Vagueness Intuitions and the Mobility of Cognitive Sortals

Bert Baumgaertner

Forthcoming in *Minds and Machines*

Abstract One feature of vague predicates is that, as far as appearances go, they lack sharp application boundaries. I argue that we would not be able to locate boundaries *even if* vague predicates had sharp boundaries. I do so by developing an idealized cognitive model of a categorization faculty which has mobile and dynamic sortals ('classes', 'concepts' or 'categories') and formally prove that the degree of precision with which boundaries of such sortals can be located is inversely constrained by their flexibility. Given the literature, it is plausible that we are appropriately like the model. Hence, an inability to locate sharp boundaries is not necessarily because there are none; boundaries could be sharp and it is plausible that we would nevertheless be unable to locate them.

Keywords Vagueness · Vagueness Intuitions · Sortal · Cognitive Sortal · Mobility · Categorization · Concept Boundary · Sorites Paradox

Bert Baumgaertner
Department of Philosophy
1240 Social Sciences and Humanities
University of California, Davis
One Shields Avenue
Davis, CA 95616
Tel.: +1-530-475-2378
E-mail: bbaum@ucdavis.edu

1 Introduction

Predicates such as 'is bald', 'is a species', 'is tall', 'is red', 'is a planet', 'is a heap', 'is hard', 'is rich', etc., are paradigmatically *vague* in that they appear to lack sharp application boundaries. This is demonstrated in our inability to locate sharp cut-offs in fine-grained series, culminating in a sorites paradox. For example, consider a rich woman. Surely, she cannot become poor simply by losing one dollar, nor by losing any given dollar we might consider between her current worth and zero. However, if we let ourselves believe that no single micro-change can turn our rich woman poor, then we can repeat the process until we have a rich woman with zero dollars. Surely, that is absurd; a woman with no money is poor.

Numerous attempts to account for vagueness have been made. Many-valued logics, for example, are motivated by the intuition that there is no fact of the matter where the boundary of a vague predicate is, and so they introduce extra truth-values such as 'indeterminate'.[1] Similarly, degree theories, which introduce continuum many truth-values, proceed from the intuition that there are no boundaries of any sort.[2] The transition from one end to the other of the spectrum of what counts as red, for example, is a continuous gradient. Supervaluationism maintains that there is a sharp boundary under a given way of making a predicate precise (called a 'precisification'); vague concepts have many possible precisifications and vagueness is indeterminacy between which ones we should pick.[3]

What such accounts have in common is that they attempt to capture vagueness vis-a-vis a semantics or logic. That is, they attempt to formulate it in word-world relations. However, the source of vagueness is unlikely to be semantic in that sense, since it emerges from an inability to complete a kind of categorization task that needn't require subjects to have linguistic capacities. A chimpanzee, for example, is not expected to be any better at picking the first yellow (ripe) banana in a series that goes from green to yellow at a fine grain of scale. Moreover, even if we grant a chimpanzee non-linguistic representational capacities, failing to complete the categorization task may not be the consequence of some feature of the representation, but of how the representation is *processed* (an important point we return to later).

If vagueness is inherited from some other source, then we need to show what else could generate it. In this respect, ontological views and epistemic views of vagueness fare much better.[4] From the ontological view, if there is no fact of the matter where the edge of the cloud is, it is no surprise that neither we nor chimpanzees can locate it; if there are no boundaries in the world, then our representations have nothing to hook onto. Ontological views, however,

[1] See Halldén (1949), Körner (1960), and Tye (1994).

[2] See Goguen (1969) and Zadeh (1975).

[3] See Fine (1975) and Lewis (2001).

[4] For discussion of the ontological view, consider van Inwagen (1990), Morreau (2002) Smith (2005), Tye (1990), and Zemach (1991). For defences of epistemicism, see Sorensen (2001) and Williamson (1996).

must make sense of what it means for vagueness to exist in the world. This is a difficult task, since giving up the idea that there are boundaries in the world leaves us with a seemingly unpalatable incoherency; singular terms such as proper names, for example, would no longer be precise.[5]

Alternatively, epistemic views suggest that our inability to locate the sharp boundaries of vague predicates is grounded in the limitations of what can be known by limited cognitive agents such as us. Even epistemic views, however, are still unsatisfactory as an explanation for why vagueness exists. Vagueness appears to emerge from something more basic than a lack of knowledge. It seems that the problem is not just that I don't *know* where to draw the boundary in a sorites series but that I can't even maintain a *belief* or *opinion* about it. If I look at a colour spectrum from orange to red, I cannot get myself to draw a line such that I believe the line indicates the end of orange and the beginning of non-orange. This suggests that the problem of vagueness is grounded in something more fundamental than knowledge, i.e., vagueness does not merely involve the inability to form a justified belief of where a boundary is, it involves the inability to form a belief at all.[6]

