The sorites paradox

The paradox we're discussing today is not a single argument, but a family of arguments. Here are some examples of this sort of argument:

1. Someone who is 7 feet in height is tall.
2. If someone who is 7 feet in height is tall, then someone 6'11.9" in height is tall.
3. If someone who is 6'11.9" in height is tall, then someone 6'11.8" in height is tall.

   ......

C. Someone who is 3' in height is tall.

The "..." stands for a long list of premises that we are not writing down; but the pattern makes it pretty clear what they would be. We could also, rather than giving a long list of premises 'sum them up" with the following **sorites premise**:

> For any height h, if someone's height is h and he is tall, then someone whose height is $h - 0.1''$ is also tall.

This is a universal claim about all heights. Each of the premises 2, 3, ... is an instance of this universal claim. Since universal claims imply their instances, each of premises 2, 3, ... follows from the sorites premise.

This is a paradox, since it looks like each of the premises is true, but the conclusion is clearly false. Nonetheless, the reasoning certainly appears to be valid.

Once we see this, it is easy to come up with other instances of the paradox:

1. 10,000 grains of sand is a heap of sand.
2. 10,000 grains of sand is a heap of sand, then 9999 grains of sand is a heap of sand.
3. 9999 grains of sand is a heap of sand, then 9998 grains of sand is a heap of sand.

   ......

C. 1 grain of sand is a heap of sand.

Once we see this, it is easy to come up with other instances of the paradox:

1. 10,000 grains of sand is a heap of sand.
2. 10,000 grains of sand is a heap of sand, then 9999 grains of sand is a heap of sand.
3. 9999 grains of sand is a heap of sand, then 9998 grains of sand is a heap of sand.

   ……

C. 1 grain of sand is a heap of sand.

Here the 'sorites premise" is:

> For any number n, if n grains of sand is a heap, then $n-1$ grains of sand is a heap.

Another example is:

1. A man with 1 hair on his head is bald.
2. If a man with 1 hair on his head is bald, a man with 2 hairs on his head is bald.
3. If a man with 2 hairs on his head is bald, a man with 3 hairs on his head is bald.

   ……

C. A man with 100,000 hairs on his head is bald.

Here the 'sorites premise" is:

> For any number n, if someone with n hairs on his head is bald, then someone with $n+1$ hairs on their head is bald.

There's also a special case of this sort of argument that has to do with our powers of perceptual discrimination. Suppose that we line up 10,000 color swatches, which range from bright red (swatch 1) and the beginning to bright orange at the end (swatch 10,000). It seems as though, with 10,000 swatches, there will be no discernible difference between any two adjacent swatches, so that every swatch will look the same as the one next to it in the series. We can then construct the following argument:

There’s also a special case of this sort of argument that has to do with our powers of perceptual discrimination. Suppose that we line up 10,000 color swatches, which range from bright red (swatch 1) and the beginning to bright orange at the end (swatch 10,000). It seems as though, with 10,000 swatches, there will be no discernible difference between any two adjacent swatches, so that every swatch will look the same as the one next to it in the series. We can then construct the following argument:

1. Swatch 1 is red.
2. If swatch 1 is red, then swatch 2 is red.
3. If swatch 2 is red, then swatch 3 is red.

   ......

C. Swatch 10,000 is red.

If we assume that each swatch looks the same as the one next to it, then this version of the sorites argument can be thought of as having a rather special sorites premise: For any objects x, y, if x looks the same as y and x is red, then y is red. This certainly seems hard to deny.

This is enough to show that the paradox does not arise from any very special feature of, for example, tallness, since the paradox can be recreated for many different predicates: not just ‘tall”, but also “bald”, ‘red”, etc.

One important feature of ‘tall” and the other predicates which can be used to generate sorites-type paradoxes is that they all admit of **borderline cases**: each of these predicates is such that there are things to which we aren’t sure whether or not the predicate applies, no matter how much we know about the thing.

It seems that a response to the sorites paradox will fall into one of three categories:

1. Rejecting the initial premise.
2. Rejecting one of the other premises, and/or the sorites premise.
3. Rejecting the validity of the argument.

The problem is that none of these looks initially promising.

Let’s consider the first option first: the option of rejecting the first premise of every sorites argument.

This is the simplest but also the most drastic response to sorites arguments. This involves, for example, denying that a 7-foot man is tall (and that he is tall for a man, and tall for a ...), that a man with one hair is bald, etc. Since we can set the values of the relevant quantities however we like in the first premise of a sorites argument, this fairly clearly involves denying that **anything** is bald, tall, etc. Yet more generally, this involves saying that no vague predicate — i.e., no predicate that admits of borderline cases — really applies to anything.

