Intransitivity and Vagueness

Joseph Y. Halpern*
Cornell University
Computer Science Department
Ithaca, NY 14853
halpern@cs.cornell.edu
http://www.cs.cornell.edu/home/halpern

Abstract

There are many examples in the literature that suggest that indistinguishability is intransitive, despite the fact that the indistinguishability relation is typically taken to be an equivalence relation (and thus transitive). It is shown that if the uncertainty perception and the question of when an agent *reports* that two things are indistinguishable are both carefully modeled, the problems disappear, and indistinguishability can indeed be taken to be an equivalence relation. Moreover, this model also suggests a logic of *vagueness* that seems to solve many of the problems related to vagueness discussed in the philosophical literature. In particular, it is shown here how the logic can handle the *Sorites Paradox*.

1 Introduction

While it seems that indistinguishability should be an equivalence relation and thus, in particular, transitive, there are many examples in the literature that suggest otherwise. For example, tasters cannot distinguish a cup of coffee with one grain of sugar from one without sugar, nor, more generally, a cup with $n+1$ grains of sugar from one with n grains of sugar. But they can certainly distinguish a cup with 1,000 grains of sugar from one with no sugar at all.

These intransitivities in indistinguishability lead to intransitivities in preference. For example, consider someone who prefers coffee with a teaspoon of sugar to one with no sugar. Since she cannot distinguish a cup with n grains from a cup with $n+1$ grains, she is clearly indifferent between them. Yet, if a teaspoon of sugar is 1,000 grains, then she clearly prefers a cup with 1,000 grains to a cup with no sugar.

There is a strong intuition that the indistinguishability relation should be transitive, as should the relation of equivalence on preferences. Indeed, transitivity is implicit in our use of the word "equivalence" to describe the relation on preferences. Moreover, it is this intuition that forms the basis of the partitional model for knowledge used in game theory (see, e.g., [Aumann 1976]) and in the distributed systems community [Fagin, Halpern, Moses, and Vardi 1995]. On the other hand, besides the obvious experimental observations, there have been arguments going back to at least Poincaré [1902] that the physical world is not transitive in this sense. In this paper, I try to reconcile our intuitions about indistinguishability with the experimental observations, in a way that seems (at least to me) both intuitively appealing and psychologically plausible. I then go on to apply the ideas developed to the problem of *vagueness*.

To understand the vagueness problem, consider the well-known *Sorites Paradox*: If $n+1$ grains of sand make a heap, then so do n. But 1,000,000 grains of sand are clearly a heap, and 1 grain of sand does not constitute a heap. Let **Heap** to be a predicate such that **Heap**(n) holds if n grains of sand make a heap. What is the extension of **Heap**? That is, for what subset of natural numbers does **Heap** hold? Is this even well defined? Clearly the set of numbers for which **Heap** holds is upward closed: if n grains of sand is a heap, then surely $n+1$ grains of sand is a heap. Similarly, the set of grains of sand which are not a heap is downward closed: if n grains of sand is not a heap, then $n-1$ grains of sand is not a heap. However, there is a fuzzy middle ground, which is in part the reason for the paradox. The relationship of the vagueness of **Heap** to indistinguishability should be clear: n grains of sand are indistinguishable from $n+1$ grains. Indeed, just as

Heap is a vague predicate, so is the predicate **Sweet**, where **Sweet**(n) holds if a cup of coffee with n grains of sugar is sweet. So it is not surprising that an approach to dealing with intransitivity has something to say about vagueness.

The rest of this paper is organized as follows. In Section 2 I discuss my solution to the intransitivity problem. In Section 3, I show how how this solution can be applied to the problem of vagueness. There is a huge literature on the vagueness problem. Perhaps the best-known approach in the AI literature involves fuzzy logic, but fuzzy logic repre-

*Work supported in part by NSF under grant CTC-0208535, by ONR under grants N00014-00-1-03-41 and N00014-01-10-511, by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the ONR under grant N00014-01-1-0795, and by AFOSR under grant F49620-02-1-0101.

sents only a small part of the picture; the number of recent book-length treatments, including [Keefe 2000; Keefe and Smith 1996; Sorenson 2001; Williamson 1994], give a sense of the activity in the area. I formalize the intuitions discussed in Section 2 using a logic for reasoning about vague propositions, provide a sound a complete axiomatization for the logic, and show how it can deal with problems like the Sorites Paradox. I compare my approach to vagueness to some of the leading alternatives in Section 4. Finally, I conclude with some discussion in Section 5.

2 Intransitivity

Clearly part of the explanation for the apparent intransitivity in the sugar example involves differences that are too small to be detected. But this can't be the whole story. To understand the issues, imagine a robot with a simple sensor for sweetness. The robot "drinks" a cup of coffee and measures how sweet it is. Further imagine that the robot's sensor is sensitive only at the 10-grain level. Formally, this means that a cup with 0–9 grains results in a sensor reading of 0, 10–19 grains results in a sensor reading of 1, and so on. If the situation were indeed that simple, then indistinguishability would in fact be an equivalence relation. All cups of coffee with 0–9 grains of sugar would be indistinguishable, as would cups of coffee with 10–19 grains, and so on. However, in this simple setting, a cup of coffee with 9 grains of sugar would be distinguishable from cups with 10 grains.

To recover intransitivity requires two more steps. The first involves dropping the assumption that the number of grains of sugar uniquely determines the reading of the sensor. There are many reasons to drop this assumption. For one thing, the robot's sensor may not be completely reliable; for example, 12 grains of sugar may occasionally lead to a reading of 0; 8 grains may lead to a reading of 1. A second reason is that the reading may depend in part on the robot's state. After drinking three cups of sweet coffee, the robot's perception of sweetness may be dulled somewhat, and a cup with 112 grains of sugar may result in a reading of 10. A third reason may be due to problems in the robot's vision system, so that the robot may "read" 1 when the sensor actually says 2. It is easy to imagine other reasons; the details do not matter here. All that matters is what is done about this indeterminacy. This leads to the second step of my "solution".

