Diagnosing Sorites arguments*

Robert Stalnaker

Received: 03/05/2018
Final version: 22/06/2018

BIBLID 0495-4548(2018)33:3p.509-520
DOI: 10.1387/theoria.19625

ABSTRACT: This is a discussion of Delia Fara's theory of vagueness, and of its solution to the sorites paradox, criticizing some of the details of the account, but agreeing that its central insight will be a part of any solution to the problem. I also consider a wider range of philosophical puzzles that involve arguments that are structurally similar to the argument of the sorites paradox, and argue that the main ideas of her account of vagueness helps to respond to some of those puzzles.

Keywords: vagueness, Sorites paradox, context-dependence, supervaluations, epistemicism.

RESUMEN: Esta es una discusión sobre la teoría de la vaguedad de Delia Fara, y de su solución a la paradoja de Sorites, criticando alguno de los detalles de su aproximación, pero de acuerdo en que la intuición central es parte de cualquier solución al problema. También considero un rango más amplio de puzles filosóficos que incluyen argumentos estructuralmente similares al argumento de la paradoja de Sorites, y defiendo que las ideas principales de su aproximación a la vaguedad ayudan a responder algunos de esos puzles.

Palabras clave: vaguedad, paradoja de Sorites, dependencia de context, supervaluaciones, epistemicismo.

The Sorites paradox, like the Liar, has an ancient origin, and like the paradoxes about truth it continues to provide both puzzlement and the opportunity to clarify philosophical issues about the relationship between our representations and what they are about. Delia Graff Fara's paper, 'Shifting sands: an interest-relative theory of vagueness', is a rich and insightful contribution to this literature. It is a paper with a complex dialectic (As she said, the solution to the sorites that she proposes "will unfold in layers.") and it is easy to get lost in its details. One of the reasons this is easy is that the details are so interesting. Fara was an Austinian philosopher with a gift for, and a love of, subtle linguistic distinctions. The paper is full of apt examples that reveal the complexity of the phenomena, but that also help to motivate an argument that illuminates the general shape of the problem. While I am not satisfied by every element in the theory she developed, I think her central insight needs to be

* One benefit of having really talented students as dissertation advisees is that one gets to take pride in their accomplishments, but a more important benefit is that one learns so much from working with them. I have long admired Delia's work on vagueness, but I find new insights whenever I return to it, and new applications of her clarifying ideas. Her death is a great loss to philosophy, but she leaves us a valuable body of work.

a part of any solution of the problem of vagueness, and that it also helps to point to a solution of some other paradoxes that have a similar form.

Here is my plan for this paper: In section 1, I will sketch the structure of the sorites argument, and describe Fara's criticisms of some responses to it. In section 2, I will sketch the main idea of her proposed solution to the paradox, agreeing that this idea should be central to a full solution, but expressing some reservations about some aspects of her account. In section 3 I will point to a broader category of arguments that have a sorites-like structure, but that do not necessarily involve vagueness, and that are used to support different conclusions. I will look at one example of an argument of this kind, a puzzle about knowledge and chance that has recently been posed, and argue that the central insight of Fara's solution to the paradox about vagueness should help to provide a response to this puzzle.

1. The sorites paradox, and two proposed strategies for a solution

We have an instance of a sorites paradox if we have a predicate F for which the following three premises all seem plausible:

(1) There are individuals a and b such that Fa, but ~Fb.
(2) There is a binary relation R meeting this condition: $(x)(y)((Fx\&Rxy) \rightarrow Fy)$
(3) There is a finite sequence of individuals, $x_1 \ldots x_n$ such that $x_1 = a$ and $x_n = b$, and for all k from 1 to n-1, Rx_kx_{k+1}.

But the three premises are inconsistent with each other.

(1) is just the assumption that there are clear cases of things that are F, and clear cases of things that are not. (2) is a kind of tolerance principle that seems to hold for vague predicates, perhaps even to be what constitutes the vagueness of the predicate. R is a kind of relevant similarity relation, so (2) says that things that are similar enough to something that is F must also be F. The denial of (2), $(\exists x)(\exists y)(Fx\&Rxy\&{\sim}Fy)$, is what Fara calls the 'sharp boundaries' claim. (3) is a kind of plenitude principle. While in practice there might be gaps in an actual sequence (For example, the girls on the basketball team might in fact divide into those that are five feet tall or less, and those that are five and a half feet tall or more), there *could* be a sequence of girls similar enough in height to fill in the gap.

