Categorization and Vagueness

Tamina Stephenson
Class paper for 24.729, Topics in the Philosophy of Language (Vagueness, Rayo)
Fall 2005 [Revised January 2006]

1. Introduction

In this paper I'll look at vague predicates such as *tall*, *red* and *bald*. My main claim will be that a notion like the class or category of red things, or the property or category "red" is not a primitive. The primitive notion is rather that of categorizing objects, for example, into "red" and "not red" or "red" and "orange." An object has the property of being red only by virtue of being categorized as red, generally in contrast to another color. More specifically, I'll propose that utterances using vague terms carry a presupposition that some particular categorization is being used in the context which the meaning of the vague term refers to. The various properties of vague terms will be described in terms of what kinds of categorizations they can invoke, and how the particular categorization in use can be recovered from the context.

2. Defining Categorizations

A crucial notion here will be that of a "categorization," which I will use in a certain defined sense. I will define this first in an intuitive way, and then extend the definition to include less intuitive examples of a categorization.

2.1. Example of categorizing: The book collection

Let's suppose I'm organizing my small book collection, and I've already decided that their only relevant properties are author and date of publication. For concreteness, let's suppose the books in my collection are listed in (1); each book is notated with a capital letter standing for the author followed by the year of publication.

(1) Book collection in chronological order (read down columns):

A-1593	A-1610	B-1828	C-1849	D-1990
A-1595	B-1820	C-1844	C-1850	D-1995
A-1596	B-1822	C-1846	C-1851	D-2003
A-1598	B-1825	B-1848	D-1987	

Now, there are number of ways I could sort these books. One reasonable way would be by author (A, B, C, or D). Another would be by century (16^{th}, 17^{th}, 19^{th}, 20^{th}, or 21^{st}). Or I might put them in three categories "early" (1593-1610), "middle" (1820-1851), and "late" (1987-2003). Or I might choose some year, say 1840, as an important point and separate the books into those written before 1840 and those written after 1840. Or perhaps I'd pick only the books written from 1593-1610 as "early" and the rest as "late" or only the books written from 1987-2003 as "late" and the rest as "early." Of course there are many unreasonable ways I could sort the books as well, such as according to whether they were written in an odd-numbered year or an even-numbered year. For that matter I could just randomly assign them to two or more groups.

How I'll actually sort the books depends on what I'm sorting them for. If I'm sorting them in order to arrange them in a bookcase, then the size and shape of the bookcase will be relevant. For example, if the bookcase has four shelves that each fit only five or six books (a pretty small bookcase, I admit), then I might want to sort them into categories that fit onto these shelves. One sensible way to do this would be to give one shelf to each author; or, if I were more concerned about time period, I might notice that the books by authors B and C fall into two clusters – those from the 1820s and those from the 1840's and 1850's – and divide those books between two shelves accordingly. And, of course, if I just cared about putting the books away, I could arrange them randomly.

2.2. Definition of a categorization

On my view, a categorization is any way of sorting a set of objects into groups. Formally, we can think of this as a pairing of a set of objects S and a finite set $\{s_1,\ldots s_n\}$ of subsets of S, such that each member of S is a member of exactly one of $\{s_1,\ldots s_n\}$.[1] (This is the same as a partition of a set.) The subsets s_i can intuitively be thought of as categories. For example, the categorization of books from (1) by author is represented in (2.a), and one of the categorizations by time period is represented in (2.b). (Let S = the set of books in (1) above.)

[1] That is, $\{s_1,\ldots s_n\}$ must have a finite number of member sets; nothing stops the set S or its subsets from themselves being infinite.

(2) (a) **A categorization by author:**

$$S \Rightarrow \left\{ \begin{array}{l} \{A\text{-}1593, A\text{-}1595, A\text{-}1596, A\text{-}1598, A\text{-}1610\}, \\ \{B\text{-}1820, B\text{-}1822, B\text{-}1825, B\text{-}1828, B\text{-}1848\}, \\ \{C\text{-}1844, C\text{-}1846, C\text{-}1849, C\text{-}1850, C\text{-}1851\}, \\ \{D\text{-}1987, D\text{-}1990, D\text{-}1995, D\text{-}2003\} \end{array} \right\}$$

(b) **A categorization by time period:**

$$S \Rightarrow \left\{ \begin{array}{l} \{A\text{-}1593, A\text{-}1595, A\text{-}1596, A\text{-}1598, A\text{-}1610\}, \\ \{B\text{-}1820, B\text{-}1822, B\text{-}1825, B\text{-}1828\}, \\ \{C\text{-}1844, C\text{-}1846, B\text{-}1848, C\text{-}1849, C\text{-}1850, C\text{-}1851\}, \\ \{D\text{-}1987, D\text{-}1990, D\text{-}1995, D\text{-}2003\} \end{array} \right\}$$

The pairings in (3) also represent categorizations of S although they don't follow any apparent pattern.

(3) (a)

$$S \Rightarrow \left\{ \begin{array}{l} \{A\text{-}1593, B\text{-}1820, C\text{-}1844, C\text{-}1846, D\text{-}1987, D\text{-}1995, D\text{-}2003\}, \\ \{A\text{-}1596, A\text{-}1598, A\text{-}1610, B\text{-}1822, D\text{-}1990\}, \\ \{A\text{-}1595, B\text{-}1825, B\text{-}1828, B\text{-}1848, C\text{-}1849, C\text{-}1850, C\text{-}1851\} \end{array} \right\}$$

(b)

$$S \Rightarrow \left\{ \begin{array}{l} \{A\text{-}1593, A\text{-}1598, A\text{-}1610, D\text{-}1987\}, \\ \{A\text{-}1595, B\text{-}1820, C\text{-}1846\}, \\ \{B\text{-}1825, B\text{-}1828, C\text{-}1844, B\text{-}1848, C\text{-}1849, C\text{-}1850\}, \\ \{A\text{-}1596, B\text{-}1822, C\text{-}1851, D\text{-}1990, D\text{-}1995, D\text{-}2003\} \end{array} \right\}$$

Although the categorizations in (3) are apparently random, that doesn't mean they couldn't be useful. Remember, it could be that my only concern is putting the books away in a bookcase, which requires deciding which books will go on which shelves. If there were three shelves I might end up with something like (3.a), and if there were four shelves I might end up with (3.b). Of course we couldn't describe these categorizations as being "according to author" or "according to time period," but we could describe them, if we needed to, as "according to which shelf they're going to go on."

Here's one more example of a categorization. Suppose someone is pouring you a glass of water, and tells you to say when to stop pouring. At some point you say "that's enough" but it takes a few seconds for the pourer to react and stop pouring. (Fortunately you expect this to happen and leave enough time to prevent the glass from overflowing.) Now, consider the set of all amounts of water that the glass contains at some point during the pouring, from none to the amount there when the pouring stops. If the pouring is continuous, this is an infinite set (unless we take molecular structure into account, in which case it's just very large.) Your pronouncement "that's enough" defines a categorization of this set into those amounts that are sufficient for your needs and desires and those that are insufficient. One subset contains the amounts that came before your pronouncement, and the other contains the ones from your pronouncement on.[2]

At this point, categorizations have nothing to do with language, or even with properties of the things being categorized. We can describe a categorization as relating to the sufficiency of an amount, or the author of a book, but so far these are just convenient descriptions. In the next section I'll introduce a more meaningful way of describing categorizations.