To simplify the rest of the discussion, assume that the "indeterminacy" is less than 4 grains of sugar, so that if there are actually n grains of sugar, the sensor reading is between $\lfloor (n-4)/10 \rfloor$ and $\lfloor (n+4)/10 \rfloor$.[1] It follows that two cups of coffee with the same number of grains may result in readings that are not the same, but they will be at most one apart. Moreover, two cups of coffee which differ by one grain of sugar will also result in readings that differ by at most one.

The robot is asked to compare the sweetness of cups, not sensor readings. Thus, we must ask when the robot *reports* two cups of coffee as being of equivalent sweetness. Given the indeterminacy of the reading, it seems reasonable that two cups of sugar that result in a sensor reading that differ by no more than one are reported as indistinguishable, since they could have come from cups of coffee with the same number of grains of sugar. It is immediate that reports of indistinguishability will be intransitive, even if the sweetness readings themselves clearly determine an equivalence relation. Indeed, if the number of grains in two cups of coffee differs by one, then the two cups will be reported as equivalent. But if the number of grains differs by at least eighteen, then they will be reported as inequivalent.

To sum up, reports of relative sweetness (and, more generally, reports about perceptions) exhibit intransitivity; it may well be that there are three cups of sugar such that a and b are reported as being equivalent in sweetness, as are b and c, but c is reported as being sweeter than a. Nevertheless, the underlying "perceived sweetness" relation can be taken to be transitive. However, "perceived sweetness" must then be taken to be a relation on the taste of a cup of coffee tried at a particular time, not on the number of grains of sugar in a cup. That is, rather than considering a **Sweeter-Than** relation where **Sweeter-Than**(n, n') holds if a cup of coffee with n grains is sweeter than one with n' grains of sugar, we should consider a **Sweeter-Than**$'$ relation, where **Sweeter-Than**$'((c, s), (c', s'))$ holds if cup of coffee c tried by the agent in (subjective) state s (where the state includes the time, and other features of the agent's state, such as how many cups of coffee she has had recently) is perceived as sweeter than cup of coffee c' tried by the agent in state s'. The former relation may not be transitive; the latter is. But note that the latter relation does not completely determine when the agent *reports* c as being sweeter than c'. Intransitivity in reports of perceptions does not necessarily imply intransitivity in actual perceptions.

3 Vagueness

The term "vagueness" has been used somewhat vaguely in the literature. Roughly speaking, a term is said to be vague if its use varies both between and within speakers. (According to Williamson [1994], this interpretation of vagueness goes back at least to Peirce [1956], and was also used by Black [1937] and Hempel [1939].) In the language of the previous section, P is vague if, for some a, some agents may report $P(a)$ while others may report $\neg P(a)$ and, indeed, the same agent may sometimes report $P(a)$ and sometimes $\neg P(a)$.

Vagueness has been applied to what seem to me to be two distinct, but related, phenomena. For one thing, it has been applied to predicates like **Red**, where the different reports may be attributed in part to there not being an objective notion of what counts as red. That is, two agents looking at

[1] $\lfloor x \rfloor$, the floor of x, is the largest integer less than or equal to x. Thus, for example, $\lfloor 3.2 \rfloor = 3$.

the same object (under the same lighting conditions) may disagree as to whether an object is red, although typically they will agree. Vagueness is also applied to situations with epistemic uncertainty, as in the case of a predicate **Crowd**, where **Crowd**(n) holds if there are at least n people in a stadium at a particular time.[2] Here there may be different responses because agents have trouble estimating the size of a crowd. I present a model that distinguishes these two sources of vagueness. Because vagueness is rather slippery, I also present a formal logic of vagueness.

3.1 A Modal Logic of Vagueness: Syntax and Semantics

To reason about vagueness, I consider a modal logic $\mathcal{L}_n^{DR}$ with two families of modal operators: $R_1, \ldots, R_n$, where $R_i\varphi$ is interpreted as "agent i reports φ", and $D_1, \ldots, D_n$, where $D_i\varphi$ is interpreted as "according to agent i, φ is definitely the case". For simplicity, I consider only a propositional logic; there are no difficulties extending to the first-order case. As the notation makes clear, I allow multiple agents, since some issues regarding vagueness (in particular, the fact that different agents may interpret a vague predicate differently) are best considered in a multi-agent setting.

Start with a (possibly infinite) set of primitive propositions. More complicated formulas are formed by closing off under conjunction, negation, and the modal operators $R_1, \ldots, R_n$ and $D_1, \ldots, D_n$.

A *vagueness structure* M has the form $(W, P_1, \ldots, P_n, \pi_1, \ldots, \pi_n)$, where P_i is a nonempty subset of W for $i = 1, \ldots, n$, and π_i is an interpretation, which associates with each primitive proposition a subset of W. Intuitively, P_i consists of the worlds that agent i initially considers plausible. For those used to thinking probabilistically, the worlds in P_i can be thought of as those that have prior probability greater than ϵ according to agent i, for some fixed $\epsilon \geq 0$.[3] A simple class of models is obtained by taking $P_i = W$ for $i = 1, \ldots, n$; however, as we shall see, in the case of multiple agents, there are advantages to allowing $P_i \neq W$. Turning to the truth assignments π_i, note that it is somewhat nonstandard in modal logic to have a different truth assignment for each agent; this different truth assignment is intended to capture the intuition that the truth of formulas like **Sweet** is, to some extent, dependent on the agent, and not just on objective features of the world.