Fara began by considering two responses to the paradox (Kit Fine's supervaluation theory, and Timothy Williamson's epistemicism), both of which deny premise (2), and so accept the sharp boundaries claim. She agrees that this is the premise to be denied, and most of her discussion is about ways of defending the sharp boundaries condition, but I will argue her own strategy for responding to the problem should lead us to look more carefully at the third premise of the argument—the plenitude principle.

Fara poses three questions that a defender of the sharp boundaries claim needs to answer:

(I) **The semantic question**: How is the sharp boundaries claim compatible with the fact that vague predicates have borderline cases? (In effect, this is asking, "what, on your account, is a borderline case?")
(II) **The epistemological question**: If the sharp boundaries claim is true, why are we unable to identify the point at which the boundary is located?

(III) **The psychological question**: If premise (2) (the tolerance principle) is false, why were we inclined to accept it in the first place? In particular, why are we inclined to believe of each point in the sequence, that the cut-off point is not there?

I would like to add a fourth question to this list:

(IV) **The metaphysical question**: If there is a fact of the matter about where the cut-off point is located, what are the facts about our use of the predicate, or about the things to which the predicate is applied that ground the fact that the cutoff point is where it is?

The supervaluation theory says that statements involving vague predicates are true if and only if they are true for all admissible ways of extending the interpretation of those predicates so that they have precise boundaries, and false if false for all admissible ways. This account has a clear answer to the semantic question: borderline cases are individuals where it is neither true nor false that the predicate applies. One might be temped to think that this account also has a satisfactory answer to the epistemological question for the following reason: The account implies that the sharp boundaries claim is true since every admissible precisification will have a sharp cutoff, but there is no fact of the matter about where that boundary is, so it is easy to see why we are unable to identify that point. But as Fara notes, the supervaluation theory is still committed to a sharp line between clear cases and borderline cases, and she argues that it has no satisfactory answer to the question why we are unable to identify *that* point.

On the epistemicist theory, borderline cases are explained as cases where there is a sharp cutoff point for the application of the predicate, but it is unknowable. That is the epistemicist's answer to the semantic question, what is a borderline case? In a sense the account gives a straightforward answer to the epistemological question since it is constitutive of being a borderline case that we are unable to identify it. But one would like to hear more about just what prevents us from knowing the semantic facts that determine where the sharp boundary lies.

2. *Fara's strategy for a solution*

It is the psychological question that is crucial to Fara's strategy for responding to the problem. She argues that neither the epistemicist nor the supervaluationist has an answer to this question, though she takes this not as a reason for thinking either of those views is false, but only a reason for thinking they need to be supplemented. Her response to the problem involves the thesis that the boundary between cases that satisfy the predicate and those that do not is both context-dependent and interest relative in a way that ensures that the boundary is not where we are looking for it. That is, the boundary will shift away from any point that is salient, given the interests of those in the context in which the predicate is being applied. She makes a persuasive case for the context and interest dependence of vague expressions with examples that both motivate the thesis and help to fill in details about the roles that salience and practical interests play in the interpretation of vague expressions. Let me look at just one of her more playful and arresting examples, which helps to illustrate the interplay of interest, decision, and knowledge in the interpretation of a simple predicate, in context. But I will suggest (taking the example more seriously than was probably intended) that what she said about it is not quite right.