There is, however, another possible source of vagueness. Let us suppose that, despite appearances and intuitions, our vague predicates are sharp. We now ask, is there any possible explanation for why we would fail to find pairwise cases where the relevant predicate applies to one but not the other? Consider the following. It is uncontentious that when we are presented with cases to classify (either by ostension or by invoking our imagination) the cognitive decision procedure requires the processing of information. Such information processing occurs over some period of time and possibly between the classification of different cases. So, it is possible that certain changes can occur after the classification of one case but before another, e.g., a sharp boundary might *move* in the interim. Given that finding a sharp boundary in a sorites series depends on us finding pairwise cases where the relevant predicate applies to one but not the other, it is thus possible that no such pairs can be found with the cognitive classification procedures we possess. This is because the sharp boundary may shift after the classification of the first member of the pair in such a way that the classification of the second member is in agreement with the classification of the first.

[5] Evans (1978) has probably the most well known argument against the view that objects can be vague, which is discussed again in Lewis (1988). Briefly, the argument is that if there are vague objects, then there could also be vague identities. A vague object like the Sahara desert, for example, would be indefinitely identical to its sharply bounded counterpart. It turns out however, that if objects a and b are indefinitely identical, that entails that the two objects are not identical. This is an odd result in and of itself, but further strengthing of the assumptions with a definitely operator leads to a flat-out contradiction. Defenses of the view exist (Tye, 1990), but has not gained widespread agreement. See Prinz (1998) for an overview.

[6] An epistemicist may insist that I can't maintain a belief because: (i) I can't maintain a *justified* belief, and (ii) good epistemic agents refuse to maintain unjustified beliefs. Then the point being made can be recapitulated in terms of (i). That is, we need an explanation for why I cannot maintain a justified belief, where justification is something that can be entangled internally with the psychological inability we are attempting to explain.

Notice that the above line of reasoning needn't make a commitment that vague predicates have sharp boundaries. Regardless of whether boundaries are sharp, our inability to locate them may be because of the mobility of the cognitive sortal (where 'sortal' stands in for 'class', 'concept', 'category', 'predicate', or 'representation').

Put succinctly, we name this line of thought as follows:

(Cognitive) Sortal Mobility: The mobility of cognitive sortals explains our inability to find sharp boundaries.

This paper argues that Cognitive Sortal Mobility (heretofore abbreviated as *Sortal Mobility*) is not just a mere possibility, but also a *plausibility*. A fast route to this claim is to point to the fact that we, even upon trying, do not (and maybe even cannot) find such pairwise cases. So, by inference to the best explanation, we are instantiations of systems that process information in the way just outlined. Such an argument, though, is admittedly too quick.

A more thorough argument is presented in this paper, which has three main threads, briefly summarized here. The first makes use of a synthetic methodology: we construct a (theoretical) model that is based on mental mechanisms which, quite plausibly, underlie our classification faculties (outlined in the paragraph below).[7] The second establishes that agents implementing the model are able to adapt their classification dispositions to incoming data streams, which, given our everyday changing environment, makes them better off than agents with static classifications. The third thread looks at the kind of data that cognitively mobile agents would produce and argues that humans produce such data as well, particularly in the context of the sorites paradox.

As far as the inability of locating supposedly sharp boundaries goes, this turns out to be a by-product. When an agent implementing the model considers certain cases in a sorites series, the relevant classes (and hence their boundaries) are updated. Considering certain cases can thereby bring about a kind of interaction effect: searching for a boundary can cause the boundary to move.[8] In fact, I will prove from the model, that the interaction effect renders an agent unable to locate a boundary beyond a limited degree of precision. Assuming that an agent must proceed through some indirect method of investigation to get information about their sortals (i.e., that neither the cognitive representation nor the updating processes are immediately transparent to an agent) and that the agent is unable to directly control the updating of the relevant sortal (i.e., an agent cannot voluntarily fix classificatory dispositions), then the agent is not in a position to find a boundary, even if there is one.

We thereby establish that Sortal Mobility is a plausible thesis about why we are unable to locate boundaries of vague predicates by: i) arguing for the plausibility that we are implementations of the kind of model constructed; and ii) showing that we would thereby not be able to locate boundaries beyond a degree of precision *even* if they were sharp.

[7] 'Synthetic methodology' is an allusion to the work of Braitenberg (1984).

[8] This is closely related to ideas in contextualism, discussed in section 4.

The rest of this paper proceeds as follows. In Section 2 I develop an idealized cognitive model with the kind of features described. I then prove, in Section 3, that if the model were implemented in an (artificial) agent, it could locate its sortals only up to a bounded degree of precision. The model is sufficiently general; systems of the same type will be confronted with the same limitations given a particular method of investigating their sortals. In Section 4, I further present the case that humans are plausibly like the model. This is done by (i) briefly looking at the literature, and (ii) pointing out that the model (with an additional assumption about generalizations) predicts *sorites susceptibility* – the fact that the inductive step in a sorites paradox is compelling when cases are considered at a fine enough level of grain.