I assume that $W \subseteq O \times S_1 \times \ldots S_n$, where O is a set of objective states, and S_i is a set of subjective states for agent i. Thus, worlds have the form $(o, s_1, \ldots, s_n)$. Agent i's subjective state s_i represents i's perception of the world and everything else about the agent's makeup that determines the agent's report. For example, in the case of the robot with a sensor, o could be the actual number of grains of sugar in a cup of coffee and s_i could be the reading on the robot's sensor. Similarly, if the formula in question was **Thin(TW)** ("Tim Williamson is thin", a formula often considered in [Williamson 1994]), then o could represent the actual dimensions of TW, and s_i could represent the agent's perceptions. Note that s_i could also include information about other features of the situation, such as the relevant reference group. (Notions of thinness are clearly somewhat culture dependent and change over time; what counts as thin might be very different if TW is a sumo wrestler.) In addition, s_i could include the agent's cutoff points for deciding what counts as thin, or what counts as red. In the case of the robot discussed in Section 2, the subjective state could include its rule for deciding when to report something as sweet.

If p is a primitive proposition then, intuitively, $(o, s_1, \ldots, s_n) \in \pi_i(p)$ if i would consider p true if i knew exactly what the objective situation was (i.e., if i knew o), given i's possibly subjective judgment of what counts as "p-ness". Given this intuition, it should be clear that all that should matter in this evaluation is the objective part of the world, o, and (possibly) agent i's subjective state, s_i. In the case of the robot, whether $(o, s_1, \ldots, s_n) \in \pi_i(\textbf{Sweet})$ clearly depends on how many grains of sugar are in the cup of coffee, and may also depend on the robot's perception of sweetness and its cutoff points for sweetness, but does not depend on other robots' perceptions of sweetness. Note that the robot may give different answers in two different subjective states, even if the objective state is the same and the robot knows the objective state, since both its perceptions of sweetness and its cutoff point for sweetness may be different in the two subjective states.

I write $w \sim_i w'$ if w and w' agree on agent i's subjective state, and I write $w \sim_o w'$ if w and w' agree on the objective part of the state. Intuitively, the $\sim_i$ relation can be viewed as describing the worlds that agent i considers possible. Put another way, if $w \sim_i w'$, then i cannot distinguish w from w', given his current information. Note that the indistinguishability relation is transitive (indeed, it is an equivalence relation), in keeping with the discussion in Section 2. I assume that π_i depends only on the objective part of the state and i's subjective state, so that if $w \in \pi_i(p)$ for a primitive propo-

[2] Of course, there may still be objective uncertainty as to how to do the count. For example, does a pregnant woman count as one or two? If the answer is "one", then if she goes into labor, at what point does the answer become "two"? The point is that even if we assume that all these details have been worked out, so that there would be be complete agreement among all agents as to how many people are in the stadium if they had all the relevant information, there will still in general be uncertainty as to how many people are in the stadium. This uncertainty leads to vagueness.

[3] In general, the worlds that an agent considers plausible depends on the agent's subjective state. That is why I have been careful here to say that P_i consists of the worlds that agent i *initially* considers plausible. It should shortly become clear how the model takes into account the fact that the agent's set of plausible worlds changes according to the agent's subjective state.

sition p, and $w \sim_i w'$ and $w \sim_o w'$, then $w' \in \pi_i(p)$. Note that j's state (for $j \neq i$) has no effect on i's determination of the truth of p. There may be some primitive propositions whose truth depends only on the objective part of the state (for example, **Crowd**(n) is such a proposition). If p is such an objective proposition, then $\pi_i(p) = \pi_j(p)$ for all agents i and j, and, if $w \sim_o w'$, then $w \in \pi_i(p)$ iff $w' \in \pi_i(p)$.

I next define what it means for a formula to be true. The truth of formulas is relative to both the agent and the world. I write $(M, w, i) \models \varphi$ if φ is true according to agent i in world w. In the case of a primitive proposition p,

$$(M, w, i) \models p \quad \text{iff} \quad w \in \pi_i(p).$$

I define $\models$ for other formulas by induction. For conjunction and negation, the definitions are standard:

$$(M, w, i) \models \neg\varphi \quad \text{iff} \quad (M, w, i) \not\models \varphi;$$
$$(M, w, i) \models \varphi \wedge \psi \quad \text{iff} \quad (M, w, i) \models \varphi \text{ and } (M, w, i) \models \psi.$$

In the semantics for negation, I have implicitly assumed that, given the objective situation and agent i's subjective state, agent i is prepared to say, for every primitive proposition p, whether or not p holds. Thus, if $w \notin \pi_i(p)$, so that agent i would not consider p true given i's subjective state in w if i knew the objective situation at w, then I am assuming that i would consider $\neg p$ true in this world. This assumption is being made mainly for ease of exposition. It would be easy to modify the approach to allow agent i to say (given the objective state and i's subjective state), either "p holds", "p does not hold", or "I am not prepared to say whether p holds or p does not hold".[4] However, what I am explicitly avoiding here is taking a fuzzy-logic like approach of saying something like "p is true to degree .3". While the notion of degree of truth is certainly intuitively appealing, it has other problems. The most obvious in this context is where the .3 is coming from. Even if p is vague, the notion "p is true to degree .3" is precise. It is not clear that introducing a continuum of precise propositions to replace the vague proposition p really solves the problem of vagueness. Having said that, there is a natural connection between the approach I am about to present and fuzzy logic; see Section 4.1.

Next, I consider the semantics for the modal operators R_j, $j = 1, \ldots, n$. Recall that $R_j\varphi$ is interpreted as "agent j reports φ. Formally, I take $R_j\varphi$ to be true if φ is true at all plausible states j considers possible. Thus,

$$(M, w, i) \models R_j\varphi \quad \text{iff}$$
$$\quad (M, w', j) \models \varphi \text{ for all } w' \text{ such that } w \sim_j w' \text{ and } w' \in P_j.$$

[4]The resulting logic would still be two-valued; the primitive proposition p would be replaced by a family of three primitive propositions, p_y, p_n, and $p_?$, corresponding to "p holds", "p does not hold", and "I am not prepared to say whether p holds or does not hold", with a semantic requirement (which becomes an axiom in the complete axiomatization) stipulating that exactly one proposition in each such family holds at each world.