> "An eccentric art collector . . . reserves one room for her paintings that contain just red pigments, and . . . another for her paintings that contain just orange pigments. One day she is presented as a gift a painted color spectrum ranging from primary red on one end to orange on the other." Presumably, this collector normally doesn't keep paintings that contain both red and orange pigments, but in this case she decides to cut the painting into two parts so she can keep both (in the separate rooms) without violating her policy. "Now if she cuts without thinking, (…) she will most likely cut in just the right place." But "if the decision about where to cut is labored, in contrast, the collector will likely find herself unable to locate the boundary (…), the pigments on either side of any proposed cut being too similar." (Graff-Fara 2000, 59)

The suggestion seems to be that the haphazard choice will get it right because the collector is not just judging where a cutoff line is, but is *deciding* where it shall be, and so whatever choice is made (within a certain range) that choice will get it right. If that were the case, then there would be no reason for the choice to be labored—she could make it deliberately, even while recognizing that the choice is to some extent arbitrary. But the cutoff point may not be so arbitrary in this example, since there are already two rooms full of paintings, and those paintings are part of a context that puts severe constraints on where the line between red and orange can be correctly drawn. Given her eccentric interests, this collector might well worry that by inadvertence, some pigments in paintings in the orange room are redder than some pigments in paintings in the red room, and given the penumbral constraint (If a is red, and b is redder than a, then b is red.), she might already have gotten it wrong in locating some of the paintings. But even if she has done a successful job, so far, in locating the paintings in the appropriate room, the choice about where to cut the new painting might get it wrong, whether the choice is haphazard or labored. Still, I think it is right to recognize that in the case where the application of a concept is context sensitive and interest dependent, correct application conditions may be partly a matter of decision. We get to decide, to some extent, what our interests are, and what we shall presuppose in the contexts in which we apply the language we are using. But the fact that the sharp line between red and orange, in this kind of context, may in part be open to decision raises a problem for the sharp boundaries claim. To continue with the story of the eccentric collector, let's assume that there a significant gap between the reddest pigment represented in the orange room and the orangest pigment represented in the red room. She thus has some leeway in her choice of where to cut the new painting. But suppose she reconsiders the idea of cutting the painting and reframing the resulting two parts, deciding instead to hang it in the corridor between the two rooms (with the redder end closer to the red room). Now we ask, is there a fact of the matter about where the sharp line between red and orange is, in the context provided by the story, and given the interests of the eccentric collector? Could we say that it is where she *would* have chosen (perhaps haphazardly) to cut the painting, had she decided to cut it? I don't think it is plausible to think that there is a fact of the matter about where she would have cut, particularly if she would have made the cut haphazardly.

In this modified example, the context is different from the context in which she decides to cut the painting, and the extension of the interest-relative predicates ('red' and 'orange') may be different than they are in the original case, where attention is focused on where to cut the painting. So we cannot assume that the place where the line would have been if she had chosen (perhaps arbitrarily) where to cut the painting is the same place where the line is in the case where she decides not to cut it. But Fara does seem to be assuming that there is a fact of the matter in each case about the extensions of the predicates, given the context and the interests of the person ascribing them. If an arbitrary decision could have determined the extension in the case where the painting is cut, is it reasonable to assume that in the case where no such decision is made, there are facts that will determine where the line will be? Perhaps we should ask, not where she would have cut the painting, if she had, but instead consider, of each possible painting that she might have collected and that might have been put in either room (given the penumbral constraint, and the paintings already in the rooms), which room would she have put it in? It seems even less plausible that there is a fact of the matter about the answers to all of these counterfactual questions.

The general point is that it is not clear that the defender of the sharp boundary claim can give a satisfactory answer to the metaphysical question, which would require some account of the kind of fact about the situation that determine where the sharp boundary is located. It is not enough to say that the boundary is always somewhere else than where we are looking. The 'we' here refers to the participants in the context in which the predicate is being applied, with the interests that they have in making the distinctions they make. But we the theorists, looking at a context from outside, can engage in what might be called 'pragmatic assent', asking exactly where, among the various places that the participants of the context are not looking, the sharp line is to be located. The theorists are of course in a context of their own, but they are allowed to look at a context they are not in, and to focus on features of the situation that the participants of the context in question are ignoring. Knowing that the sharp cutoff is at one of the points that is being ignored is not enough to tell us which of those points is the right one.