3. Categorizations that Respect Orders

So far I've been talking informally about categorizations of books "according to author" or "according to period." In this section I'd like to make that idea a little more precise.

3.1. The role of orders

I assume that properties that can be used to categorize things (in an intuitive sense), such as a book's author or year of publication, can do this by virtue of defining one or more (weak) orders on some domain. For example, the year of publication of books defines the order illustrated in (4), defined on the set of books S from above.

(4) Order defined by year of publication in S:
A-1593 ≤ A-1595 ≤ A-1596 ≤ A-1598 ≤ A-1610 ≤ B-1820 ≤ B-1820 ≤ B-1822 ≤ B-1825 ≤ B-1828 ≤ C-1844 ≤ C-1846 ≤ B-1848 ≤ C-1849 ≤ C-1850 ≤ C-1851 ≤ D-1987 ≤ D-1990 ≤ D-1995 ≤ D-2003

I use the symbol "≤" in (4) because in general the orders I use will be weak orders, that is, orders that are reflexive and antisymmetric rather than irreflexive and asymmetric.[3] This is because I suspect that it might be desirable to make a distinction between two objects that are equivalent with respect to a certain property (e.g., two reddish color chips that are identical and therefore equally red) and two objects that cannot be compared with

[2] If you're worried about the fact that it takes time to say "that's enough" we could say that the relevant point in time is either the moment when the utterance starts or the moment it ends, but this isn't important.
[3] For mathematical definitions related to orders, one good source is Partee, ter Meulen, & Wall (1993).

respect to a property (e.g., a prototypically blue color chip and a prototypically green one, which arguably can't be compared with respect to redness). But this won't be crucial to any of the examples I'll use.

It's less obvious how an order could be defined on the books in S by who their authors are, but there's a mundane way to do this if we allow a set of orders: we can define different orders $\leq_X$ for each author X, as in (5).

(5) Orders defined by author:
Def.: The set of orders $\{\leq_X \mid X\in \{A, B, C, D\}\}$ such that for any $y,z\in S$,
(i) $y\leq_X y$
(ii) If y and z have the same author, then $y\leq_X z$ and $z\leq_X y$
(iii) If X is the author of y but not of z, then $y\leq_X z$ but **not** $z\leq_X y$

I'd like to introduce two bits of shorthand: I'll write "$a=_x b$" to mean $a\leq_x b$ and $b\leq_x a$; and I'll write "$a<_x b$" to mean that $a\leq_x b$ but that it's not the case that $b\leq_x a$. Then the order corresponding to author A can be represented as in (6).

(6) Order defined by author A ($\leq_A$):
A-1593 = A-1595 = A-1596 = A-1598 = A-1610
A-1593 < B-1820 … B-1848, C-1844 … C-1851, D-1987 … D-2003
A-1595 < B-1820 … B-1848, C-1844 … C-1851, D-1987 … D-2003
…
A-1610 < B-1820 … B-1848, C-1844 … C-1851, D-1987 … D-2003
B-1820 = … = B-1848 = C-1844 = … = C-1851 = D-1987 = … = D-2003

This kind of order simply has the effect of dividing the books into two equivalence classes – the books written by that author and all the others.

3.2. Respecting a set of orders

I'll say that a categorization of a set A respects a set of orders $\{\leq_1, \leq_2, \ldots \leq_n\}$ if, informally speaking, all the borders between categories are set so as to divide up the objects along points in the orders. I give this a more formal characterization in (7).

(7) Given a set **A**, a set **O** of orders $\{\leq_1, \ldots, \leq_n\}$,
and **a categorization C** of A into categories $c_1, \ldots c_n$,
C respects O iff for any $x, y, z\in A$:

(a) If for all orders $\leq_i\in O$, $x=_i y$, then x and y must be in the same category.
(b) If x and y are in one category c_i, and z is in a different category c_j, then there can be no order $\leq_i\in O$ such that $x<_i z<_i y$.

Of course, (7) may need to be modified or augmented, but I'll take this as a starting point. The condition in (7.a) ensures that two objects that are equivalent in all relevant ways cannot be categorized differently. For example, if only properties relating to color are being considered, then two (solid-colored) objects with the exact same color must be put in the same category. The condition in (7.b) requires that if a line is drawn at some point on an order, then nothing on one side of the line can be put into the same category as anything on the other side of the line (even if they have other properties in common). When only one order ≤ is relevant, I'll talk interchangeably of respecting the singleton set of orders {≤} and respecting the order ≤.

3.3. Examples

Using the same example of the book collection, let me give some examples of categorizations that would or would not respect certain orders. First, let's consider categorizations that respect the order imposed by year of publication, which was illustrated in (4). There are many such categorizations, a few of which have already been discussed. Two more are illustrated in (8).

(8) Categorizations of S that respect the order of year of publication

(a) S ➔

$$\left\{\begin{array}{l}\{\text{A-1593, A-1595, A-1596, A-1598, A-1610, B-1820, B-1822}\}\ ,\\ \{\text{B-1825, B-1828, C-1844, C-1846, B-1848, C-1849, C-1850}\}\ ,\\ \{\text{C-1851, D-1987, D-1990, D-1995, D-2003}\}\end{array}\right\}$$

(b) S ➔

$$\left\{\begin{array}{l}\{\text{A-1593, A-1595, A-1596, A-1598, A-1610, B-1820, B-1822,}\\ \quad\text{B-1825, B-1828, C-1844, C-1846, B-1848, C-1849, C-1850,}\\ \quad\text{C-1851}\}\ ,\\ \{\text{D-1987, D-1990, D-1995, D-2003}\}\end{array}\right\}$$

The categorization in (8.a) divides the books into three periods, with the first cutoff point being between 1822 and 1825, and the second cutoff point being between 1850 and 1851. The one in (8.b) divides the books into two periods, with the cutoff point being between 1851 and 1987. Notice that there need not be any large or significant gap around the cutoff points for a categorization to respect a order; later I'll introduce a new notion of a "natural" categorization which will include requirements about the gaps (see Section 5.1).

The categorizations in (9), on the other hand, do not respect the order defined by year of publication.

(9) Categorizations of S that <u>do not</u> respect the order of year of publication

(a) S ➔

$$\left\{ \begin{array}{l} \{\text{A-1593, A-1595, A-1596, B-1820, B-1822}\}\ , \\ \{\text{A-1598, A-1610, B-1825, B-1828, C-1844, C-1846, B-1848}\}\ , \\ \{\text{C-1849, C-1850, C-1851, D-1987, D-1990, D-1995, D-2003}\} \end{array} \right\}$$

(b) S ➔

$$\left\{ \begin{array}{l} \{\text{A-1593, A-1595, A-1596, A-1598, A-1610, D-1987, D-1990,} \\ \quad \text{D-1995, D-2003}\}\ , \\ \{\text{B-1820, B-1822, B-1825, B-1828, C-1844, C-1846, B-1848,} \\ \quad \text{C-1849, C-1850, C-1851}\} \end{array} \right\}$$

The categorization in (9.a) fails to follow the order of year of publication because some items are put in the same category even though they appear on different sides of a line that's drawn elsewhere. For example, A-1596 and B-1820 are put in the same category, and A-1610 is put in a different category, even though A-1596 < A-1610 < B-1820 – thus violating the condition in (7.b). The categorization in (9.b) fails for an exactly parallel reason; for example, we could use A-1610, D-1987, and C-1850 as a falsifying instance. Note that it doesn't matter that (9.b) makes a coherent classification relating to time period, namely, 19th-century books vs. non-19th-century books, since the order we're concerned with is the particular one defined in (4), which corresponds to temporal precedence. (If we instead defined an order in terms of whether or not a book was written in the 19th century, akin to the orders defined for authors, then (9.b) would respect that order.)