Of course, for a particular formula φ, an agent may neither report φ nor $\neg\varphi$. An agent may not be willing to say either that TW is thin or that TW is not thin. Note that, effectively, the set of plausible states according to agent j given the agent's subjective state in world w can be viewed as the worlds in in P_j that are indistinguishable to agent j from w. Essentially, the agent j is updating the worlds that she initially considers plausible by intersecting them with the worlds she considers possible, given her subjective state at world w. If $P_j = W$ for all agents $j = 1, \ldots, n$, then it is impossible for agents to give conflicting reports; that is, the formula $R_i\varphi \wedge \neg R_j\varphi$ would be inconsistent. By considering only the plausible worlds when giving the semantics for R_j, it is consistent to have conflicting reports.

Finally, φ is definitely true at state w if the truth of φ is determined by the objective state at w:

$$(M, w, i) \models D_j\varphi \quad \text{iff}$$
$$\quad (M, w', j) \models \varphi \text{ for all } w' \text{ such that } w \sim_o w'.$$

A formula is said to be *agent-independent* if its truth is independent of the agent. That is, φ is agent-independent if, for all worlds w,

$$(M, w, i) \models \varphi \quad \text{iff} \quad (M, w, j) \models \varphi.$$

As we observed earlier, objective primitive propositions (whose truth depends only on the objective part of a world) are agent-independent; it is easy to see that formulas of the form $D_j\varphi$ and $R_j\varphi$ are as well. If φ is agent-independent, then I often write $(M, w) \models \varphi$ rather than $(M, w, i) \models \varphi$.

3.2 A Modal Logic of Vagueness: Axiomatization and Complexity

It is easy to see that R_j satisfies the axioms and rules of the modal logic KD45.[5] It is also easy to see that D_j satisfies the axioms of KD45. It would seem that, in fact, D_j should satisfy the axioms of S5, since its semantics is determined by $\sim_j$, which is an equivalence relation. This is not quite true. The problem is with the so-called *truth axiom* of S5, which, in this context, would say that anything that is definitely true according to agent j is true. This would be true if there were only one agent, but is not true with many agents, because of the different π_i operators.

To see the problem, suppose that p is a primitive proposition. It is easy to see that $(M, w, i) \models D_i p \Rightarrow p$ for all worlds w. However, it is not necessarily the case that $(M, w, i) \models D_j p \Rightarrow p$ if $i \neq j$. Just because, according to agent i, p is definitely true according to agent j, it does not follow that p is true *according to agent* i. What

[5]For modal logicians, perhaps the easiest way to see this is to observe that we can define a relation $\mathcal{R}_j$ on worlds consisting of all pairs (w, w') such that $w \sim_j w'$ and $w' \in P_j$. This relation, which characterizes the modal operator R_j, is easily seen to be Euclidean and transitive, and thus determines a modal operator satisfying the axioms of KD45.

is true in general is that $D_j\varphi \Rightarrow \varphi$ is valid for *agent-independent* formulas. Unfortunately, agent independence is a semantic property. To capture this observation as an axiom, we need a syntactic condition sufficient to ensure that a formula is necessarily agent independent. I observed earlier that formulas of the form $R_j\varphi$ and $D_j\varphi$ are agent-independent. It is immediate that Boolean combination of such formulas are also agent-independent. Say that a formula is *necessarily agent-independent* if it is a Boolean combination of formulas of the form $R_j\varphi$ and $D_{j'}\varphi'$ (where the agents in the subscripts may be the same or different). Thus, for example, $(\neg R_1 D_2 p \wedge D_1 p) \vee R_2 p$ is necessarily agent-independent. Clearly, whether a formula is necessarily agent-independent depends only on the syntactic form of the formula. Moreover, $D_j\varphi \Rightarrow \varphi$ is valid for formulas that are necessarily agent-independent. However, this axiom does not capture the fact that $(M, w, i) \models D_i\varphi \Rightarrow \varphi$ for all worlds w. Indeed, this fact is not directly expressible in the logic, but something somewhat similar is. For arbitrary formulas $\varphi_1, \ldots, \varphi_n$, note that at least one of $D_i\varphi_1 \Rightarrow \varphi_1, \ldots, D_n\varphi_n \Rightarrow \varphi_n$ must be true respect to each triple (M, w, i), $i = 1, \ldots, n$. Thus, the formula $(D_1\varphi_1 \Rightarrow \varphi_1) \vee \ldots \vee (D_n\varphi_n \Rightarrow \varphi_n)$ is valid. This additional property turns out to be exactly what is needed to provide a complete axiomatization.

Let AX be the axiom system that consists of the following axioms Taut, R1–R4, and D1–D6, and rules of inference $\mathrm{Nec_R}$, $\mathrm{Nec_D}$, and MP:

Taut. All instances of propositional tautologies.

R1. $R_j(\varphi \Rightarrow \psi) \Rightarrow (R_j\varphi \Rightarrow R_j\psi)$.

R2. $R_j\varphi \Rightarrow R_jR_j\varphi$.

R3. $\neg R_j\varphi \Rightarrow R_j\neg R_j\varphi$.

R4. $\neg R_j(\mathit{false})$.

D1. $D_j(\varphi \Rightarrow \psi) \Rightarrow (D_j\varphi \Rightarrow D_j\psi)$.

D2. $D_j\varphi \Rightarrow D_jD_j\varphi$.

D3. $\neg D_j\varphi \Rightarrow D_j\neg D_j\varphi$.

D4. $\neg D_j(\mathit{false})$.

D5. $D_j\varphi \Rightarrow \varphi$ if φ is necessarily agent-independent.

D6. $(D_1\varphi_1 \Rightarrow \varphi_1) \vee \ldots \vee (D_n\varphi_n \Rightarrow \varphi_n)$.