Of course predicates are applied, in a particular context, in order to distinguish between the members of a particular population of things, and it is a familiar point that the things in such population may in fact divide into groups that are isolated from each other with respect to the relation used to order a sorites series. Suppose there are some books on the right side of a shelf, all close enough to each other in shade to be part of a sorites-like series from blue to green. On the left side of the shelf the books are also similar enough to each other to be part of the same series, but they are all closer to the green end, and there is a big difference between the closest to green in the first set and the closest to blue in the second set. There will be no problem about how to understand, in that context, a request to fetch the blue books on the shelf. But suppose one asks the counterfactual question, where should the line be drawn (*in this context*) between the members of a sequence of *possible* books that would fill out the sorites series? Suppose there is an actual book, not on the shelf but on a bedside table next door, that is a color that falls in the intermediate range (I am supposing that this book is of a shade that would be judged blue in some contexts, and green in others.) Must there be a fact of the matter (perhaps unknown and unknowable) about whether *that* book is blue or green (with respect to the context described)? I think it would be carrying bivalence too far to require a positive answer to this question. The supervaluationist can do somewhat better with this very familiar kind of case, since there is a knowable and salient line between the colors of the books on the shelf (which are all of the books of interest in the context) that are clear cases of blue, or of not blue, and the colors of books not of interest in the context, for which no answer is required. The task of the use of sentences in a context defined by information that is common ground is to express propositions that distinguish unambiguously between the possibilities compatible with that context. The sentences used will usually express propositions that also distinguish between some possibilities beyond those compatible with the context, but that is not required to do the main job that speech acts are designed to do. Sentences that contain vague predicates will do this job so long as, given the presuppositions of the context (which includes presuppositions about the interests of the participants in the discussion), they distinguish unambiguously between the objects that are under discussion in the context. (Cf. Rayo 2008)

How do these remarks bear on the sorites argument? They do not, I should emphasize, provide a solution, but the recognition of the context dependence of vague predicates does give us reason to look more carefully at premise (3) of the argument: that there is a sequence of R-related individuals connecting a clear case where the predicate F applies to clear cases where the predicate fails to apply. Given the assumption of context dependence, we need to think of the argument as given for a particular context, since both the relevant relation R and the domain of individuals that are candidates for the application of the predicate will be determined relative to the context. The upshot of the example of a sorites series with a gap is that there will be many contexts for which premise (3) is false, and for those contexts (which probably includes most normal situations) there may be no problem with the coherent application of a vague predicate. If this is right, then this upshot takes *some* of the teeth out of the sorites paradox by denying that it provides a general challenge to the coherence of the application of vague predicates.

The reason the paradox is not solved by these considerations is that one can construct contexts in which both premises (2) and (3) seem to be true for a paradigm case of a vague predicate, and a binary similarity relation. Fara was clear that a solution requires that even in a context where we have a full sequence of individuals, all similar enough to each other to make premise (2) plausible, we will need to find at least one place in the sequence where the similarity is not relevant, or salient, in the context, and given the interests.

She considers two responses to this problem: the first is to argue (using a sorites series with the predicate 'tall' as the example) that in any specific context, "there cannot be enough *salient* similarities. When we are confronted with a sorites series for 'tall', although it is salient that each adjacent pair in the series is very similar, it is not the case that each adjacent pair in the series is such that it is salient that they are similar. There are too many pairs for us to actively entertain each similarity." The second response is to reject the assumption that "on any given occasion, there will be a standard in use for the vague predicate that satisfies all of my constraints. In a situation in which there are too many salient similarities, no standard could satisfy all four constraints" (Graff-Fara 2000, 71). The first response is the ambitious one, and Fara preferred it, but I am not convinced. In fact, I am not sure exactly how to understand it.

I take the general idea of this response to be something like this: when a person is presented with a putative sorites series, and is assessing the argument, he must begin by focusing on one part of the sequence, and working from there. Usually one starts at one end of the sequence, focusing on the clear case of the application of the predicate, and the similarity relation between it and the next in the sequence. The relationships between other individuals in the sequence are not salient at that point in the assessment. As one marches through the sequence, the context changes, with different pairs, and the relation between them, becoming salient. By the time one gets to the middle of the series, where the similarity between adjacent members at that point becomes salient, the similarities considered earlier are no longer salient, so while you can conclude (from premise (2)) that *if* individual x_{17} is F, then so is x_{18}, you are no longer in a position to be confident of the conclusion previously reached (in a slightly different context) that x_{17} is F. Because the context is continually shifting as we go, with the relation of salient similarity changing, the reductio is based on equivocations.