To show how a categorization can fail to respect an order by virtue of the other condition (the one that says that equivalent items must be in the same category), let's consider the set of orders I defined for author. First, let's consider the set of all four orders $\{\leq_A, \leq_B, \leq_C, \leq_D\}$. The categorizations shown in (10) respect this set of orders.

(10) Categorizations of S that respect the set of orders $\{\leq_A, \leq_B, \leq_C, \leq_D\}$

(a) S ➔

$$\left\{ \begin{array}{l} \{\text{A-1593, A-1595, A-1596, A-1598, A-1610}\}\ , \\ \{\text{B-1820, B-1822, B-1825, B-1828, B-1848}\}\ , \\ \{\text{C-1844, C-1846, C-1849, C-1850, C-1851}\}\ , \\ \{\text{D-1987, D-1990, D-1995, D-2003}\} \end{array} \right\}$$

(b) S ➔

$$\left\{\begin{array}{l}\{\text{A-1593, A-1595, A-1596, A-1598, A-1610, B-1820, B-1822,}\\ \quad\text{B-1825, B-1828, B-1848}\}\ ,\\ \{\text{C-1844, C-1846, C-1849, C-1850, C-1851}\}\ ,\\ \{\text{D-1987, D-1990, D-1995, D-2003}\}\end{array}\right\}$$

The categorization in (10.a) is just the obvious one where each author corresponds to a category. The one in (10.b) is the same except that authors A and B are lumped together into one category. Nothing in my definition prevents this.

Two categorizations that fail to respect this set of orders are illustrated in (11).

(11) Categorizations of S that <u>do not</u> respect the set of orders $\{\leq_A, \leq_B, \leq_C, \leq_D\}$

(a) S ➔

$$\left\{\begin{array}{l}\{\text{A-1593, A-1595, A-1596, A-1598}\}\ ,\quad \{\text{A-1610}\}\ ,\\ \{\text{B-1820, B-1822, B-1825, B-1828, B-1848}\}\ ,\\ \{\text{C-1844, C-1846, C-1849, C-1850, C-1851}\}\ ,\\ \{\text{D-1987, D-1990, D-1995, D-2003}\}\end{array}\right\}$$

(b) S ➔

$$\left\{\begin{array}{l}\{\text{A-1593, A-1595, A-1596, A-1598, A-1610, B-1820}\}\ ,\\ \{\text{B-1822, B-1825, B-1828, B-1848}\}\ ,\\ \{\text{C-1844, C-1846, C-1849, C-1850, C-1851}\}\ ,\\ \{\text{D-1987, D-1990, D-1995, D-2003}\}\end{array}\right\}$$

Both categorizations in (11) violate the condition that equivalent objects must be in the same category, in (11.a) because the books by author A aren't all together and in (11.b) because the books by author B aren't all together.

If we don't include all four orders in the set, the situation is somewhat different. Consider the set of just two orders $\{\leq_A, \leq_B\}$. Under this set of orders, all the books by authors C and D are equivalent and so must be in the same category. This means, for example, that the categorization in (10.a) (which puts the books in four categories according to author) doesn't follow the set of orders $\{\leq_A, \leq_B\}$ because the books by authors C and D aren't in the same category. The categorizations in (11), which fail to respect the set of all four orders, also fail to respect the set of orders $\{\leq_A, \leq_B\}$.

Finally, let's go back to the original situation with the collection of books, and suppose that both time period and author are relevant. That is, consider the set of five orders $\{\leq_A, \leq_B, \leq_C, \leq_D, \leq_t\}$, where $\leq_t$ is the order defined by year of publication (from (4) above). The categorizations in (12) all respect this set of orders.

(12) Categorizations of S that respect the set of orders $\{\leq_A, \leq_B, \leq_C, \leq_D, \leq_t\}$

(a) S ➔

$$\left\{\begin{array}{l} \{\text{A-1593, A-1595, A-1596, A-1598, A-1610}\}, \\ \{\text{B-1820, B-1822, B-1825, B-1828, B-1848, C-1844, C-1846,} \\ \quad \text{C-1849, C-1850, C-1851}\}, \\ \{\text{D-1987, D-1990, D-1995, D-2003}\} \end{array}\right\}$$

(b) S ➔

$$\left\{\begin{array}{ll} \{\text{A-1593, A-1595, A-1596}\}, & \{\text{A-1598, A-1610}\}, \\ \{\text{B-1820, B-1822, B-1825, B-1828}\}, & \{\text{C-1844, C-1846}\}, \\ \{\text{B-1848}\}, & \{\text{C-1849, C-1850, C-1851}\}, \\ \{\text{D-1987, D-1990}\}, & \{\text{D-1995, D-2003}\} \end{array}\right\}$$

(c) S ➔

$$\left\{\begin{array}{l} \{\text{A-1593, A-1595, A-1596, A-1598, A-1610}\}, \\ \{\text{B-1820, B-1822, B-1825, B-1828}\}, \\ \{\text{C-1844, C-1846, B-1848, C-1849, C-1850, C-1851}\}, \\ \{\text{D-1987, D-1990, D-1995, D-2003}\} \end{array}\right\}$$

Notice that now books by the same author need not be put in the same category, as long as they're divided by time. This is illustrated with the books by authors A and D in (12.b). In fact, the condition requiring equivalent items to be in the same category in effect no longer applies, because no two items are equivalent under the order $\leq_t$. (Now two books would only be equivalent if they were written in the same year by the same author.) This means that even a categorization like (12.c) now respects the orders even though the single item B-1848 is put with books by author C and separated from the other books by author B, since it respects time.

On the other hand, any categorization that fails to respect the order $\leq_t$ will also fail to respect the set of orders $\{\leq_A, \leq_B, \leq_C, \leq_D, \leq_t\}$. For example, the categorizations in (13) fail to respect this set of orders.

(13) Categorizations of S that <u>do not</u> respect $\{\leq_A, \leq_B, \leq_C, \leq_D, \leq_t\}$

(a) S ➔

$$\left\{\begin{array}{l} \{\text{A-1593, A-1595, A-1596, A-1598, A-1610}\}, \\ \{\text{B-1820, B-1822, B-1825, B-1828, B-1848}\}, \\ \{\text{C-1844, C-1846, C-1849, C-1850, C-1851}\}, \\ \{\text{D-1987, D-1990, D-1995, D-2003}\} \end{array}\right\}$$

(b) $S \Rightarrow$

$$\left\{ \begin{array}{l} \{\text{A-1593, A-1610}\}, \quad \{\text{A-1595, A-1596, A-1598}\}, \\ \{\text{B-1820, B-1822, B-1825, B-1828, B-1848, C-1844, C-1846,} \\ \quad \text{C-1849, C-1850, C-1851}\}, \\ \{\text{D-1987, D-1990, D-1995, D-2003}\} \end{array} \right\}$$

(13.a) fails to respect the order $\leq_t$ because, for example, B-1828 and B-1848 are put into the same category with C-1844 in another, where B-1828 $<_t$ C-1844 $<_t$ B-1848. (13.c) fails similarly because, for example, A-1593 $<_t$ A-1596 $<_t$ A-1610.