Nec$_\mathrm{R}$. From φ infer $R_j\varphi$.

Nec$_\mathrm{D}$. From φ infer $D_j\varphi$.

MP. From φ and $\varphi \Rightarrow \psi$ infer ψ.

Using standard techniques of modal logic, it is can be shown that AX characterizes $\mathcal{L}_n^{DR}$.

Theorem 3.1: *AX is a sound and complete axiomatization with respect to vagueness structures for the language $\mathcal{L}_n^{DR}$.*

This shows that the semantics that I have given implicitly assumes that agents have perfect introspection and are logically omniscient. Introspection and logical omniscience are both strong requirements. There are standard techniques in modal logic that make it possible to give semantics to R_j that is appropriate for non-introspective agents. With more effort, it is also possible to avoid logical omniscience. (See, for example, the discussion of logical omniscience in [Fagin, Halpern, Moses, and Vardi 1995].) In any case, very little of my treatment of vagueness depends on these properties of R_j.

The complexity of the validity and satisfiability problem for the $\mathcal{L}_n^{DR}$ can also be determined using standard techniques.

Theorem 3.2: *For all $n \geq 1$, determining the problem of determining the validity (or satisfiability) of formulas in $\mathcal{L}_n^{DR}$ is PSPACE-complete.*

Proof: The validity and satisfiability problems for KD45 and S5 in the case of two or more agents is known to be PSPACE-complete [Halpern and Moses 1992]. The modal operators R_j and D_j act essentially like KD45 and S5 operators, respectively. Thus, even if there is only one agent, there are two modal operators, and a straightforward modification of the lower bound argument in [Halpern and Moses 1992] gives the PSPACE lower bound. The techniques of [Halpern and Moses 1992] also give the upper bound, for any number of agents. ∎

3.3 Capturing Vagueness and the Sorites Paradox

Although I have described this logic as one for capturing features of vagueness, the question still remains as to what it means to say that a proposition φ is vague. I suggested earlier that a standard view has been to take φ to be vague if, in some situations, some agents report φ while others report $\neg\varphi$, or if the same agent may sometimes report φ and sometimes report $\neg\varphi$ in the same situation. Both intuitions can be captured in the logic. It is perfectly consistent that $(M, w) \models R_i\varphi \wedge \neg R_j\varphi$ if $i \neq j$; that is, the logic makes it easy to express that two agents may report different things regarding φ. Expressing the second intuition requires a little more care; it is certainly not consistent to have $(M, w) \models R_j\varphi \wedge \neg R_j\varphi$. However, a more reasonable interpretation of the second intuition is to say that in the same *objective* situation, an agent i may both report φ and $\neg\varphi$. It is certainly consistent that there are two worlds w and w' such that $w \sim_o w'$, $(M, w) \models R_j\varphi$, and $(M, w') \models \neg R_j\varphi$. Note that this is true iff $(M, w) \models \neg D_jR_j\varphi$. Thus, in the case of one agent, under this interpretation, φ is taken to be vague if $\neg DR\varphi$ holds at some world. I return to this point in Section 4.4.

Although, by design, the logic and the associated semantics can capture features of vagueness, the question still remains as to whether it gives any insight into the problems associated with vagueness. I defer the discussion of some of the problems (e.g., higher-order vagueness) to Section 4.

Here I show how it can deal with the Sorites Paradox. Before going into details, it seems to me that there should be two components to a solution to the Sorites Paradox. The first is to show where the reasoning that leads to the paradox goes wrong in whatever formalism is being used. The second is to explain why, nevertheless, the argument seems so reasonable and natural to most people.

The Sorites Paradox is typically formalized as follows:

1. **Heap**(1,000,000).
2. $\forall n > 1($**Heap**$(n) \Rightarrow$ **Heap**$(n-1))$.
3. $\neg$**Heap**(1).

It is hard to argue with statements 1 and 3, so the obvious place to look for a problem is in statement 2, the inductive step. And, indeed, most authors have, for various reasons, rejected this step (see, for example, [Dummett 1975; Sorenson 2001; Williamson 1994] for typical discussions). As I suggested in the introduction, it appears that rejecting the inductive step requires committing to the existence of an n such that n grains of sand is a heap and $n-1$ is not. While I too reject the inductive step, it does *not* follow that there is such an n in the framework I have introduced here, because I do not assume an objective notion of heap (whose extension is the set of natural numbers n such that n grains of sands form a heap). What constitutes a heap in my framework depends not only on the objective aspects of the world (i.e., the number of grains of sand), but also on the agent and her subjective state.

To be somewhat more formal, assume for simplicity that there is only one agent. Consider models where the objective part of the world includes the number of grains of sand in a particular pile of sand being observed by the agent, and the agent's subjective state includes how many times the agent has been asked whether a particular pile of sand constitutes a heap. What I have in mind here is that the sand is repeatedly added to or removed from the pile, and each time this is done, the agent is asked "Is this a heap?". Of course, the objective part of the world may also include the shape of the pile and the lighting conditions, while the agent's subjective state may include things like the agent's sense perception of the pile under some suitable representation. Exactly what is included in the objective and subjective parts of the world do not matter for this analysis.

In this setup, rather than being interested in whether a pile of n grains of sand constitutes a heap, we are interested in the question of whether, when viewing a pile of n grains of sand, the agent would report that it is a heap. That is, we are interested in the formula $S(n)$, which I take to be an abbreviation of **Pile**$(n) \Rightarrow R($**Heap**$)$. The formula **Pile**(n) is true at a world w if, according to the objective component of w, there are in fact n grains of sand in the pile. Note that **Pile** is not a vague predicate at all, but an objective statement about the number of grains of sand present.[6] By way of contrast, **Heap** is vague; its truth depends on both the objective situation (how many grains of sand there actually are) and the agent's subjective state.