Before looking at why I don't find this response persuasive, let me ask, is it premise (2) or (3) that is false, on this diagnosis? The answer will depend on how we understand the relation R in the premises, and I see three different possibilities. Suppose R* is the underlying similarity relation, defined independently of salience (in Fara's tallness example, *being 1 mm apart in height*), and that S is the salience property, applied to pairs of individuals. On the first interpretation, the relation is 'being *saliently* similar', and that relation is defined as holding between things that are both similar and salient. That is, $Rxy =_{df} (R^*xy \& S(\langle x,y\rangle))$. On this interpretation, it is premise (3) rather than premise (2) that is false.

The second interpretation also takes *salient similarity* to be the relation in the premises, but this interpretation puts salience and similarity together in a different way. It says that R means *being the same for present purposes*, where two things are distinguishable for present purposes only if the similarities and differences between those two things are salient, and they are the same for present purposes if and only if they are not distinguishable for present purposes. That is, we should define R in terms of R* this way: $Rxy =_{df} (S(\langle x,y\rangle) \rightarrow R^*xy)$. On this interpretation it is premise (2) that is false, but on this interpretation the label for the denial of (2), $\exists x\exists y(Fx\&Rxy\&{\sim}Fy)$ ('the sharp boundaries claim'), is wholly inappropriate since the label seems to imply, falsely, that there is a sharp point of transition (even if one that is not salient). But an x and y that instantiate the so-called 'sharp boundaries' claim, using the second definition, can be *any* two individuals that are not salient in the context in question. So for example, while we are focusing on the difference between being 1.222 meters tall and being 1.223 meters tall, a pair of things, one that is 1.500 meters tall and the other that is 2.000 meters tall will be the same for present purposes.

On the third interpretation, R is just R*. Salience does not enter into the sorites argument at all. Its only role is to explain why we are tempted to accept the second premise of the argument. On this interpretation, 'the sharp boundaries claim' is only slightly misleading. As Fara suggested, the boundary is in a sense not sharp, since it is out of focus, but the rejection of premise (2) does imply, on this interpretation, that at each moment in the extended examination of the sorites argument there is exactly one particular place, out of view, such that the boundary is there. Compare the street magician who somehow ensures that the pea is not under the cup we choose to look under. It is always somewhere else, and of course some other *particular* place—under a particular one of the other cups. It is just that we don't know which one, and it may shift to a different cup when we take another guess. But in the case of the sorites series. is it really plausible to think that there is a fact about the short-lived context in which we are considering, say, the similarity between x_{17} and x_{18} that somehow determines a particular non-salient pair—perhaps x_{29} and x_{30}, where the transition from not tall to tall takes place? This is the metaphysical question, which I think is particularly hard to answer with the kind of fragile and shifting contexts that this response depends on. The other two interpretations do not have the same problem with this question.

I still find this response unpromising even if we interpret Fara's favored response in one of the first two ways, because it makes the context more fragile than is plausible. There should be a context for the whole argument, one in which the assessor of the argument need not focus in on a particular pair for the similarity of the members of that pair to be salient. The person assessing the argument recognizes that the members of each adjacent pair in the series are similar in exactly the same way. As Fara says, it is salient that they are similar, and that should be enough to ensure that the similarity of the members of each adjacent pair is salient and relevant.

Fara's alternative response, however, I find persuasive. The idea here is to reject the presupposition that every possible context must allow for a coherent application of a given vague predicate. One allows that there may be contexts in which there is a sequence of individuals and a relation R that together satisfy both premises (2) and (3), but claims that those are contexts in which the predicate has no coherent application. As Fara emphasized, the interests of the participants in the context should be expected to help provide the line between the cases where the predicate applies and the cases where it does not, but there is no reason to suppose the participants will always have interests that do this job.