3.4. Summary

At this point I have explained the notion of categorization that I'd like to use, and given a sense in which a categorization can be coherent or sensible, in terms of respecting orders defined on sets of objects. So far I've said nothing about the semantics of vague predicates, and in fact nothing about language or any other kind of representation where vagueness might show up. That's what I will turn to next.

4. The Semantics of Vague Predicates

In this section I'll present my proposed semantics for vague predicates such as *tall*, *red*, and *bald*. I'll start by giving the general form and then say more about specific predicates.

4.1. The general form

The semantics I propose for predicates such as *tall*, *red*, or *bald* has the general form shown in (14):

(14) For a (vague) predicate P and individual x:
Let C = the categorization in use (of the right kind).

"x is P" is true iff C puts x in the right category.

Of course the immediate question to ask about (14) is what is meant by the "right kind" of categorization and the "right category." These may share some general properties, but for the most part will depend on lexical properties of each particular predicate, for which I'll give examples.

I'd also like to bring attention to the definite description "the categorization in use." I intend this to express a presupposition that such a categorization exists. In other words, my claim is that, for any vague predicate P, an assertion of the form "x is P" is only felicitous if there's a categorization of the right kind for P that's understood as the one being used in the context of utterance. I'll come back to the question of what is required for a categorization to satisfy this condition. Right now the crucial point is just that asserting "x is P" requires the existence of a categorization in the same way that asserting "x's mother is Q" requires that x have a mother or that asserting "she is R" requires that there be a contextually salient female individual for *she* to refer to.

4.2. Scalar predicates: *tall* and *big*

I propose that *tall* has the semantics in (15), a specific instance of the schema in (14).

(15) <u>Semantics for *tall*</u>
Let S = a (finite) set of relevant individuals including x.
Let C = the categorization of S in use that respects the order $\leq_{\text{height}}$
where for any y and z, y $\leq_{\text{height}}$ z iff z is at least as tall as y.

"x is tall" is true iff C assigns x to a category containing a maximal element(s) in S according to the order $\leq_{\text{height}}$.

The meaning of *tall*, then, requires a categorization that divides the set of relevant individuals up into one or more groups according to height, and an individual is tall if it is assigned to the same category as the tallest individual(s). Given the requirement that the set be finite and the way I've defined the order $\leq_{\text{height}}$, there must be one or more maximal elements; if there's more than one, they have to be put in the same category for the order to be respected.

This meaning allows for a lot of context dependency in the comparison class, and relative freedom in what height an individual needs to have to be called tall, of the kind that has been widely observed. On the other hand, there will be some limits on this freedom given that the hearer needs to be able to figure out what categorization is being used, and therefore it must be salient. I'll defer a more thorough discussion of what makes a categorization salient until Section 5, but for now let's follow an idea in Rayo (2005) and say that a categorization is salient if there are big enough gaps between the categories. On this view, then, *tall* can only be understood if there's a big enough gap in heights between the group of individuals categorized with a maximal element and those in the next shorter category. (Actually I'll suggest that a categorization can be usable for other reasons as well, but again I'll come back to that.)

I assume that other scalar predicates have largely parallel meanings, with differences only in the types of orders that must be respected. For example, *big* can be given the meaning in (16).

(16) Semantics for *big*
Let S = a (finite) set of relevant individuals including x.
Let C = the categorization of S in use that respects a set of orders O
that relate to dimensions of size
(where $x \leq_D y$ iff y is at least as big as x along dimension D).

"x is big" is true iff C assigns x to a category containing a maximal element(s) in S according to one of the orders in O.

The predicate *big* differs from *tall* in allowing a set of orders rather than a single order, since there can be different dimensions of bigness. For example, if we take the dimensions of bigness to be weight and volume, then the set of orders would be $\{\leq_{weight}, \leq_{volume}\}$, where $x \leq_{weight} y$ iff y weighs at least as much as x, and $x \leq_{volume} y$ iff y has at least as much volume as x. If both weight and volume are relevant, then it stands to reason that the largest individual by weight and the largest by volume should both be considered "big" (even though they might be put in different categories) and so an individual is big just in case it's assigned to the same category as either of these.

4.3. An additional requirement for some vague terms

The semantics outlined for *big* and *tall* above allow for a lot of freedom in where the line is drawn for when an individual is big or tall. I think this is more or less correct for these kind of scalar predicates, because people do have a considerable amount of freedom in what to call big or what to call tall. Some other vague terms seem to allow less freedom, however. For example, for something to be called red, it really has to be close enough to a certain color. You can't call a pumpkin red just because it's only being compared to yellow squash, or call a green apple blue just because all the other apples are red or yellow. There is some requirement of a more absolute kind where color terms are concerned.

I'd like to capture this requirement as follows: first, let's assume that each (basic) color term is associated with a prototype or exemplar for that color. When a categorization is made that is to draw the line for that color, the set being categorized must include the exemplar. In other words, any comparison class for color will contain phantom elements matching the prototype or exemplar for each relevant color term. Another way of thinking of this is that whenever you use the word *red* (for example), there's an imaginary individual matching focal red that's always taken to be relevant.[4] Thus my view of the difference between highly context-dependent vague terms like *tall* and less context-dependent ones like *red* is in a sense the opposite of what has been assumed. Predicates like *tall* are often assumed to contain an extra argument or parameter in the form of a "comparison class." (See, e.g., Chierchia & McConnell-Ginet 2000, p. 464.) On my view, all vague predicates require a comparison class, but the less context-dependent ones like *red* have additional restrictions on what the comparison class can look like. In other words, context dependency is a normal byproduct of the general semantics of vague terms, but in some cases it is obscured by other properties of particular items.

Taking this additional requirement into account, the semantics I propose for *red* is as in (17).

(17) <u>Semantics for *red*</u>
Let P_{red} = the prototype for red.
Let S = a set of relevant individuals including x.
Let C = the categorization of $S \cap \{P\}$ in use that respects a set of orders O that relate to color, including any that help determine redness.

"x is red" is true iff C assigns x to the same category as P_{red}.

I won't make any claims about exactly what orders there are that are related to color. These are just whatever properties in fact determine our perception of color, and what these are is a question for the science of vision. The added requirement that the set of orders must include any that "help determine redness" is also a placeholder for something that science will determine, and just means that the term *red* may require some specific orders to be included.

If we assume, as I mentioned earlier, that the hearer can only recover the categorization being used if there are big enough gaps between categories, this means that

[4] If you're uncomfortable with the idea of having a single exemplar, we could say that there's a cluster of exemplars that all have to be included in the categorization, and all have to be put in the same category.

"x is red" can only be understood to be true if x is close enough to the prototype for red that there can be a big enough gap between the category containing x and the prototype and the next category.

I assume that *bald* is like a color term in having a prototype or exemplar – namely, a person with no hair on their head at all. Thus we can give *bald* the meaning in (18).

(18) Semantics for *bald*
Let P_{bald} = the prototype for bald (a person with no hair on their head)
Let S = a set of relevant individuals including x.
Let C = the categorization of $S \cap \{P\}$ in use that respects a set of orders O that relate to the fullness of the hair on people's heads.

"x is bald" is true iff C assigns x to the same category as P_{bald}.