There is no harm in restricting to models where $S(1,000,000)$ holds in all worlds and $S(1)$ is false in all worlds where the pile actually does consist of one grain of sand. If there are actually 1,000,000 grains of sand in the pile, then the agent's subjective state is surely such that she would report that there is a heap; and if there is actually only one grain of sand, then the agent would surely report that there is not a heap. We would get the paradox if the inductive step, $\forall n > 1(S(n) \Rightarrow S(n-1))$ holds in all worlds. However, it does not, for reasons that have nothing to do with vagueness. Note that in each world, $\mathbf{Pile}(n)$ holds for exactly one value of n. Consider a world w where there is 1 grain of sand in the pile and take $n = 2$. Then $S(2)$ holds vacuously (because its antecedent **Pile**(2) is false), while $S(1)$ is false, since in a world with 1 grain of sand, by assumption, the agent reports that there is not a heap.

The problem here is that the inductive statement $\forall n(n > 1(S(n) \Rightarrow S(n-1))$ does not correctly capture the intended inductive argument. Really what we mean is more like "if there are n grains of sand and the agent reports a heap, then when one grain of sand is removed, the agent will still report a heap".

Note that removing a grain of sand changes both the objective and subjective components of the world. It changes the objective component because there is one less grain of sand; it changes the subjective component even if the agent's sense impression of the pile remains the same, because the agent has been asked one more question regarding piles of sand. The change in the agent's subjective state may not be uniquely determined, since the agent's perception of a pile of $n-1$ grains of sand is not necessarily always the same. But even if it is uniquely determined, the rest of my analysis holds. In any case, given that the world changes, a reasonable reinterpretation of the inductive statement might be "For all worlds w, if there are n grains of sand in the pile in w, and the agent reports that there is a heap in w, then the agent would report that there is a heap in all the worlds that may result after removing one grain of sand." This reinterpretation of the inductive hypothesis cannot be expressed in the logic, but the logic could easily be extended with dynamic-logic like operators so as to be able to express it, using a formula such as

$$\textbf{Pile}(n) \land R(\textbf{Heap}) \Rightarrow [\text{remove 1 grain}](\textbf{Pile}(n-1) \land R(\textbf{Heap}).$$

Indeed, with this way of expressing the inductive step, there is no need to include **Pile**(n) or **Pile**$(n-1)$ in the formula; it suffices to write $R($**Heap**$) \Rightarrow [\text{remove 1 grain}]R($**Heap**$)$.

Is this revised inductive step valid? Again, it is not hard to see it is not. Consider a world where there is a pile of

[6] While I am not assuming that the agent knows the number of grains of sand present, it would actually not affect my analysis at all if the agent was told the exact number.

1,000,000 grains of sand, and the agent is asked for the first time whether this is a heap. By assumption, the agent reports that it is. As more and more grains of sand are removed, at some point the agent (assuming that she has the patience to stick around for all the questions) is bound to say that it is no longer a heap.[7]

Although the framework makes it clear that the induction fails (as it does in other approaches to formulating the problem), the question still remains as to why people accept the inductive argument so quickly. One possible answer may be that it is "almost always" true, although making this precise would require having a probability measure or some other measure of uncertainty on possible worlds. While this may be part of the answer, I think there is another, more natural explanation. When people are asked the question, they do not consider worlds where they have been asked the question many times before. They are more likely to interpret it as "If, in a world where I have before never been asked whether there is a heap, I report that there is a heap, then I will still report that there is a heap after one grain of sand is removed." Observe that this interpretation of the induction hypothesis is consistent with the agent always reporting that a pile of 1,000,000 grains of sand is a heap and always reporting that a pile of 1 grain is not a heap. More importantly, I suspect that the inductive hypothesis is in fact empirically true. After an agent has invested the "psychic energy" to determine whether there is a heap for the first time, it seems to me quite likely that she will not change her mind after one grain of sand is removed. While this will not continue to be true as more and more grains of sand are removed, it does not seem to me that this is not what people think of when they answer the question.

Bottom line: not only does the framework introduced here make it clear why the inductive step, interpreted in the naive way, fails, it also gives a psychologically plausible interpretation that is consistent. Of course, a similar analysis applies to all the other sorites-like paradoxes in the literature.

4 Relations to Other Approaches

In this section I consider how the approach to vagueness sketched in the previous section is related to other approaches to vagueness that have been discussed in the literature.

4.1 Fuzzy Logic

Fuzzy logic [Zadeh 1975] seems like a natural approach to dealing with vagueness, since it does not require a predicate be necessarily true or false; rather, it can be true to a certain degree. As I suggested earlier, this does not immediately resolve the problem of vagueness, since a statement like "this cup of coffee is sweet to degree .8" is itself a crisp statement, when the intuition suggests it should also be vague.

Although I have based my approach on a two-valued logic, there is a rather natural connection between my approach and fuzzy logic. We can take the degree of truth of a formula φ in world w to be the fraction of agents i such that $(M, w, i) \models \varphi$. We expect that, in most worlds, the degree of truth of a formula will be close to either 0 or 1. We can have meaningful communication precisely because there is a large degree of agreement in how agents interpret subjective notions thinness, tallness, sweetness.

Note that the degree of truth of φ in $(o, s_1, \ldots, s_n)$ does not depend just on o, since $s_1, \ldots, s_n$ are not deterministic functions of o. But if we assume that each objective situation o determines a probability distribution on tuples $(s_1, \ldots, s_n)$, then if n is large, for many predicates of interest (e.g., **Thin**, **Sweet**, **Tall**), I expect that, as an empirical matter, the distribution will be normally distributed with a very small variance. In this case, the degree of truth of such a predicate in an objective situation o can be taken to be the expected degree of truth of P, taken over all worlds $(o, s_1, \ldots, s_n)$ whose first component is o.