1. Rejecting the initial premise.
2. Rejecting one of the other premises, and/or the sorites premise.
3. Rejecting the validity of the argument.

Let's consider the first option first: the option of rejecting the first premise of every sorites argument.

This is the simplest but also the most drastic response to sorites arguments. This involves, for example, denying that a 7-foot man is tall (and that he is tall for a man, and tall for a ...), that a man with one hair is bald, etc. Since we can set the values of the relevant quantities however we like in the first premise of a sorites argument, this fairly clearly involves denying that **anything** is bald, tall, etc. Yet more generally, this involves saying that no vague predicate — i.e., no predicate that admits of borderline cases — really applies to anything.

This looks pretty clearly like a last-ditch solution; before adopting such an extreme view, we should want to see if there are any better options.

A natural thought is that there must be, and that the fault must lie with one of the premises in the sorites argument other than the first one. Consider, for illustration, the instance of the sorites argument concerned with baldness:

1. A man with 1 hair on his head is bald.
2. If a man with 1 hair on his head is bald, a man with 2 hairs on his head is bald.
3. If a man with 2 hairs on his head is bald, a man with 3 hairs on his head is bald.

......

C. A man with 100,000 hairs on his head is bald.

The simplest response to this argument other than rejecting the first premise is simply to reject one of the premises other than the first. (After all, there are a lot to choose from.)

One way to pursue this strategy is to simply say that there is, somewhere, a cut off between the bald and the non-bald men; and that the false premise is the one which crosses this crucial cut-off line. So, for example, the false premise might be #125:

125. If a man with 124 hairs on his head is bald, a man with 125 hairs on his head is bald.

1. Rejecting the initial premise.
2. Rejecting one of the other premises, and/or the sorites premise.
3. Rejecting the validity of the argument.

1. A man with 1 hair on his head is bald.
2. If a man with 1 hair on his head is bald, a man with 2 hairs on his head is bald.
3. If a man with 2 hairs on his head is bald, a man with 3 hairs on his head is bald.

   ……

C. A man with 100,000 hairs on his head is bald.

The simplest response to this argument other than rejecting the first premise is simply to reject one of the premises other than the first. (After all, there are a lot to choose from.)

One way to pursue this strategy is to simply say that there is, somewhere, a cut off between the bald and the non-bald men; and that the false premise is the one which crosses this crucial cut-off line. So, for example, the false premise might be #125:

125. If a man with 124 hairs on his head is bald, a man with 125 hairs on his head is bald.

The view that the right solution to the sorites paradox is to reject the “border-crossing” premise is called **the epistemic view** - since, on this view, the only real problem with a sorites argument is that we do not know which premise is the false one.

If the epistemic view is right, then it seems that men with 125 hairs on their head (or whatever the magic number is) who don’t want to be bald should be extremely careful: if they lose just one more hair, that will push them over the edge into baldness. This sort of thought makes many want to reject the epistemic view, since they doubt that there could be such a magic number.

Further, as noted above, if one premise of this sort is false, then it is fair to say that no one knows which premise it is. Moreover, it is hard to see what we could do to find out what it is. So facts about whether our word “bald” applies to someone with, say 130 hairs is forever unknowable. But is it plausible to think that there are unknowable facts of this sort about the application of our own words?

Presumably, words like “bald” have the meanings they do because of the way that we use them. But how could we use our words in ways which determined standards which even we not only don’t know, but couldn’t know?

1. Rejecting the initial premise.
2. Rejecting one of the other premises, and/or the sorites premise.
3. Rejecting the validity of the argument.

1. A man with 1 hair on his head is bald.
2. If a man with 1 hair on his head is bald, a man with 2 hairs on his head is bald.
3. If a man with 2 hairs on his head is bald, a man with 3 hairs on his head is bald.

   ……

C. A man with 100,000 hairs on his head is bald.

The most natural response to these questions, many think, is that the reason why we cannot know whether the sharp cut-off point for words like "bald" falls is that there is nothing of that sort to know: there just is no sharp cut-off point.

The intuitive idea is this: there is one group of people of whom it is simply true to say that they are bald. There's another group of people of whom it is simply false to say that they are bald — of these people, it is true to say that they are not bald.

But there are also some people in the middle. If you say that one of them is bald, you haven't said anything true; but you haven't said anything false, either. Just so, if you say that one of them is not bald, you haven't said anything true, but you haven't said anything false, either. The rules for applying the word "bald" just don't deliver a verdict for these people — it is "undefined" when it comes to them. This view is called a **truth-value gap** view, because it says that there is a gap between truth and falsity into which the "borderline" cases can be put.

