *Cuts and Clouds: Vagueness, its Nature and its Logic*, Oxford: Oxford University Press.

[26] Priest, G. (2014), *One*, Oxford: Oxford University Press.

[27] Priest, G. (2017), 'What *If*: the Exploration of an Idea', *Australasian Journal of Logic* 14, https://ojs.victoria.ac.nz/ajl/article/view/4028/3574.

[28] Ripley, D. (2005), *Sorting out the Sorites*, MA Thesis, University of North Carolina at Chapel Hill.

[29] Ripley, D. (2013): 'Sorting out the Sorites' ch. 18 of F. Berto, E. Mares, and K. Tanaka (eds), *Paraconsistency: Logic and Applications*, Dordrecht: Springer.

[30] Russell, B. (1923), 'Vagueness', *Australasian Journal of Philosophy and Psychology* 1: 84-92.

[31] Sorensen, R. (1988), *Blindspots*, Oxford: Oxford University Press.

[32] Sorensen R. (2012), 'Vagueness', in E. Zalta (ed), *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/vagueness/.

[33] Weber, Z. (2010), 'A Paraconsistent Model of Vagueness', *Mind* 119: 1025-1045.

[34] Weber, Z., and Colyvan, M. (2010), 'A Topological Sorites', *Journal of Philosophy* 107: 311-25.

[35] Williamson, T. (1994), *Vagueness*, London: Routledge.

[36] Williamson, T. (1997), 'Imagination, Stipulation and Vagueness', *Philosophical Issues* 8: 215-28.

[37] Wright, C. (1995), 'The Epistemic Conception of Vagueness', *Southern Journal of Philosophy* 33 (S1): 133-59.