2 Adaptive Categorization

We begin by developing an idealized model of a system that, if implemented, would exhibit classification behaviour akin to the cognitive systems of humans. We will see how a system (artificial or natural) with limited resources can significantly shift its classificatory dispositions. This is because of a system's ability to adapt to changes in its environment. (Where no philosophical scruples arise, we say that the system shifts its *concepts* or *categories*, instead of 'classificatory dispositions'. Such terminology allows a more natural way of talking, but is not required. What matters to our construction is that there are changes in classificatory dispositions.)

2.1 Economical Representations

To start, let us give our system a way of categorizing an infinite number of cases, including those that have not yet been encountered. Of course given finite resources, it is impossible (and even if it were possible, it would be grossly uneconomical) to represent an infinite collection of cases with an equinumerous collection of encodings. Apples, to take one example, can vary infinitely by their colour, shape, etc., so it is plainly impossible to pre-program the system with an explicit encoding for each case.[9] So what we seek is a way to represent an infinite number of cases by some finitary means.

One method of representing an infinite number of cases (either countable or uncountable) through finitary means has been presented by prototype and exemplar theories. The core idea is to take some cases (i.e. a prototype or an exemplar) and a rule that determines how much, and in what ways, new cases

[9] In fact, that would be to miss the purpose of representation altogether; the generality of a representation, i.e., that it can apply to numerous cases, is what makes it useful to an agent. We see this idea already back in Kant in his distinction between concepts (which are general) and intuitions (which are singular) (Kant, 1996, B377).

are allowed to vary from these paradigms.[10] In this way, a new case can be recognized even if that case has never been explicitly encoded in the system.

We can illustrate how such a system might look with a toy example. Suppose the task of our artificial system is to classify people that are deemed as MIDDLE AGED (MA). We could have the system represent the age range by explicitly encoding each of the years, so that MA = (35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55). Then, when presented with a case, this system checks that it satisfies MIDDLE AGED by looking to see whether it appears on the list. Alternatively, we can save resources by having the system put a few cases on a list and apply a rule to it.[11] For example, we could have $MA = (40, 47, 50) \pm 5$, where the list is designated by "()" and the rule operating on that list is here indicated by $\pm$. The rule '± 5' says to include those cases that are 5 more than the highest member on the list and 5 less than the lowest member on the list. (The rule ± 5 is obviously very simple and purposely so. It is meant to be illustrative only and not representative of the rules operating in more complex systems.) Admittedly, the age scale chosen, i.e. years, is discrete and fairly coarse grained. This means that the economical advantage of the second route is only slightly better. The advantage increases however if we change to a scale that is finer grained, e.g. days. This is because we would now require more cases to be on the list, i.e., we would have to put 48 years and 234 days on the list, 48 years and 235 days, etc. In the extreme case however, notice that, if the scale were dense (and hence 'infinitely fine grained'), then only the second route could finitely represent the infinitely varying cases.

It should be noted that our toy example involves a category that, at least in part, can be represented quantitatively. It is not obvious how we would encode representations for predicates such as 'is funny' without having corresponding quantitative metrics. This is an important but separate issue. We continue to develop our model using quantifiable metrics, leaving aside the task of applying the same principles to predicates that don't obviously have metrics of this sort. After all, the predicates of primary interest are those that are

[10] Of course both theories, either broadly construed or in their more detailed developments, differ on how exactly cases are encoded in the system. Prototype theories rely on a process of abstraction to generate a summary representation of statistically significant properties which are then used to determine the concept. In this way, the "prototype" need not correspond to specific cases. Exemplar theories do not rely on an abstract summary, but rather take exemplars to govern a concept. Exemplars are encoded property descriptions of a case or cases. The essential difference between exemplars and prototypes is that the former are encodings of many encountered cases where prototypes are encodings of parameters that characterize those cases. See Smith and Medin (1981) for a general overview and Smith and Medin (1999) for further discussion of the differences between prototypes and exemplar theories. See Malt (1989) who argues for both prototypes and exemplars and that subjects can use either in categorization tasks. See Barsalou (1990) who argues that we cannot empirically distinguish between exemplar and abstracted representations.

[11] Lists don't literally contain cases like Dorothy, Lassie the dog, or grandma. Talking about adding or taking cases off a list is shorthand for the representational operations in a system of encodings.

sorites susceptible, and a sorites series is usually presented with predicates that have explicitly quantifiable metrics.

2.2 Waste versus Recycle

Now suppose our system has gone through the process of classifying a new case. What might happen with the results of that processing? One option is that the system produces a single-serve representation only to then forget it (by not storing it or any traces of it). The alternative is that the system adds the new case (or information about it) to the ones it already encodes as explicit exemplars. To track these two options, let us consider two separate systems: one that trashes information after it uses it, the other that updates itself in reaction to the explicitly considered case. Let these be called *Waste* and *Recycle* respectively.