I allow for a set of orders here since there could be more than one dimension to take into account for baldness, such as the thickness and distribution of hair, or the number of hairs and their coarseness, with any of these properties possibly relative to the size of the head. Again, it's an empirical question just what properties are relevant to our perception of the amount of hair people have on their heads, and I take the answer not to be particularly relevant to linguistics or the philosophy of language.

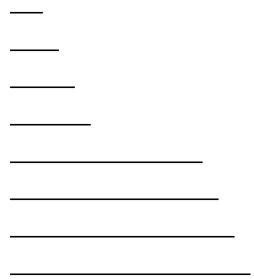
5. Usable Categorizations

On the view I'm developing, any use of a vague predicate carries a presupposition that some particular kind of categorization is being used. Thus in order for communication with vague terms to be successful, the hearer must be able to figure out which categorization this is. I suggest that there are two main ways this can happen. The first is if a particular categorization is salient, and the second is if the use of the vague term is serving to define the categorization rather than simply referring to it. When neither of these conditions hold, a third strategy is used that is reminiscent of supervaluation. I'll discuss these one at a time.

5.1. Natural categorizations

I've already suggested, following Rayo (2005), that a categorization can be salient if there are big enough "gaps" between the categories. I'd like to try to elaborate and clarify this idea. Consider the set of lines drawn in (19). An order $\leq_{length}$ is defined on this set as in (20).

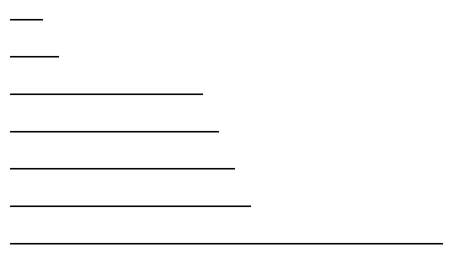
(19) (a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)



(20) Order $\leq_{length}$ defined on (19):
$(a)<_{length}(b)<_{length}(c)<_{length}(d)<_{length}(e)<_{length}(f)<_{length}(g)<_{length}(h)$

Now, any categorization that simply puts boundaries in one or more places between consecutive lines will respect the order $\leq_{length}$. But there's one particular categorization that jumps out at you, namely the one that puts the lines in (a)-(d) in one category and (e)-(h) in another. This is because the lines in (d) and (e) differ much more in length than the other pairs of consecutive lines. I'll call this kind of categorization a "natural categorization," or "natural with respect to length." (Technically, the naturalness will not be with respect to length but rather to a measurement of length.) Of course, there can be natural categorizations that divide a set into three or more categories. For example, the lines in (21) have a natural categorization into three groups: (a)-(b), (c)-(f), and (g)-(h).

(21) (a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)



What makes these categorizations natural is not just that there are large gaps between the categories, but also that there are no large gaps within categories. In the case of lines ordered by length, that is, consecutive lines within the same category have only small differences in length, whereas consecutive lines in different categories have large differences in length. There could even be a category where the longest line in that category was much longer than the shortest line in that category, as long as they were connected by a sequence of lines that differed little from their neighbors. For example, I think the lines in (22) have a natural categorization in two groups, (a)-(b) and (c)-(l), even though the difference in length between (c) and (l) is at least as great as that between (b) and (c).

(22) (a) ——
(b) ———
(c) ——————————
(d) ———————————
(e) ————————————
(f) —————————————
(g) ——————————————
(h) ————————————————
(i) —————————————————
(j) ——————————————————
(k) ———————————————————
(l) ————————————————————

Taking all these considerations into account, we can give a rough formal definition of a "natural" categorization as in (23). This applies to cases where only one order is relevant, such as the examples above using lines.

(23) <u>Definition of a natural categorization</u> [preliminary]:
Let S be a set on which an order $\leq_1$ is defined.
Let C be a categorization of S that respects the order $\leq_1$.
Let M be a measurement function from pairs $<x,y>$, such that
    M is defined iff $x \in S$, $y \in S$, and $x \leq_1 y$, and for any such pair $<x,y>$,
    $M(<x,y>)$ is a real number that represents the difference
    between x and y along the dimension of $\leq_1$.
For any number ε, let R^{ε} be a relation such that $xR^{\varepsilon}y$ iff $M(<x,y>) < \varepsilon$
    (and is defined), and let TC(R) be the transitive closure of R.
Then:

C is natural with respect to M iff there is some number ε such that
    C puts x and y in the same category iff $<x,y> \in TC(R^{\varepsilon})$.

This essentially means that if you arrange the items so that pairs with the least difference are next to each other, then a categorization is natural just in case it puts boundaries in places where adjacent items differ by more than a certain amount. Informally speaking, this is done by setting up a relation between items that differ by no more than ε (which will be a reflexive and symmetric relation). Then this relation is closed under transitivity.

This yields a reflexive, symmetric, and transitive relation, also called an equivalence relation. Thus a categorization is natural if it matches up with an equivalence relation that can be set up in this way.

To extend this definition to cases where more than one order is relevant, the definition can be generalized along the lines of (24).

(24) Definition of a natural categorization [general]:
Let S be a set on which the set of orders $\{\leq_1, \ldots, \leq_n\}$ are defined.
Let C be a categorization of S that respects the set of orders $\{\leq_1, \ldots, \leq_n\}$.
For i=1, …, n, let M_i be a function from pairs $<x,y>$, such that
M_i is defined iff $x \in S$, $y \in S$, and $x \leq_i y$, and for any such pair $<x,y>$, $M_i(<x,y>)$ is a real number that represents the difference between x and y along the dimension of $\leq_i$.
For any number ε, let R_i^{ε} be a relation such that $x R_i^{\varepsilon} y$ iff $M_i(<x,y>) < \varepsilon$ (and is defined), and let TC(R) be the transitive closure of R.
Then:

C is natural w.r.t. $\{M_1, \ldots, M_n\}$ iff there is some set $\{\varepsilon_1, \ldots \varepsilon_n\}$ such that C puts x and y in the same category iff $<x,y> \in TC(\cup\{R_1^{\varepsilon_1}, \ldots, R_n^{\varepsilon_n}\})$.

This can be thought of as setting up a set of (reflexive and symmetric) relations that hold between items that differ no more than a certain amount on a particular dimension. Then these relations are combined together into one relation (which is thus also reflexive and symmetric). Finally, this combined relation is closed under transitivity, yielding an equivalence relation. In the general case, then, a categorization is natural if it matches up with an equivalence relation that can be set up in this more general way.

I've had to introduce the notion of a measurement function into my definitions in (23) and (24). This is because some way of measuring is needed to make sense of two individuals differing by no more than a certain amount. For a predicate like *tall*, the measurement function can straightforwardly be identified with one of the ways that we actually measure height – for example, using feet. It's less, clear, though, what kind of measurement function would be used for a predicate like *(is a) table*.[5] While some artifacts are clearly more table-like than others, it's hard to see how the differences could be quantified, and thus how an equivalence relation could be established that would give rise to a natural categorization. I don't believe this poses a serious problem for my view, however. All that's really needed to set up the right kind of equivalence relation is a set

[5] Thanks to Agustín Rayo for pointing this out.

of similarity relations corresponding to the different orders being used. Each similarity relation has to follow its corresponding order in a certain way; specifically, if the similarity relation for order $<_i$ holds between a and c, and $a<_i b<_i c$, then the relation has to hold between a and b and between b and c as well. (Of course, the similarity relations must also be reflexive and symmetric.) Similarly relations could take the place of the relations $R_i^{\varepsilon_i}$. I assume that we do have ways in practice of establishing similarity relations on, for example, sets of table-like artifacts. How we actually do that is another question for cognitive science.