What might the interests be that are relevant in a case where there is an apparent sorites series? Suppose Alice and Bert are in a context like the one described above, with books on a shelf, ordered by color, with a clear case of blue on one end and a clear case of not-blue on the other, but with a salient gap part way through. Alice asks Bert to bring her the blue books on the shelf. Her purpose is to get a certain set of books brought to her, and if she knows that there is a gap in the sequence, then she knows that Bert will be in a position to tell which books she wants. (Why she wants just that set of books brought to her doesn't matter. Perhaps she is looking for a particular book, but has forgotten its title, and remembers only that it is on the shelf, and on the right side of the gap.) But now change the example so that there is no gap. Now we have a standard sorites series, and in this case, there is no set of books that, given the semantics for the predicate plus the context plus the facts, will be determinately picked out by the phrase, 'the blue books on the shelf'. Alice cannot reasonably expect that Bert will be able to tell just which set of books will satisfy her request. And it would not seem right to say that there *are* in fact correct necessary and sufficient conditions for doing what she asked, but that Bert and we are ignorant as to what they are. I suppose Alice might have had a certain particular set of books in mind, but if she knew that the set of books on the shelf formed a full sorites series, she would have been misunderstanding the semantics and pragmatics of vague expressions if she thought that she had requested that Bert bring her just that set. The situation is similar to a case where Alice, knowing that there are two books on the table, asks Bert to bring her *the* book on the table. She might have a particular one of them in mind that she wants brought to her, but she misunderstands the semantics and pragmatics of singular definite descriptions if she thinks she can ask for it, in that context, in that way.

3. *Sorites beyond vagueness*

There is a variety of arguments with a structure that is very similar to the argument of the sorites paradox, but where vagueness does not seem to be the source of the problem. Even if we could defuse the paradox about vagueness by saying, as suggested above, that a simple sorites series is a case where vague predicates cannot be coherently applied, it is not clear that an analogous response is available in most of these other cases. I will look in detail at just one example of the class of arguments I have in mind, suggesting that Fara's 'shifting sands' idea can help to resolve some of the problems it poses. But first I will just mention some of the others to show the diversity of philosophical issues to which this kind of argument has been applied.

First, and most obviously, there are 'slippery slope' arguments in moral philosophy where the relation that plays the role of salient similarity is moral irrelevance. Suppose we agree that infanticide is always impermissible, and that taking the morning-after pill (which prevents the implantation of a fertilized egg) is permissible. But the critic constructs a sequence of possible situations linking these two cases where the differences between adjacent situations in the series seem all to be morally insignificant. The structure of the argument is the same as the sorites, but there is no escaping the problem by saying that the predicate 'morally permissible' just does not have application in this kind of case. One can avoid the problem by denying the premise about the end points or by finding moral significance in some salient point in the course of the series, but many will judge that both the pressure to accept an extreme conclusion, and the demand to find a salient sharp cutoff point distort the moral phenomena.

Second, some puzzles about identity over time, or across possible worlds can be posed with a sorites-like argument. When the object in question is a ship, one might think that the puzzle has its source in a kind of vagueness, though it is not easy to identify the source of the vagueness—it is surely not the identity relation that is vague. And when it is *personal* identity that is in question, many have a strong inclination to think that something substantive is at stake.

Third, Timothy Williamson's signature argument for the thesis that there are no mental states that are *luminous* takes a sorites-like form, but he argues that it does not essentially involve vagueness. A luminous state S is one that meets the condition that whenever one is in S, one is in a position to know one is in S. The thesis that if one knows that ϕ, then one is in a position to know that one knows that ϕ implies that knowledge itself is luminous. Williamson's argument is like the sorites in that it begins with a case of someone who in state S at a certain time, but not in state S at a later time. Over a period of time between the two, the subject gradually changes from S to not-S. The sorites-like sequence is a sequence of times close enough together to justify the claim that if the agent *knows* that she is in state S at one time, then she is in state S at the adjacent time. But this analogue of the second premise of the sorites argument (Williamson's margin-of-error principle) is compatible with there being a sharp line between cases in which the subject is in state S, and situations in which she is not.[1]

The puzzle I will look at in a little more detail is a puzzle about knowledge and chance posed in a recent paper by Cian Dorr, Jeremy Goodman and John Hawthorne (DGH).[2] I will describe it in a way that exactly fits the abstract form of the sorites argument about vagueness sketched at the beginning of this paper. Here is the set up: there is a sequence of a thousand fair coins that will be flipped in order until either one of them lands heads (ending the experiment), or all of them have been flipped. Alice knows that this is the set up.