As a way of illustrating the difference between Waste and Recycle, we consider again the toy example (keeping in mind that they are oversimplified versions of much more complex systems like ourselves). Suppose both systems have the task of classifying people as MIDDLE AGED, and that each have the following hard-represented in them: $MA = (40, 47, 50) \pm 5$. We now put them to work. Each time Waste encounters a case in the interval [35,55], e.g. 46, it will classify that case as 'middle-aged'. Cnce that determination is made, and once it has had the appropriate causal effects, it then trashes the data just produced, so that further classifications are unaffected by it. Recycle, on the other hand, has the same classification process, but instead of deleting the data, it assimilates it by adding it to the list. The list is updated so as to reflect the fact that a concrete instance of 46 was classified as a positive case. Hence, after this change, $MA = (40, 46, 47, 50) \pm 5$. This will turn out to be advantageous.

Recycle's updating procedure will cause its behaviour to deviate over time. For example, given that Recycle's initial list and rules is $MA = (40, 47, 50) \pm 5$, when we give it cases like 42, 49, or 44, the list begins to expand, but the cases it would classify as MIDDLE AGED, namely those that fall in the interval [35, 55], stay the same. But the interval would change if we give it a case like 53; because it is a positive case that is nonetheless higher than any case explicitly on the list, Recycle updates, and then the rule '± 5' takes 53 as the highest and not 50, changing the interval of cases it would classify as MIDDLE AGED to [35, 58]. (For simplicity, nothing has been said about how cases are taken off the list. This can be represented by a competing list (such as ELDERLY) which, as it expands, takes cases away from MIDDLE AGED.)

Waste's dispositions are static in that it will always classify 55 as MIDDLE AGED, but never 56 since it falls outside [35,55]. Recycle on the other hand will differ because its dispositions will change as a result of cases it has previously classified. So where Waste's classificatory dispositions are independent of its past, Recycle's are dependent on it.

Recycle's capacity to change its dispositions gives it a significant advantage over Waste when operating in a changing environment. For if new cases keep continuing to be only *somewhat* like old cases, then Recycle will update to match that trend, allowing it to intelligently adapt.[12] This advantage is discussed in more detail in section 4.

In the next section, we regiment ideas about updating with an idealized cognitive model. We can then examine some relevant consequences with more rigour. The fundamental result will be that boundaries are not findable when searched for at a fine grain of scale.

2.3 Generalization and Adaptive Lists

Let us develop a model based on *adaptive lists*. An adaptive list is a list of encodings of cases (list of cases for shorthand) that can update, like Recycle's. Adaptive lists are intended to capture two primary ideas. The first is that a list governs what falls in an extension in the same way that a prototype or exemplar does – by encoding certain cases which then, together with a deviation rule, determine the membership conditions. The second is that adaptive lists change the system's classificatory dispositions by updating in response to incoming data. As such, the account is distinguished from other archetype-theories that posit unchanging archetypes or exemplars.

We use adaptive lists to help us generalize from our Recycle system. To begin, let us distinguish between the *adaptive list* and the *projection*. The adaptive list A contains those cases that are explicitly encoded in the system at a time (e.g. Recycle's list at a time for MIDDLE AGED contained 40, 47, 50). A projection, on the other hand, is the "projected" output of a rule that takes the explicit cases as input – it is all the admissible cases that the system would classify positively (e.g. all the cases in the range of $(40, 47, 50) \pm 5$).[13] Now, we need a way of capturing how information about new cases is assimilated, so let us introduce an update procedure. An update is a two-step process: First, certain cases of the projection are selected and a 'new' adaptive list A' is reconstructed. (How cases of the projection are selected can occur in a variety of ways, but none are essential to the model.) Second, a 'new' projection is recalculated from the adaptive list A', from which cases will again be selected, and so on.[14] In this way, updating an adaptive list can

[12] One might be inclined to object that such trend-matching could lead to complete expansion or contraction, which is neither advantageous, nor an accurate reflection of our categorization faculty. This objection is addressed in section 4.3.1.

[13] I talk as if an projection were a *projected entity* of a (mental) representation. However, a projection is just the set of cases that an agent is disposed to accept as being equivalent.

[14] It may turn out that some updates are redundant. In this case, the projection at t_n is extensionally identical to the projection at t_{n+1} because they include the same cases. An update is non-redundant when this fails to hold.

shift its projection, which is what allows the system to adapt to changes like those we have considered.[15]

The main idea can be understood as an instance of more pervasive phenomena: some concrete facts about a system determine its dispositions which, when triggered, alter the concrete facts about the system, which in turn alter its dispositions. Such cycling appears to be ubiquitous in nature and organized systems generally. Consider your eating cycle as an example. Your recent meals influence what you would be disposed to eat. What you are disposed to eat constrains (or even determines) what you do eat. What you do eat determines what your recent meals are. And what your recent meals are influence what you are disposed to eat. So, triggering different eating dispositions in the past can result in different eating dispositions in the future. The same is true of adaptive lists. The cases it is disposed to admit constrain the cases it does admit, which then influence the cases it is disposed to admit.