When you think about the purposes for which we use the word "bald" and other vague terms, this can seem quite plausible. We want to be able to use the word to be able to distinguish one group of people, the bald ones, and to say of some other people that they don't belong to that group. But it's not as though we have a big interest in providing an exhaustive division of the world's people into two groups, the bald and the non-bald.

What does this have to do with the sorites argument? One thing that the proponent of truth-value gaps can say is that some number of the premises in a typical sorites argument will fail to be true. Consider, for example, the one we considered earlier:

125. If a man with 124 hairs on his head is bald, a man with 125 hairs on his head is bald.

1. Rejecting the initial premise.
2. Rejecting one of the other premises, and/or the sorites premise.
3. Rejecting the validity of the argument.

1. A man with 1 hair on his head is bald.
2. If a man with 1 hair on his head is bald, a man with 2 hairs on his head is bald.
3. If a man with 2 hairs on his head is bald, a man with 3 hairs on his head is bald.

   ......

C. A man with 100,000 hairs on his head is bald.

What does this have to do with the sorites argument? One thing that the proponent of truth-value gaps can say is that some number of the premises in a typical sorites argument will fail to be true. Consider, for example, the one we considered earlier:

125. If a man with 124 hairs on his head is bald, a man with 125 hairs on his head is bald.

Let's suppose that it is neither true nor false to say that someone with 124 hairs, or 125, is bald. Then this premise is an example of an "if-then" sentence both of whose constituent sentences are neither true nor false, but rather "undefined." The proponent of truth-value gaps might say that sentences of this sort are also undefined. Since these sentences are undefined rather than true, not all premises of the argument are true; perhaps this is enough to explain why the conclusion of the sorites argument is false.

But this approach to sorites arguments also has some curious features. Consider, for example, the sentence

Either it is raining or it is not raining.

Ordinarily, we think of sentences of this sort as logical truths — they are true no matter what the weather. We think the same of

If it is raining, then it is raining.

But on the present approach, it looks like these sentences can, in some cases, be untrue. For presumably there are borderline cases of rain; "is raining" is a vague predicate. Suppose that (as often in South Bend) it is a borderline case of rain. Then both "It is raining" and "It is not raining" will be undefined, rather than true or false; but then it looks like neither of the above sentences will be true.

1. Rejecting the initial premise.
2. Rejecting one of the other premises, and/or the sorites premise.
3. Rejecting the validity of the argument.

1. A man with 1 hair on his head is bald.
2. If a man with 1 hair on his head is bald, a man with 2 hairs on his head is bald.
3. If a man with 2 hairs on his head is bald, a man with 3 hairs on his head is bald.

   ......

C. A man with 100,000 hairs on his head is bald.

Let's suppose that it is neither true nor false to say that someone with 124 hairs, or 125, is bald. Then this premise is an example of an "if-then" sentence both of whose constituent sentences are neither true nor false, but rather "undefined." The proponent of truth-value gaps might say that sentences of this sort are also undefined. Since these sentences are undefined rather than true, not all premises of the argument are true; perhaps this is enough to explain why the conclusion of the sorites argument is false.

A different sort of problem comes from a seeming asymmetry between the following two sentences (suppose that Bob has 125 hairs on his head — or whatever number you think would make him a "borderline case" of baldness):

> If Bob is bald, then with one less hair he would still be bald.
>
> If Bob is bald, then with one more hair he would still be bald.

It looks like, intuitively, the first one is definitely true; the second looks quite plausible, but surely not as clearly true as the first. In any case, there appears to be a definite asymmetry between them. But what would the proponent of truth-value gaps say about these "if-then" sentences?

This might lead us to think that we want some approach to the sorites paradox which captures the idea that there are "middle cases" for which words like "bald" are undefined, but which avoids the problematic results discussed above. This is the aim of the proponent of **supervaluationist** approaches to vagueness.

The core idea behind supervaluationism is as follows: as above, there are a host of middle cases of thinly haired men which are such that the rules for "bald" don't dictate that it is true to say of them that they are bald, but also don't dictate that it would be false to say this of them. So, in a certain sense, it is "up to us" to say what we want about such cases. Let's call the act of 'drawing the line" between the bald and non-bald a **sharpening** of "bald." Then we can say that there are many possible sharpenings of "bald" which are consistent with the rules governing the word.

1. Rejecting the initial premise.
2. Rejecting one of the other premises, and/or the sorites premise.
3. Rejecting the validity of the argument.