This discussion shows that my approach to vagueness is compatible with assigning a degree of truth in the interval $[0, 1]$ to vague propositions, as is done in fuzzy logic. Moreover non-vague propositions (called *crisp* in the fuzzy logic literature) get degree of truth either 0 or 1. However, while this is a way of giving a natural interpretation to degrees of truth, and it supports the degree of truth of $\neg\varphi$ being 1 minus the degree of truth of φ, as is done in fuzzy logic, it does not support the semantics for $\wedge$ typically taken in fuzzy logic, where the degree of truth of $\varphi \wedge \psi$ is taken to be the minimum of the degree of truth of φ and the degree of truth of ψ. Indeed, under my interpretation of degree of truth, there is no functional connection between the degree of truth of φ, ψ, and $\varphi \wedge \psi$

4.2 Supervaluations

The D operator also has close relations to the notion of *supervaluations* [Fine 1975; van Fraassen 1968]. Roughly speaking, the intuition behind supervaluations is that language is not completely precise. There are various ways of "extending" a world to make it precise. A formula is then taken to be true at a world w under this approach if it is true under all ways of extending the world. Both the R_j and D_i operators have some of the flavor of supervaluations. If we consider just the objective component of a world o, there are various ways of extending it with subjective components $(s_1, \ldots, s_n)$. $D_i\varphi$ is true at an objective world o if $(M, w, i) \models \varphi$ for all worlds w that extend o. (Note that the truth of $D_j\varphi$ depends only on the objective component

[7] There may well be an in-between period where the agent is uncomfortable about having to decide whether the pile is a heap. As I observed earlier, the semantics implicitly assumes that the agent is willing to answer all questions with a "Yes" or "No", but it is easy to modify things so as to allow "I'm not prepared to say". The problem of vagueness still remains: At what point does the agent first start to say "I'm not prepared to say"?

of a world.) Similarly, given just a subjective component s_j of a world, $R_j\varphi$ is true of s_j if $(M, w, i) \models \varphi$ for all worlds that extend s_i. Not surprisingly, properties of supervaluations can be expressed using R_j or D_j. Bennett [1998] has defined a modal logic that formalizes the supervaluation approach.

4.3 Higher-Order Vagueness

In many approaches towards vagueness, there has been discussion of *higher-order vagueness* (see, for example, [Fine 1975; Williamson 1994]). In the context of the supervaluation approach, we can say that $D\varphi$ ("definitely φ") holds at a world w if φ is true in all extensions of w. Then $D\varphi$ is not vague; at each world, either $D\varphi$ or $\neg D\varphi$ (and $D\neg D\varphi$) is true (in the supervaluation sense). But using this semantics for definitely, it seems that there is a problem. For under this semantics, "definitely φ" implies "definitely definitely φ" (for essentially the same reasons that $D_i\varphi \Rightarrow D_iD_i\varphi$ in the semantics that I have given). But, goes the argument, this does not allow the statement "This is definitely red" to be vague. A rather awkward approach is taken to dealing with this by Fine [1975] (see also [Williamson 1994]), which allows different levels of interpretation. Here the situation is much simpler. Higher-order vagueness is represented by considering combinations of R_i and D_i. We can consider when agent i would report that something is red (R_i**Red**), when he definitely would report it (D_iR_i**Red**), when we would report that he would definitely report it ($R_iD_iR_i$**Red**), and so on. It is easy to see that D_iR_ip does not imply $D_iR_iD_iR_ip$; lower-order vagueness does not imply higher-oder vagueness. However, $D_iR_iD_iR_i\varphi$ does imply $D_iR_i\varphi$ (this follows using the fact that $D_i\varphi \Rightarrow \varphi$ and $R_iR_i\varphi \Rightarrow R_i\varphi$ are both valid). That is, higher-order vagueness does imply lower-order vagueness. However, this depends on the assumption that agents are introspective.

4.4 Williamson's Approach

One of the leading approaches to vagueness in the recent literature is that of Williamson; see [Williamson 1994, Chapters 7 and 8] for an introduction. Williamson considers an epistemic approach, viewing vagueness as ignorance. Very roughly speaking, he uses "know" where I use "report". However, he insists that it cannot be the case that if you know something, then you know you know it, whereas my notion of reporting has the property that R_i implies R_iR_i. It is instructive to examine the example that Williamson uses to argue that you cannot know what you know, to see where his argument breaks down in the framework I have presented.

Williamson considers a situation where you look at a crowd and do not know the number of people in it. He makes what seem to be a number of reasonable assumptions. Among them is the following:

> I know that if there are exactly n people, then I do not know that there are not exactly $n-1$ people.

This may not hold in my framework. This is perhaps easier to see if we think of a robot with sensors. If there are n grains of sugar in the cup, it is possible that a sensor reading compatible with n grains will preclude there being $n-1$ grains. For example, suppose that, as in Section 2, if there are n grains of sugar, and the robot's sensor reading is between $\lfloor(n-4)/10\rfloor$ and $\lfloor(n+4)/10\rfloor$. If there are in fact 16 grains of sugar, then the sensor reading could be 2 ($= \lfloor(16+4)/10\rfloor$). But if the robot knows how its sensor works, then if its sensor reading is 2, then it knows that if there are exactly 16 grains of sand, then (it knows that) there are not exactly 15 grains of sugar. Of course, it is possible to change the semantics of R_i so as to validate Williamson's assumptions. But this point seems to be orthogonal to dealing with vagueness.

Quite apart from his treatment of epistemic matters, Williamson seems to implicitly assume that there is an objective notion of what I have been calling subjectively vague notions, such as red, sweet, and thin. This is captured by what he calls the *supervenience thesis*, which roughly says that if two worlds agree on their objective part, then they must agree on how they interpret what I have called subjective propositions. Williamson focuses on the example of thinness, in which case his notion of supervenience implies that "If x has exactly the same physical measurements in a possible situation s and y has in a possible situation t, then x is thin in s if and only if y is thin in t" [Williamson 1994, p. 203]. I have rejected this viewpoint here, since, for me, whether x is this depends also on the agent's subjective state. Indeed, rejecting this viewpoint is a central component of my approach to intransitivity and vagueness.