3 Sortal Mobility and Bounded Precision

In this section some basic limitations about how we can investigate our categorization faculties are briefly discussed. I then prove that when such limitations are reflected in the model, further limitations follow. The most important limitation we are interested in is that the location of a boundary cannot be found when the degree of adaptability (i.e., the amount that a projection shifts from updating the adaptive list) is greater than the grain of the scale on which the boundary is searched. Hence, *even if* boundaries were sharp, they would not be findable if our conceptual mechanisms were governed by an adaptive dispositional mechanism of the sort suggested in the previous section.

3.1 Learning by Testing

We want our model to be informative about us and how we investigate our mental states. An integral part of learning about one's own mental states is testing. Our dispositions to reason, act, or make decisions in a particular way are manifested in counterfactual circumstances, and so we imagine ourselves in such circumstances to acquire information.

This is also reflected in the way we access and acquire knowledge about the content of our concepts. Consider the post-Gettier investigation of our

[15] One might ask whether adaptive lists are sets. Adaptive lists persist through time, they are not "destroyed" when the encodings are changed. This means their identities are not determined extensionally, and consequently that they are not sets. What ontological commitments we have regarding adaptive lists, especially given that a projection can shift, is a question that deserves more exploration in its own right. The current objection however, is to represent the operational nature of a classification system. For that reason I leave the question about ontology to the side.

concept of knowledge.[16] We learn about the concept of knowledge by testing our dispositions to accept (hypothetical) examples as instances of it.[17]

So, let us make the same hold true of Recycle. That means that the way Recycle will learn about its own categorization faculty is by testing its dispositions to accept cases. Note that, although Recycle is initially ignorant of how it classifies, this does not presuppose that a boundary could not be located. We are simply stipulating that Recycle does not *start* by being omniscient about its inner states, and that it learns by doing tests on itself.

Now, the performance of such tests have an interesting interaction effect. Given Recycle's architecture, whenever it tests to see if it is disposed to accept a case, and that case is accepted, then the relevant adaptive list is updated to reflect that. Consequently, a test is an indication of some of Recycle's classificatory dispositions at the time of the testing, but the testing itself affects its classificatory dispositions. The result is that tests can affect the initial conditions of subsequent tests.

This is like a combination of the observer-effect (where the phenomenon being observed is affected by the very act of observing) and the primacy or recency effect (where earlier stimuli can affect responses to later stimuli). These classification tests have this interaction-effect because they require the use of the very faculty they are testing – the categorization faculty. And the categorization faculty is "in the dark" about whether it is being deployed for mere testing or genuine use – it functions just the same whether it is "in the wild" or in the comfort of an armchair. So such a test has two consequences: one is that it generates data and another is that it updates the system.

3.2 The Bounded Precision Theorem

We can now prove that certain systems like Recycle have interesting limitations given their architecture. The most significant result we show is that it is not possible to find a boundary (i.e., a pair of neighbouring cases such that they are given different classification results) when the level of search grain is equal to or finer than the level at which our model updates. To show this, we require a little more terminology, guided by our construction of Recycle and adaptive lists.

Case Testing: To test a case is to check whether it is a member of a list's projection. A test, O, is taken at a particular time or from a particular state, t, and takes cases, $a, b, c, d, \ldots$, as input. A test has two consequences. First, it gives either a positive result for membership, $\top$, or a negative result for non-membership, $\bot$. E.g. $O_t(a) = \bot$, which can be read as "Testing case a at time or state t yields a negative result". Second, a test can signal an update for the relevant tested list.

[16] See Sturgeon (1993) and Goldman (1988) for discussion.

[17] Interestingly, peoples' dispositions, and accordingly their intuitions, vary widely about which examples are and are not instances of knowledge (Goldman, 2003), which is evidence that people encode different membership conditions.

Updating: When a case is close to the boundary, the boundary of the relevant list is shifted by some non-zero magnitude. The amount of shift is determined by some multiple of a *grain*, which is the amount that a list can update given the relevant factors. Finer grained updates are shifts by smaller amounts than coarser grained updates. Every (non-redundant) update has a minimal unit of shift, which is equal to the grain.[18] For example, if the degree of grain went to the first decimal place, then a minimal unit of shift would be a tenth.