The core idea behind supervaluationism is as follows: as above, there are a host of middle cases of thinly haired men which are such that the rules for “bald” don’t dictate that it is true to say of them that they are bald, but also don’t dictate that it would be false to say this of them. So, in a certain sense, it is “up to us” to say what we want about such cases. Let’s call the act of ‘drawing the line” between the bald and non-bald a **sharpening** of “bald.” Then we can say that there are many possible sharpenings of “bald” which are consistent with the rules governing the word.

Then the supervaluationist can give the following definitions:

A sentence is true if and only if it is true with respect to every sharpening.

A sentence is false is and only if it is false with respect to every sharpening.

A sentence is undefined if and only if it is true with respect to some sharpenings, and false with respect to others.

Since the supervaluationist believes that some sentences are undefined rather than true or false, this is a version of the truth-value gap family of solutions to the sorites paradox — but it seems to handle the problem cases for truth-value gap views discussed above.

Consider first the supervaluationist treatment of this sentence:

Either it is raining or it is not raining.

If we are in a “borderline raining” scenario, will this be true, false, or undefined?

How about our pair of sentences (where Bob is borderline bald):

If Bob is bald, then with one less hair he would still be bald.

If Bob is bald, then with one more hair he would still be bald.

Will the supervaluationist see a difference between them, as it seems we should?

How does this view escape the sorites paradox? Many premises in the typical instance of the sorites paradox will be true on some sharpenings, but false on others. So, as above, some of these premises will be undefined; this makes room for the view that the reasoning is valid and the conclusion false, which is the result for which all truth-value gap approaches aim.

1. Rejecting the initial premise.
2. Rejecting one of the other premises, and/or the sorites premise.
3. Rejecting the validity of the argument.

How does this view escape the sorites paradox? Many premises in the typical instance of the sorites paradox will be true on some sharpenings, but false on others. So, as above, some of these premises will be undefined; this makes room for the view that the reasoning is valid and the conclusion false, which is the result for which all truth-value gap approaches aim.

But think again about this sentence:

> Either it is raining or it is not raining.

As noted above, the supervaluationist treats this as true. But according to the supervaluationist **neither** of the following sentences will be true, which seems odd:

> It is raining.
> It is not raining.

And consider this sentence:

> There is some number such that if you have that number of hairs you are not bald, but if you have one fewer you are bald.

Will the supervaluationist count this sentence as true, false, or undefined? Is there any sharpening on which it is false?

What is odd is that the supervaluationist will say that **every** sentence of the following form is false:

> If you have N hairs you are not bald, but if you have N-1 hairs you are bald.

This means that there are true generalizations of the form ‘something is …” for which there are no corresponding particular truths of the form “N …” But this seems weird, since normally we think that what it is for such a generalization to be true is for there to be at least one corresponding particular truth.

Futher, all truth-value gap approaches — whether the simple version or the supervaluationist version — face the problem of **higher-order vagueness**.

This problem is that just as there are borderline cases between “bald” and “not bald”, there are also borderline cases between cases where “bald” applies **and those borderline cases**. But both simple truth-value gap theories and supervaluationist theories assume that there is a dividing line between the cases where “bald” applies and the cases in which it is undefined. (That is needed to state the theory.)

1. Rejecting the initial premise.
2. Rejecting one of the other premises, and/or the sorites premise.
3. Rejecting the validity of the argument.

Futher, all truth-value gap approaches — whether the simple version or the supervaluationist version — face the problem of **higher-order vagueness**.

This problem is that just as there are borderline cases between “bald” and “not bald”, there are also borderline cases between cases where “bald” applies **and those borderline cases**. But both simple truth-value gap theories and supervaluationist theories assume that there is a dividing line between the cases where “bald” applies and the cases in which it is undefined. (That is needed to state the theory.)

One way to see this is to define ‘definitely bald” as a predicate that applies to everything to which “bald” applies, and does not apply to everything to which “not bald” applies and everything with respect to which “bald” is undefined. The idea that there is higher-order vagueness can then be expressed as the idea that ‘definitely bald” is itself vague.

The epistemic theorist might then reply to the truth-value gap approaches by saying that if they posit a sharp dividing line between the cases to which “bald” applies and the cases for which it is undefined, why not simplify the theory and posit such a sharp dividing line between “bald” and “not bald”? This would then avoid all of the problems discussed above. Of course, then we’d be stuck with the idea that there’s some number of hairs such that, if you have that number you are not bald, but if you lost just one, you would be bald.