1 I have argued, in defending the principle that knowing that ϕ puts one is a position to know that one knows it, that Williamson's margin of error principles should be rejected. See Stalnaker 2015.

2 Dorr, Goodman and Hawthorne 2014.

The sequence, α_1 to α_{1000}, is the sequence of coins. The analogue of the vague predicate F is the following condition that may hold, or not, for each of the thousand coins: *For all Alice knows (before the experiment begins), coin x will be flipped*. The analogue of the relation R in this version of the argument is the relation that holds between each coin from α_1 to α_{999} and its successor in the sequence.

It is obvious that for early coins in the sequence, for example α_2, the condition applies, and DGH argue that to avoid unacceptable skepticism, we must assume that Alice knows that the last coin will not be flipped, so that the condition fails to apply to α_{1000}. This gives us the analogue of premise (1) of the sorites argument. Furthermore, according to the set up, for any k, if α_k is flipped, then there will be a 50/50 chance that it will land tails, in which case α_{k+1} will be flipped as well, and this seems to be enough to give us the analogue of premise (2): For each k, if for all Alice knows coin α_k will be flipped, then for all Alice knows, coin α_{k+1} will be flipped. The analogue of premise (3) is stipulated in the description of the setup. But as with the original sorites argument, this triad of premises is inconsistent.

In this case (unlike the sorites about vagueness) there is room to deny premise (1) by claiming that while it is extremely improbable (one chance in 2^{999}) that the last coin will be flipped, it is still compatible with Alice's knowledge. But here is an argument against this way out: It is assumed that Alice knows that the set up is as described, but whatever her basis for initially taking herself to know this, if she were to learn that all 1,000 coins were flipped (or even if she were to learn that the first 500 coins were flipped), she would have sufficient reason to doubt her hypothesis about the set up. All empirical hypotheses are subject to defeat, and it would be excessively dogmatic for Alice to continue holding onto this hypothesis in the face of ever growing evidence against it. If this is right, then we can conclude that if Alice learned that all 1,000 coins were flipped, she would *then* not know that the set up was as described. One might be tempted to say that in this scenario, Alice still knew (before any coins were flipped) that all the coins were fair, but later lost that knowledge.[3] This, however, seems implausible, since Alice could consider, at the start, what she *would* conclude if she learned (against all expectation) that all the coins landed heads. She would, if rational, judge that in that case, she was probably (or certainly) wrong about the set up, and this seems enough for her (at the start) to conclude, '*if* it is true that all the coins will be flipped, *then* I don't now know that the set up is as it was described. From this she can infer, by modus tollens, from the hypothesis that she *does* know (at the start) that the set up is as described, that it is false that all the coins will be flipped. So she is in a position to know this.

One can question this argument at various points, but let's consider the alternative response: to deny premise (2), which is to accept the analogue of the sharp boundaries claim. Suppose there is a sharp boundary—a first coin in the series, α_k meeting the condition that it is compatible with Alice's knowledge that α_k be flipped, but incompatible with her knowledge that α_{k+1} be flipped. Suppose Alice considers the hypothesis, for any particular k, that k is the place where the boundary is located. She reasons as follows: if k marks the boundary, then I know that α_{k+1} won't be flipped, but also that it *would* be flipped if α_k were to land tails, so I know that if α_k is flipped, it will land heads. But I don't know that,

3 This is the conclusion Williamson reaches about an example with a similar structure. See Williamson 2000, 205.

so k cannot mark the boundary. If Alice can give this argument for any hypothesis about the value of k, how can there be a k that marks the boundary?[4]

Suppose that, for a certain k such that it is compatible with Alice's knowledge that α_k is flipped, Alice is offered the following conditional bet: the condition is that α_k is flipped (so the bet is called off if the game ends before that coin is reached). If α_k lands tails, Alice gets $100, but if it lands heads, she loses $10. Intuitively, she ought to take the bet, even though if k is large, she will have little chance of either winning or losing (since the bet will very likely be called off). If the bet is *not* called off, it seems that she has a 50% chance of winning $100, and a 50% chance of losing $10, which looks like a good deal. But if k is the cut-off point, then it follows that she knows that α_k will land heads if flipped, so that she knows she will lose the bet unless it is called off. The conditional bet has small but negative expected utility. If you agree that a rational person should accept the bet, you should conclude that k cannot be the number that marks the boundary, and this will be true for any k.