Finding a Boundary: To find a boundary on some relevant scale involves two steps: i) A case a is picked for a test O_t with a positive result, $\top$; ii) another case b for a subsequent test O_{t+1} is picked and has a negative result, $\bot$. The result of these two steps is an *interval* that approximates the location of the boundary.[19]

Finding boundaries can be done to varying degrees of precision, where precision is determined by the size of the interval. For example, the (upper) boundary of MIDDLE AGED on a scale of years would be found if the two-step process produces the interval [43,63]. Precision increases inversely with the magnitude of the interval, e.g. [44,61] is more precise than [43,63]. *Maximum precision* is achieved when the interval is equivalent to a minimal unit of grain of the scale in question. For example, on the scale of years as expressed with positive integers, [55,56] would be maximally precise. (Where the scale is dense, there is no maximal precision, since there is no smallest grain. Instead, there are arbitrarily high degrees of precision, since intervals can be arbitrarily small.)

We can now prove that whenever the grain of the scale on which a boundary is considered is equal to or finer than the amount it shifts from updating, the boundary cannot be found with maximal precision. To see this briefly and intuitively, suppose we give our model a pair of cases to test, n and $n+1$, where the distance between them is equal to or finer than a grain on a scale of updating. If n is tested first and it is a positive case near enough to a boundary, then a boundary shift is signaled in the direction of expansion. Consequently, the boundary will shift so that it includes both n and $n+1$ (since the distance of a shift is at least as large as the distance between n and $n+1$). Then $n+1$ will be given the same classification as n. The same holds if $n+1$ is tested first and is a negative case near enough to a boundary (where the boundary shifts in the direction of contraction). Hence, testing any pair of cases n and $n+1$ at a level of grain equal to or finer than the level of update will not reveal a boundary.

[18] Any time we speak of an update, we mean a non-redundant one.

[19] With the way we have designed the system, updating occurs when a test yields a positive result. To handle cases of updating where the first test yields a negative result, we consider the complementary concept for which it is a positive result and then proceed likewise. This is not an ad hoc amendment, for it is quite plausible that a concept and its complement are connected by a rule so that updating one automatically updates the other. We simply consider tests with positive results first for ease of proof.

More thoroughly, to claim that a boundary *can* be found with maximal precision on some scale is to have two conditions satisfied:

(C1) testing a pair of cases results in an interval, i.e. a space between which the boundary is found, and
(C2) the magnitude of the interval is equal to (or finer than) the grain of the scale on which the cases are tested.

Theorem 1 (Bounded Precision Theorem) *For any j grained updating and a k grained scale, if $j \geq k$, then the boundary cannot be found with maximal precision on the k grained scale.*

Proof We proceed by assuming that both C1 and C2 are met when $j \geq k$. C1 says we have tested a pair of cases with one of the following results:

(i) $O_t(a) = \top$, $O_{t+1}(b) = \bot$, or
(ii) $O_t(b) = \top$, $O_{t+1}(a) = \bot$

Consider (i). Since by assumption C2 is also met, then a and b must define an interval equal to (or smaller than) $1k$ grain of scale. That also means a is closest to the boundary (otherwise, we would have to suppose that there is another case 'wedged' in between, but then j would have to be finer than k, contrary to our assumption). If anything signals an update by being close enough to the boundary, the closest case does. Hence, $O_t(a)$ would have signaled an update at t. An update causes the boundary to shift by at least the magnitude of $1j$ grain unit of updating. Given that the boundary shifts away from accepted cases, b must be at least $1j$ unit of antecedent distance away from a given that $j \geq k$ and $O_{t+1}(b) = \bot$. Hence, after the shift, it is at least $1j+1k$ distance away. $1j$ unit + $1k$ unit is greater than $1k$ unit (whenever j is positive, which it is since we're dealing with absolute values). But then C2 is not satisfied, since the distance between a and b is greater than a single k grain unit, i.e., it is not maximally precise. Hence (i) is not a possible result given our assumptions. The proof goes likewise for result (ii). Since (i) and (ii) are exhaustive, satisfying C1 excludes the satisfaction of C2.

Example 1 Consider a slightly modifed system, Recycle*, where MIDDLE AGED is represented as $MA = (46, 48, 50) \pm 1$. Suppose that when Recycle* tests cases near the boundary (where conservatively, 'near' means at the very least 'touching the boundary') it updates by adding them to the list. Now suppose we gave Recycle* the task of finding a boundary for MIDDLE AGED on the scale of years as expressed in natural numbers. Then Recycle is able to successfully find a boundary, which happens to be between 51 and 52 before testing, with an interval that has a magnitude of at least 2. This is because, according to the supposed updating procedure, when 51 is tested the boundary shifts so that the highest member of the list is now 51. Consequently, MA comes to include 52. This means that the next available case that would give a negative test result is 53. So, either Recycle tests 52 for the second test, in which case it doesn't get the contrastive result, or it tests 53 or greater and gets the contrastive result, in which case the interval is not maximally precise.

Recycle's updating limits the level of precision with which a boundary can be found, which is the Bounded Precision Theorem.

In contrast to our inability to find boundaries on fine grains of scale, when we consider cases on coarse grains, it turns out that we *can* locate boundaries. For example, we can find the boundary for RED when we consider a series consisting of 8 crayons that go from red to green. In other words, an (in)ability to find boundaries is *grain sensitive*; finding a boundary requires a coarse-enough grain of scale. This turns out to be a trivial corollary of the Bounded Precision Theorem.