This notion of a natural categorization is still incomplete. For example, ultimately we would need to say more about how measurement functions work and what their properties are. But it seems clear that there is some sense in which a categorization can be natural, or obvious, or salient, by virtue of the kinds of properties I've described. Given some notion of naturalness along these lines, we can say that a categorization is usable in a context if it is natural. Put differently, when a speaker makes an assertion using a vague predicate, the hearer will generally be able to recover which categorization the speaker is referring to if the categorization is natural.

5.2. Assertions defining categorizations

I suggest that there's another, completely different kind of situation where a categorization can be usable. This is when an assertion referring to the categorization is itself serving to define or set up the categorization rather than describing it. For example, let's go back to the example of the book collection. Suppose that I was putting my books away, but that I had to divide them between two different shelves in two different rooms. I decide that I want to arrange them chronologically, so that all the books in one room are from an earlier period and all the books in the other room are from a later period. It turns out that the books fit exactly on the two shelves, so I don't have any choice as to where to draw the boundary between "early" and "late" books. Let's suppose that the last book that fits on the "early" shelf is from 1822, and the next book in chronological order is from 1825. Now, imagine that a friend is at my house and wants to find an (imaginary) book titled *Domestic Tranquility*. She knows that this book was written in 1850, and knows that I've arranged the books chronologically, but doesn't know where I've set the boundary between early and late. Furthermore, all these things are common ground. We could imagine that a dialogue like (25) would ensue.

(25) A: Is *Domestic Tranquility* early or late?
B: It's late.

It seems to me that speaker B isn't making any particular claims about any inherent property of the book. More importantly, it's not necessary that the categorization of books into "early" and "late" be natural – in fact, in the example I set up earlier, it's clear that a categorization of books into those written in 1822 or earlier versus those written after 1822 could not be natural, since there are much larger gaps in time at other points in my collection. (For example, there are no books between 1610 and 1820, and no books between 1851 and 1987.) This isn't a problem in this situation because speaker A doesn't need to recover which categorization speaker B is referring to. Rather speaker B's assertion tells speaker A something about the relevant categorization, namely that it puts a particular item in a particular category.

This can also account for the situation where someone is pouring a glass of water for someone else and asks them to say when they've poured enough. If "enough" is a vague predicate, then it ought to presuppose that there is a usable categorization of the different amounts of water into "enough" and "not enough." But if the water is being poured at a constant rate, there couldn't possibly be a natural categorization of the amounts. However, in this case when the speaker says, "That's enough!," he is thereby telling the hearer what the relevant categorization is – all amounts that came before his assertion are not enough, and those that came during or after the assertion are enough.

At this point, there are two completely different ways that a categorization can be made usable. It would presumably be desirable if they could be unified, or at least if their differences could be explained independently. I'll return to this in Section 5.4.

5.3. The fallback strategy

I've claimed that a categorization is usable if it's natural or if it's being defined by the relevant assertion. For now I'll assume that there are no other conditions where a categorization can be usable, but if additional cases were identified that would not affect what I'm going to say. I suggest that there is one additional interpretive strategy that can be used when there is no usable categorization. In this case, the hearer assumes that whatever categorization is being used agrees to a great extent with categorizations that would be usable in other contexts containing some of the relevant items. For example, suppose a speaker says "Patch #1 is red," and the hearer doesn't know what patch #1 or the other patches being categorized look like. The hearer falls back on the assumption that this categorization agrees with a large majority of natural categorizations that could be made on sets containing patch #1. This must mean, then, that patch #1 is very close to the prototype for red, so that it would be hard to imagine a situation where there would be a large gap in color between the prototype and patch #1.

This fallback strategy is reminiscent of supervaluation accounts of vague scalar terms in that a set of different categorizations are being compared and used to draw a conclusion about a certain item. This is crucially different from introducing supervaluations, however, in that they are not part of the semantics at all, on my view, but rather a strategy used by language users to draw conclusions from utterances that are in a certain sense defective – which refer to an object that the hearer can't pick out from the context.[6] The option of the fallback strategy will be crucial to my explanation of the sorites paradox in Section 6.

Of course, speakers know what tools hearers have at their disposal, including the fallback strategy. Therefore the availability of this strategy also allows a speaker to make an assertion using a vague term even knowing that the assertion will suffer from reference failure. If patch #1 is in fact very close to prototypical red, for example, then a speaker can assert that it is red without going to the trouble of deciding which categorization to refer to and making sure it's salient. The speaker knows that the hearer will conclude that patch #1 is very close to prototypical red, the hearer draws this conclusion, and communication is successful despite the failure to refer to a particular categorization. This is analogous to a situation where a person who is known to have two dogs tells a friend, "My dog died yesterday." If the friend doesn't know the dogs well enough to be particularly attached to either of them (or to be particularly surprised that one of them died instead of the other) then it may not matter to him which dog died, because either way he'll draw the same conclusions: that the speaker has lost a beloved pet and needs his sympathy. Under the assumption that the possessive contains a hidden

[6] Another possibility might be to say that in the kind of contexts where I'm claiming that this fallback strategy is used, the vague term actually has a generic interpretation along the lines of "in normal contexts, the categorization into P and not P classifies x as P," where the presupposition that there is a usable categorization restricts the quantification. (It's widely argued that presuppositions restrict quantification: see, e.g., Schubert & Pelletier 1987.) I'm hesitant to make this move, however, because this would seem to involve quantification over contexts of utterance rather than situations in any standard sense.

definite article or for some other reason presupposes a unique referent, the same thing is going on here: the speaker hasn't taken the trouble to specify which dog it is that died, but she knows that the hearer will draw the same conclusions as if she had.

Important questions remain about the fallback strategy. For example, we would need to find a way to constrain its use so that it only applies in appropriate contexts; in particular, we would want to make sure that it can't apply in contexts where a natural categorization is available or where the categorization is being defined by an utterance. (Merely calling it a "fallback" strategy obviously isn't enough to ensure this.) For that matter, we would need to determine exactly what are the appropriate contexts for use of the strategy; I'll claim that contexts involving sorites paradoxes are among them, but at this point I have little to say beyond that. These questions are very important, but will have to be left for future work.

5.4. Two-dimensional semantics

On the face of it, the two ways of making a categorization usable seem to be completely different. In one case, a categorization is usable because of an inherent property – being "natural" in the sense defined earlier (and hence salient). In the other case, a categorization is usable because of its role in the assertion, with the assertion serving to define the categorization. One might ask, though, if it would be possible to unify the two under one overarching principle of usability. I can't give a thorough answer to this, but some progress can be made in this direction if we think about the two cases from the perspective of two-dimensional semantics and diagonalization discussed in Stalnaker (1978, 2004).[7] On this view, a two-dimensional proposition, or propositional concept, is represented by a matrix like that in (26).

(26) A 2-dimensional proposition:

	w_1	w_2	w_3	…
w_1				
w_2				
w_3				
…				

The worlds on each axis are those in the context set. The matrix represents the two ways that facts determine whether an utterance is true: first, they help determine what

[7] Agustín Rayo suggested this.

proposition is expressed (after all, it's a contingent fact that words even have the meanings that they do), and second, they determine whether that proposition is true. The empty cells would contain truth values (0 or 1).