All of the responses we have been discussing - the epistemic view, the simple truth-value gap view, and the supervaluationist view - are versions of a type 2 solution: all reject sorites arguments as unsound because of their containing an untrue premise other than the first one. Could we plausibly take the third route, and say that the problem with the sorites arguments is that they are **invalid**?

A quick look at the form of a sorites argument shows that this is going to be tough to do:

1. A man with 1 hair on his head is bald.
2. If a man with 1 hair on his head is bald, a man with 2 hairs on his head is bald.
3. If a man with 2 hairs on his head is bald, a man with 3 hairs on his head is bald.

   ......

C. A man with 100,000 hairs on his head is bald.

It looks we are just moving from “P” and “If P, then Q” to “Q” — and this is about the most obviously valid form of reasoning there is.

1. Rejecting the initial premise.
2. Rejecting one of the other premises, and/or the sorites premise.
3. Rejecting the validity of the argument.

All of the responses we have been discussing - the epistemic view, the simple truth-value gap view, and the supervaluationist view - are versions of a type 2 solution: all reject sorites arguments as unsound because of their containing an untrue premise other than the first one. Could we plausibly take the third route, and say that the problem with the sorites arguments is that they are **invalid**?

A quick look at the form of a sorites argument shows that this is going to be tough to do:

1. A man with 1 hair on his head is bald.
2. If a man with 1 hair on his head is bald, a man with 2 hairs on his head is bald.
3. If a man with 2 hairs on his head is bald, a man with 3 hairs on his head is bald.

   ......

C. A man with 100,000 hairs on his head is bald.

It looks we are just moving from "P" and "If P, then Q" to "Q" — and this is about the most obviously valid form of reasoning there is.

However, this is rejected by the friend of **degrees of truth**. According to this view, the sentences in a sorites series are not simply true or false, but rather true to a certain degree.

How can this help with an instance of the sorites paradox? The idea would be that a sentence like

A 7' tall man is tall for an adult person.

is true to a very high degree — Let's say, degree .95. However, a sentence like

A 6'11" tall man is tall for an adult person.

is true to a slightly lesser degree — say, degree .94. Now consider a typical premise in a sorites argument, like

If a 7' tall man is tall for an adult person, then a 6'11" tall man is tall for an adult person.

This is an "if-then" statement in which the "if" part is more true than the 'then" part. So the intuitive idea is that since we are going from a more true statement to a less true one, the whole "if-then" statement is true to some degree less than 1. It is more true than most statements, surely — but not perfectly true.

On this view, what should we expect when we have a very long string of "if-then" statements, none of which are perfectly true? How might this help explain how the reasoning in the sorites paradox can lead us from (almost complete) truth to (almost complete) falsity?

1. Rejecting the initial premise.
2. Rejecting one of the other premises, and/or the sorites premise.
3. Rejecting the validity of the argument.

However, this is rejected by the friend of **degrees of truth**. According to this view, the sentences in a sorites series are not simply true or false, but rather true to a certain degree.

There are three central problems for this sort of solution to the sorites paradox. The first is simply making sense of what degrees of truth mean; intuitively, truth is saying how things are, and falsehood is saying how they aren't — but what is truth to degree 0.61 even supposed to mean?

The second concerns certain compound statements. Suppose that the following two claims are both true to degree 0.5:

It is raining.
It is not raining.

It is very plausible that the truth-values of statements of the form "P and Q" and "If P, then Q" just depend on the truth-values of the constituent sentences, P and Q. But then it follows that the degree-theorist will have to treat the following pairs of sentences as having the same truth-value:

If it is raining, then it is raining.	If it is raining, then it is not raining.
It is raining and it is raining.	It is raining and it is not raining.
It is raining or it is not raining.	It is raining or it is raining.

Could this be right?

Finally, it seems like the degree-theorist, like the proponent of truth-value gaps, faces the problem of higher-order vagueness. The degree-theorist, after all, uses predicates like 'true to degree 1." But what is the cut-off point between those adult human beings of whom it is true to degree 1 to say that they are tall, and those of whom it is true to merely degree 0.99 to say this? It does not seem that this question has an easy answer. But this means that the sorites paradox can be recreated in the framework which the degree theorist designs to solve it. Consider a sorites argument with sorites premise: "If it is true to degree 1 to say that someone of X height is tall, then it is true to degree 1 to say that someone of X-1 height is tall." Can the degree-theorist solve this paradox by talking about the degree of truth enjoyed by various claims about the degree to which certain other statements are true? Will this procedure of ever higher-level talk of degrees of truth ever come to an end?

The sorites paradox has no easy solution.