Corollary 2 (Grain Sensitivity) *If a boundary can be found with maximal precision for some k grained scale, then $j < k$, i.e. the j grain of the update procedure is finer than the k grained scale.*

Proof Contrapositive of Bounded Precision Theorem.

Example 2 If the grain of updating were somehow more fine grained than the scale on which it operates, e.g. all else from example 1 were the same but the minimal update was 0.5 years, then a boundary could be located because 56 would give a negative result (at least for the first pair of tests). Corollary 2 says that when a boundary is found with maximal precision, then the grain of the update rule is finer than the scale in question. That is in fact the case - a scale of half-years is finer than a grain on the scale of years. So boundaries can be located with maximal precision, but only when considered on coarse scales – we, for example, can locate the boundary for MIDDLE AGED when thinking in terms of decades (a scale with ten-year units), but not in minutes (a scale with units consisting of 0.000001903 years).

Corollary 3 *If a boundary can be located on all scales with non-zero units of grain, then the update rule is null, i.e. there is no shifting.*

Proof We proceed by reductio. Suppose a boundary can be located on all scales with non-zero units of grain and an update rule is not null. Then $i < k$ for any non-zero k (from corollary 2). But that means $i = 0$, since it is smaller than all possible non-zero values of k, which contradicts the assumption that the update grain is not null.

Example 3 Suppose we have some class that never updates, e.g., 1HOUR. Then we could select any arbitrarily fine grained scale, e.g. milliseconds, and still locate the boundary. This is a good result since mathematical concepts are presumably like this for humans. Such concepts, being perfectly stable, are perfectly precise.

3.3 Our Model and Sortal Mobility

Let us take stock. It has been suggested that a system with adaptive concepts recycles information whenever it classifies a case. The corresponding result

of this recycling process is that the boundary of the concept shifts, i.e. the dispositions of the system deploying the concept change. As we will see below, this is on the whole advantageous, since it allows a system to adapt to the incoming data. But it also means there are certain limitations pertaining to the investigation of the nature of an adaptive concept. For example, in the toy examples we have considered with Recycle and Waste, there are tasks that Recycle cannot complete that a static system like Waste can. If we ask Recycle, for example, to find the boundary for MIDDLE AGED with a precision equal to the scale in question, it will fail. But we also learned that this is not so if Recycle were only required to find the boundary on the scale of decades. This is because the minimum update on Recycle's scale of years is much smaller than the precision demanded by the task, so the shifting does not disrupt the search for maximal precision.

Sortal Mobility says that an inability to locate sharp boundaries is explained by the mobility of cognitive sortals (where for ease we have interchanged 'sortal', 'concept', and 'class'). On the model we have considered, we can assume that a boundary is sharp and show that nevertheless it can *not* be found beyond a degree of precision. So, an inability to locate sharp boundaries is not necessarily because there are none, it could be because finding them requires a degree of precision that is equal to or dwarfed by their degree of mobility. We have thus established that *if* we are like the model, then we should not expect to locate sharp boundaries (even if those boundaries are indeed sharp). In the next section we provide reasons for thinking that we *are* indeed like the model.

4 How Plausible Is It That We Are Like the Model?

We have developed Recycle, an idealized model on which we can suppose that boundaries are sharp but are nevertheless not findable beyond a degree of precision (where that degree is set by the mobility of the sortal). We now consider the question of whether it is *plausible* that we are implementations of Recycyle-like systems. The literature, which includes philosophy, artificial intelligence, and psychology, suggests an affirmative answer – we are such systems. In addition, we show that on the assumption that we are such systems (along with a methodological assumption about refuting generalizations), we get an explanation for why the sorites paradox is compelling. Hence Cognitive Sortal Mobility is a plausible thesis.

4.1 We Classify Dynamically

In philosophy, contextualism about vagueness is a good example of a view which holds that we classify dynamically (Keefe, 2000). A contextualist claims that the boundary of a vague predicate cannot be found because its content or extension *shifts* as the speaker considers cases in a sorites series (i.e., series of cases where each case is similar to its immediate neighbours but not

necessarily its extended neighbours). One reason for why the shift occurs is because the speaker or agent updates the relevant (implicit) *comparison class* (Ludlow, 1989). For example, when uttering the sentence, 'that elephant is heavy and that feather is too' the speaker updates the relevant comparison class throughout the utterance, making the content of it something akin to *that elephant is heavy for an elephant and that feather is heavy for a feather*. According to a contextualist, vague predicates do something similar when we apply them through a sorites series.