Despite these differences, there is one significant point of contact between Williamson's approach and that presented here. Williamson suggests modeling vagueness using a modal operator C for *clarity*. Formally, he takes a model M to be a quadruple (W, d, α, π), where W is a set of worlds and π is an interpretation as above (Williamson seems to implicitly assume that there is a single agent), where d is a metric on W (so that d is a symmetric function mapping $W \times W$ to $[0, \infty)$ such that $d(w, w') = 0$ iff $w = w'$ and $d(w_1, w_2) + d(w_2, w_3) \leq d(w_1, w_3)$), and α is a nonnegative real number. The semantics of formulas is defined in the usual way; the one interesting clause is that for C:

$(M, w) \models C\varphi$ iff
  $(M, w') \models \varphi$ for all w' such that $d(w, w') \leq \alpha$.

Thus, $C\varphi$ is true at a world w if φ is true at all worlds within α of w.

The intuition for this model is perhaps best illustrated by considering it in the framework discussed in the previous section, assuming that there is only one proposition, say **Tall(TW)**, and one agent. Suppose that **Tall(TW)** is taken to hold if TW is above some threshold height t^*. Since **Tall(TW)** is the only primitive proposition, we can take the objective part of a world to be determined by the actual height of TW. For simplicity, assume that the agent's subjective state is determined by the agent's subjective estimate of TW's height (perhaps as a result of a measurement). Thus, a world can be taken to be a tuple (t, t'), where t is TW's height and t' is the agent's subjective estimate of the height. Suppose that the agent's estimate is within $\alpha/2$ of TW's actual height. Thus, the $\mathcal{R}$ relation is taken to consist of all pairs $((t, t'), (u, u'))$ of worlds such that $|t' - u'| \leq \alpha/2$. Assume that all worlds are plausible (so that $P = W$). It is then easy to check that $(M, (t, t')) \models DR(\mathbf{Tall}(\mathbf{TW}))$ iff $t \geq t^* + \alpha$. That is, the agent will definitely say that TW is Tall iff TW's true height is at least α more than the threshold t^* for tallness, since in such worlds, the agent's subjective estimate of TW's height is guaranteed to be at least $t^* + \alpha/2$.

To connect this to Williamson's model, suppose that the metric d is such that $d((t, t'), (u, u')) = |t - u|$; that is, the distance between worlds is taken to be the difference between TW's actual height in these worlds. Then it is immediate that $(M, (t, t')) \models C(\mathbf{Tall(TW)})$ iff $t \geq t^* + \alpha$. Bottom line: given all the assumptions here, the semantics of C agrees with that of DR.

Williamson does not give examples of how the metric should be interpreted. Under my interpretation, it would be more reasonable to have a different metric for each proposition. But ignoring this point, the agreement between C is DR is more than just a superficial one. For example, Williamson provides a complete axiomatization of C; it is not hard to show that if all worlds are plausible (that is, if $P = W$, so that R becomes an S5 relation), then the combination DR is characterized by exactly the same axioms as C.

Williamson suggests that a proposition φ should be taken to be vague if $\neg C\varphi$ holds is satisfiable. In Section 3, I suggested that $\neg DR\varphi$ could also be taken to be a reasonable interpretation of vagueness. Thus, I can capture much the same intuition for vagueness as Williamson, without having to make what seem to me unwarranted epistemic assumptions.

5 Discussion

I have introduced what seems to me a natural approach to dealing with intransitivity of preference and vagueness. Although various pieces of the approach seem certainly have appeared elsewhere, it seems that this particular packaging of the pieces is novel. The approach leads to a straightforward logic of vagueness, while avoiding many of the problems that have plagued other approaches.

Acknowledgments: I'd like to thank Rohit Parikh, Riccardo Pucella, and Tim Williamson for comments on a previous draft of the paper.

References

Aumann, R. J. (1976). Agreeing to disagree. *Annals of Statistics 4*(6), 1236–1239.

Bennett, B. (1998). Modal semantics for knowledge bases dealing with vague concepts. In *Principles of Knowledge Representation and Reasoning: Proc. Sixth International Conference (KR '98)*, pp. 234–244.

Black, M. (1937). Vagueness: an exercise in logical analysis. *Philosophy of Science 4*, 427–455.

Dummett, M. (1975). Wang's paradox. *Synthese 30*, 301–324.

Fagin, R., J. Y. Halpern, Y. Moses, and M. Y. Vardi (1995). *Reasoning about Knowledge*. Cambridge, Mass.: MIT Press.

Fine, K. (1975). Vagueness, truth, and logic. *Synthese 30*, 265–300.

Halpern, J. Y. and Y. Moses (1992). A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence 54*, 319–379.

Hempel, C. G. (1939). Vagueness and logic. *Philosophy of Science 6*, 163–180.

Keefe, R. (2000). *Theories of Vaguness*. Cmabridge, UK: Cambridge University Press.

Keefe, R. and P. Smith (1996). *Vaguness: A Reader*. Cambridge, Mass.: MIT Press.

Peirce, C. S. (1931–1956). *Collected Writings of Charles Sanders Peirce*. Cambridge, Mass.: Harvard University Press. Edited by C. Hartshorne, P. Weiss, et al.

Poincareé, H. (1902). *La Science et l'Hypothèse*. Paris: Flamarrion Press.

Sorenson, R. (2001). *Vagueness and Contradiction*. Oxford, UK: Oxford University Press.

van Fraassen, B. C. (1968). Presuppositions, implications, and self-reference. *Journal of Philosophy 65*, 136–152.

Williamson, T. (1994). *Vagueness*. London/New York: Routledge.

Zadeh, L. A. (1975). Fuzzy logics and approximate reasoning. *Synthese 30*, 407–428.