Now, let's consider a context that has a natural categorization. I'll use height and the predicate *tall* as an example. Imagine that we're looking at a group of six women, numbered x_1 to x_6. Suppose that x_1 and x_2 are both 5'0" in height, x_3 is 5'2", x_4 is 5'8", and both x_5 and x_6 are 5'10". In this context, the categorization of $\{x_1, x_2, x_3\}$ into "not tall" and $\{x_4, x_5, x_6\}$ into "tall" is natural, and is the only natural categorization. So this categorization would be usable in this context. But we have to ask what kinds of assertions could be made using this categorization. There would be little point, for example, in pointing to a particular woman, say x_4, and saying, "That woman is tall." The hearer presumably already knows that x_4 is tall according to the one available natural categorization. Suppose, however, that one of the women x_1 to x_6 is named Mary, and the speaker is trying to help the hearer identify which one she is. (This might be in Rayo's gameshow context, for example.) It's common ground that Mary is among these six women, and that the speaker knows which one she is but the hearer doesn't. Now suppose the speaker asserts, "Mary is tall." In this context this corresponds to the two-dimensional proposition illustrated in (27).

(27) The 2-dimensional proposition expressed by "Mary is tall"
Let w_1 = a world where Mary = x_1, w_2 = a world where Mary = x_2, etc.

	w_1	w_2	w_3	w_4	w_5	w_6
w_1	0	0	0	1	1	1
w_2	0	0	0	1	1	1
w_3	0	0	0	1	1	1
w_4	0	0	0	1	1	1
w_5	0	0	0	1	1	1
w_6	0	0	0	1	1	1

The proposition expressed is the same in each world in the context set – namely, the proposition that is true in worlds w_4, w_5, and w_6 and false in worlds w_1, w_2, and w_3. That means that the hearer can simply conclude that Mary is either x_4, x_5, or x_6. In other words, the assertion serves to remove worlds w_1, w_2, and w_3 from the context set.

The preceding context was one where the categorization being used was known (along with the order that it respected), and the identity of one of the individuals being categorized was at issue. Now consider a context where the identities of the individuals are known, and so is the order that the categorization needs to respect, but the cutoff point is in question. One case like this is the context, discussed above, where a book collection has been divided between two shelves according to year of publication, where the cutoff point was determined only by how many books could fit on each shelf. We'll call the books on the first shelf "early" and those on the second shelf "late," and this categorization respects the order of year of publication. Suppose that it's common ground that the book collection has been categorized this way, and that the speaker knows where the cutoff point is but the hearer doesn't. This means that all the worlds in the context set are the same in terms of what books are in the collection and what year each one was published; the only difference between worlds is in what categorization the words *early* and *late* refer to. To simplify things, let's suppose that there are only four books, a, b, c, and d, where a is the earliest, b was written later than a, c was written later than b, and d was written later than c, giving us the order a<b<c<d. Assuming that there must be at least one book in each category, there are three places the cutoff point could be: between a and b, between b and c, or between c and d. Now, suppose the speaker asserts, "Book b is early." This corresponds to the two-dimensional proposition in (28).

(28) The 2-dimensional proposition expressed by "Book b is early."
Let w_1 = a world where the categorization being used is { {a}, {b,c,d} }.
Let w_2 = a world where the categorization being used is { {a,b}, {c,d} }.
Let w_3 = a world where the categorization being used is { {a,b,c}, {d} }.

	w_1	w_2	w_3
w_1	0	0	0
w_2	1	1	1
w_3	1	1	1

In this case, the proposition expressed in each world is either necessarily false or necessarily true. Stalnaker (1978) argues that in this case the proposition is reinterpreted as its "diagonal," i.e., the proposition expressed in each world, as interpreted in that same world. This is the proposition obtained by taking only the cells in the matrix where the world on the horizontal axis is the same as the one of the vertical axis (shown shaded). By reinterpreting the assertion this way using diagonalization, the hearer can now use the speaker's assertion to conclude that the categorization in use is either the one where the cutoff point is between b and c, or the one where it's between c and d. In other words, world w_1 can be eliminated from the context set.

This suggests, then, that the two ways of making a categorization usable that I've discussed might correspond to two different ways of interpreting utterances that are generally available. Specifically, the case where there is a natural categorization is an example of an utterance which expresses the same proposition in every world in the context set. On the other hand, the case where a categorization is being defined corresponds to an utterance like "The Morning Star is the Evening Star," where the proposition expressed in each world is either necessarily true or necessarily false, depending on what are essentially semantic facts in that world. Therefore, while I can't exactly unify the two different ways of making a categorization usable, it might be possible to derive both from independently motivated principles of interpretation. However, this opens up a host of new questions about how principles of conversation affect the interpretation of vague predicates. For example, it's not immediately clear what Stalnaker's type of theory would predict about cases where the orders, cutoff points, and identity of individuals were all only partially known. There are also more specific questions, such as how (or whether) categorizations could be made intensional – in other words, how categorizations could be identified across worlds where the relevant orders, for example, might differ. The project of embedding a theory of vagueness using categorizations into a full-fledged theory of conversation is outside the scope of this paper.

5.5. Summary: The three strategies

In this section I've discussed three ways that a hearer can interpret an utterance containing a vague predicate. The preferred strategies are to either find a natural categorization or try to understand the utterance as serving to define a categorization. If these fail, then the hearer can fall back on the assumption that whatever categorization the speaker is using agrees with most natural categorizations that could be made in different contexts.

This completes my discussion of the semantic analysis of vague terms. In the next section I'll explain how this analysis characterizes vagueness and how it can give an answer to the sorites paradox.

6. The Characterization of Vagueness and the Sorites Paradox

Before I turn to the sorites paradox, let me clarify what I think vagueness consists in on my view. Here are the requirements for a vague term, as I see it: first, it must make reference to categorizations that follow some kind of order or orders. Second, this reference must be similar to pronouns and definite descriptions in being able to vary by context. (That is, a term wouldn't be vague if it uniquely defined a categorization for every context.) Third, any order used must be along a dimension that is continuous, in the sense that there can be pairs x and y such that $x<y$, and for any such pair it would generally be possible to imagine or construct a z such that $x<z<y$.[8] Finally, it must be possible in some contexts to find a natural categorization of the right kind.

Finally, I'll explain how my view provides an answer to the sorites paradox. Let's consider a version using the colors *red* and *orange*. To set up the paradox, we first imagine that we have a sequence of color patches – say, 100 of them – going from one that's clearly red to another that's clearly orange, where each patch is just barely less red (and more orange) than the previous one. The crucial "sorites" premise, then, is that if some patch is red, then so is the patch that comes after it (since adjacent patches are very similar in color). But this, together with the assumption that patch #1 is red, leads to the conclusion that all the patches are red, contradicting the assumption that patch #100 is orange. This is summarized in (29).

(29) Sorites paradox:
- (a) Premise 1: Patch #1 is red.
- (b) Premise 2: Patch #100 is orange.
- (c) Premise 3: For any n, if patch #n is red, then patch #n+1 is red as well.

- (d) Conclusion: Patch # 100 is red.