Both of these arguments against the sharp boundaries claim take the form of zeroing in on a particular candidate for the boundary point, and arguing that the point cannot be there. But if our account of knowledge is context and interest dependent, then this kind of argument will be compatible with the possibility that there is a boundary, but one that shifts with shifts in salience, and in interests in exactly the way that Fara's shifting sands strategy for solving the problem about vagueness suggests. When a particular pair of adjacent coins is in focus, we see that if we reach the first, there will be a good chance we will reach the second as well, so either both are epistemically possible, or neither is. The boundary point must be somewhere else—either earlier, so we take ourselves to know that neither coin will be flipped, or later, so that it is open whether either will be.[5]

The second argument, involving a conditional bet, introduces a practical interest to which the location of the boundary point is relevant. It is not just where one's attention is focused that is important, but which points in the potential sequence of possible events are relevant to a decision that has to be made. The story provides a specific model for a way in which practical interests shift a cutoff point away from a point that is relevant to the consequences of a certain decision.

My cautious conclusion is that Fara's shifting sands strategy should be a part of a solution to this puzzle, and that it helps to point the way to a well-motivated account of knowledge that is both context-dependent, and sensitive to practical interests.[6] But much more

---

4 One could object to Alice's reductio by saying that it might be, for some k, that she *does* know that if α_k is flipped, it will land heads—she just doesn't know which k it is for which she has this knowledge. This requires a rejection of the KK thesis, which some would welcome, but even without the KK thesis, it seems implausible to think that Alice could know of a fair coin that might for all she know be flipped that if it is flipped, it will land heads. See Goodman and Salow 2018 for discussion of the KK thesis, in the context of this paradox.

5 The dynamic dimension of the story—the way our knowledge and belief would change as we actually go through the process of flipping coins—is relevant, and it complicates the story, but it is important to keep in mind that the puzzle is a static one about the knowledge and beliefs that one has at the beginning, as one reflects from that perspective about what will and might happen as the sequence of events unfolds.

6 I defend a contextualist account of knowledge in the information-theoretic tradition in Stalnaker (forthcoming), which briefly discusses this paradox, but provides only some sketchy ideas about a solution to it.

needs to be said, about this case and about the other analogues to the sorites arguments, before we have satisfactory solutions to the problems they raise.[7]

REFERENCES

Dorr, Cian, Jeremy Goodman and John Hawthorne. 2014. Knowing against the odds. *Philosophical Studies* 170: 277-287.

Goodman, Jeremy and Bernhard. Salow. 2018. Taking a chance on KK. *Philosophical Studies* 175: 183-196.

Graff-Fara, Delia 2000. Shifting sands: an interest-relative theory of vagueness. *Philosophical Topics* 28: 45-81.

Rayo, Agustín. 2008. Vague representation. *Mind* 117: 329-373.

Stalnaker, Robert. 2015. Luminosity and the KK principle. In S. Goldberg, ed., *Externalism Self-knowledge and Skepticism*, pp. 17-40. Oxford: Oxford University Press.

— (forthcoming) Contextualism and the logic of knowledge. In Robert Stalnaker. *Knowledge and conditionals*. Oxford: Oxford University Press.

Williamson, Timothy 2000. *Knowledge and its limits*. Oxford: Oxford University Press.

Robert Stalnaker is Professor of philosophy emeritus at MIT. He earlier taught at Yale University, the University of Illinois at Urbana-Champaign, and at Cornell University. He is the author of six books and collections of papers, including *Inquiry* (MIT Press, 1984), *Our Knowledge of the Internal World* (OUP, 2008), and *Mere Possibilities* (Princeton U. Press, 2012).

Address: Professor of philosophy, emeritus. MIT, 32D808. 77 Massachusetts Avenue. Cambridge, MA 02139, USA.
Email: stal@mit.edu

[7] Thanks to two anonymous referees for helpful comments.