On my view, the words *red* and *orange* refer to categorizations according to color. Moreover, it seems reasonable that in any context that just involved color patches ranging from red to orange, both *red* and *orange* would be expected to use the same categorization, whether it was a categorization into two groups, with redder objects in one and more orange ones in the other, or a categorization into three or more groups with

[8] This requirement is probably too strong. For example, we arguably can't construct potential heaps of sand that differ by less than a whole grain. However, this problem of granularity applies to many diverse areas of natural language semantics, and I'll put it aside.

one or more "reddish orange" categories in the middle. The question, then, is what categorization *red* and *orange* refer to as used in (29). On my approach, this is really a question about discourse: if a speaker asserted, for example, (29.a), how could the hearer figure out which color categorization was being used? The immediate answer is that they couldn't, because in the context given there can be no natural categorization, and there is no sense that the speaker is using the assertion to define a categorization. Thus, in interpreting the first two premises in (29), we have to fall back on the strategy of assuming that the speaker is using a categorization that agrees in large part with natural categorizations that could be made in other contexts. Since by assumption patch #1 and patch #100 are very close to prototypical red and prototypical orange, respectively, it's clear that this is the case. That is, I claim that we accept the first two premises of the sorites paradox not because they're true per se, but because if we heard them in the context described, we would draw true conclusions from them.

Now for the sorites premise. Here, again, we are forced to fall back on the strategy of comparing natural categorizations that could be made in other contexts. And again it's true that for any particular pair of adjacent color patches, any imaginable natural categorization would put them in the same category. This just follows from the properties of natural categorizations and the way we've set up the context.

So what's the problem with the argument in (29)? The problem is that although it's true that all natural categorizations that satisfied the lexical constraints of *red* and *orange* would have to classify patch #1 as red, classify patch #100 as orange, and put adjacent patches in the same category, nothing ensures that such a categorization exists. In fact, given that it's impossible for a categorization to include all 100 patches and satisfy all three of these constraints, this simply proves that there can be no natural categorization of objects in a sorites series – a very unsurprising conclusion considering that the context was expressly set up so as to make a natural categorization impossible. If everything were made explicit, the sorites "paradox" would look like (30).

(30) The sorites argument revised:
- (a) Premise 1: All natural categorizations classify patch #1 as red.
- (b) Premise 2: All natural categorizations classify patch #100 as orange.
- (c) Premise 3: All natural categorizations put pairs of adjacent patches into the same category.

(d) Conclusion: All natural categorizations classify patch #100 as red.
(e) Real conclusion: There are no natural categorizations.

The three premises in (30) are only contradictory if we know that there are natural categorizations. Once we recognize the sorites series as really being an argument of this form, we should realize that the real conclusion is simply that there can be no natural categorization of a sorites series of color patches into "red" and "orange."

The obvious question, then, is why we don't easily recognize the sorites paradox as really having the form of (30). I think the reason is that the universal quantification is not actually part of the meanings of the premises in (29). In using the fallback strategy of considering what natural categorizations in other contexts would look like, we're still tacitly assuming that there is a natural categorization that the speaker is referring to in using, for example, the word *red*. Thus with each premise, we're using this strategy to draw a conclusion about that particular categorization, whatever it is: whatever it is, it must classify patch #1 as red; whatever it is, it must classify patch #100 as orange; and whatever it is, it must classify adjacent patches together. Thus we're left with the assumption that *red* in this context refers to some particular natural categorization that has these contradictory properties. Recall that the existence of a categorization to refer to is a presupposition of a word like *red*, so if we tried to simply drop the assumption that *red* refers to some particular categorization, then all of our premises involving *red* would suffer from presupposition failure. So on the one hand we have to assume that there is a categorization being referred to in order to understand the premises at all, and on the other hand any such categorization would constitute a contradiction. If you wanted to boil this down to a slogan, then, the sorites paradox doesn't come from a contradiction in the semantics of vague terms, but rather from a blind spot in our fallback strategies for interpreting sentences when we're under a certain amount of duress.

7. Conclusions

Under the analysis of vagueness that I've presented, vague terms don't denote properties in the normal sense. An assertion of the form "x is P" for a vague predicate P presupposes that there's a categorization of a certain kind that's being used, and asserts that this categorization puts x in a certain category. In a sense, then, this is a contextual theory: an individual is only red, or bald, or tall relative to a particular categorization. On the other hand, not every context uniquely determines a categorization. In particular, the kind of context that gives rise to a sorites paradox cannot in principle determine a unique categorization. The apparent paradox, on my view, actually reflects a kind of defectiveness in the context, and not any problem in the denotation of vague terms.

I'll conclude by making some brief comments about the implications of my account for the role of context in the interpretation of language. My proposed account of vagueness involves more context dependence than many alternative accounts. Specifically, my account makes all vague terms dependent on a contextually supplied categorization, although one is only available under certain conditions. Considering the great variety and number of vague expressions – it has even been claimed that almost all lexical items in natural language are vague – this apparently adds a substantial layer of context dependence to the interpretation of language. One might well ask, then, whether it's worth complicating our semantic theory with this additional layer. I think it is, and the main reason is that it isn't as great a complication as it may seem, because for a large number of vague expressions, my added source of context dependence takes the place of another source of context dependence that is otherwise needed. As I mentioned in Section 4, it's commonly assumed that scalar items such as *tall*, *big*, and so on are interpreted relative to a comparison class supplied by the context. For example, a sentence like *John is tall* might be true if the comparison class consists of American adults, but false if it consists of NBA basketball players. Under this view, *tall* is still vague apart from its dependence on a comparison class, because even given a comparison class such as American adults or NBA basketball players, it's not clear where to draw the line between "tall" and "not tall." Under my view, the contextually-supplied categorization takes the place of a comparison class, so that both kinds of properties of vague scalar items – those taken to be due to vagueness and those taken to be due to context dependence – are captured by the same mechanism. This means that the additional context dependence required by my account is limited to two parts: first, the contextually supplied object required for vague scalar terms under my account (a categorization, or set of sets of individuals) is more complex than the one required under the alternative view (a comparison class, or set of individuals); and second, this contextually supplied object is also required for non-scalar vague terms, assuming that these exist. That is, for one class of vague terms I'm just slightly complicating an existing context dependency, and for another class I'm adding context dependency in a way that unifies it with the first class. To the extent that my account of vagueness turns out to be correct, I believe this degree of complication is worth it.

8. References

Chierchia, Gennaro, and Sally McConnell-Ginet, 2000. *Meaning and Grammar: An Introduction to Semantics*, 2nd ed. [1st edition 1990.] Cambridge, Massachusetts: MIT Press.

Partee, Barbara H., Alice ter Meulen, and Robert E. Wall, 1993. *Mathematical Methods in Linguistics* [Corrected 1st Ed.] Studies in Linguistics and Philosophy, Vol. 30. Dordrecht, The Netherlands: Kluwer.

Rayo, Augustín, 2005. Vague Representations. Ms., UC San Diego / MIT. URL: http://web.mit.edu/arayo/www/vr.pdf

Schubert, L. K. and Pelletier, F. J., 1987. Problems in the Representation of the Logical Form of Generics, Plurals, and Mass Nouns. In E. LePore (ed.), *New Directions in Semantics*, London: Academic Press, pp. 385–451.

Stalnaker, Robert C., 1978. Assertion. In P. Cole (ed.), *Syntax and Semantics, Volume 9: Pragmatics*, New York: Academic Press, pp. 315-322. Reprinted in P. Portner and B. H. Partee (eds.), *Formal Semantics: The Essential Readings*, Malden, Massachusetts: Blackwell, pp. 147-161.

Stalnaker, Robert C., 2004. Assertion Revisited: On the Interpretation of Two-Dimensional Modal Semantics. *Philosophical Studies* 118: 299-322.