An updating comparison class is not the only contextualist explanation for why vague predicates shift. Some contextualists have suggested that vague predicates involve variation in elements of 'conversational score' (Lewis, 1979; Shapiro, 2006, 2003). Others have suggested that since vagueness can occur without a shift in context (where 'context' is externally construed), we also need an account of variation in psychological context (i.e., we should construe 'context' broadly so that it also includes mental states Raffman (1996, 2004)). So different accounts of variation are on the market. Nonetheless, all of them suggest that there is some kind of dynamic *shifting* that vague predicates undergo when we apply them to sorites series.

The claim that such shifting can and does occur is substantiated by demonstrations of it, both in philosophical and psychological domains. Swain et al. (2008) found that intuitions about how to classify cases of knowledge can vary according to whether, and which, other thought experiments were considered beforehand. This has also been documented more generally in psychology when people attempt to give definitions for their own categories (Barsalou, 1987, 1993). Moreover, people's classification behaviour exhibit *hysteresis effects*. Raffman (1994) for example, suggests that when a subject classifies a series of colour patches that go from red to orange, the place where the subject switches from calling the colour red to calling it orange is different when going in the other directio. This is thought to be because of 'a kind of judgmental inertia' which is created from the starting point (Raffman, 1994). This data suggests that we are systems whose classification dispositions not only *can* change, but that they *do* change in response to recent classifications. This was the very idea behind adaptive lists.

Yet, the shifting of a predicate, in and of itself, is not an explanation for why boundaries are not found. One needs to claim that a vague expression has the effect of changing its location *so that it is not where we look*.[20] In other words, it needs to be shown how the interaction effect that occurs by searching for a boundary is the very explanation for why a boundary is not found. The proof of the Bounded Precision Theorem does this. It demonstrates how interaction effects prevent the model from finding boundaries with maximal precision. To the extent that it is plausible that predicates do shift, it is plausible that we are more complicated versions of our idealized model, and consequently that the Bounded Precision Theorem holds of us as well. Moreover, although we

[20] Variations on this idea are discussed, though not necessarily endorsed, in Graff (2002); Kamp (1981); Raffman (1994, 1996); Stanley (2003); Ludlow (1989).

do not provide a proper treatment of inertia effects here, it is easy to see that they would arise in systems that implement our model.

What is novel to this account is that no appeal is made to *context* or *content* as traditionally construed in (formal) semantics.[21] In fact, we have largely ignored views regarding the semantic content of vague predicates, but with justification. In the introduction we gave an argument for thinking that an inability to locate sharp predicate boundaries is not closed off to linguistic competence – presumably, non-linguistic creatures can still sort the world and, to the extent that their classification dispositions can be investigated, it is unlikely that they will be found sharp. So, although vagueness occurs in language, and although we have appealed to particularly salient features of language to motivate the claim that we classify dynamically, it is unlikely that the (only) source of vagueness is in the semantic content of language.

In short, Sortal Mobility is a more general thesis that considers the underlying mental mechanisms of sorting. It claims that the apparent absence of sharp boundaries in classification can follow more straightforwardly from cognitive *processing* and *flexibility*. Our proofs of the Bounded Precision Theorem and its corollaries established this claim for an idealized model, but they also lead us to expect analogous limits on the relationship between sortal mobility and discriminatory precision in more realistic sorting systems. We needn't have a thesis of the semantic contents of vague predicates to provide the relevant explanation we are interested in.[22]

[21] The work of Raffman (1996, 2004) comes closest, since mental states are included as constituents of contexts. However, this account focuses solely on the inner workings.

[22] The reader may nevertheless want to know which semantic views are compatible with Sortal Mobility. Obviously, it is compatible with contextualist views of the kind mentioned. One might even extend our model, for example, to something like what appears in the work of Barker (2002). Roughly speaking, we could let the sortal be a gradable adjective in a shared discourse and let the updating rules of our model be determined by how uses of the term update the shared knowledge in discourse.

Generally speaking, however, Sortal Mobility is compatible with a view that vague predicates express properties whose extensions have exact boundaries, where the given property a predicate expresses varies across contexts. It is also compatible with views where the properties expressed do not vary across contexts, though one needs to be more subtle about what the constituents of a context are. If mental states are constituents of context, than clearly a change in mental state is a change in context, which is the preceding contextualist view. If, however, the properties expressed by vague predicates do not vary by context, then there seem to be at least two options. One is to give up that properties have crisp boundaries. Though compatible, this would unnecessarily double up explanatory work for our inability to locate sharp boundaries. The second option is to maintain that properties are crisp and then explain why appearances are to the contrary. This would need to allow for some separation between the property and its representation, which is perfectly compatible with Sortal Mobility.

Alternatively, one may hold that predicates do not express properties at all. Semantic content may be, for example, entirely 'in the head'. The details such a view would take us too far adrift. Suffice it to say that Sortal Mobility is relatively non-committal to views on semantic content, as it is a claim about our classification dispositions and not the meanings of terms.

4.2 Dynamic Classification is Advantageous

In section 2 it was suggested that Recycle's capacity to change its dispositions gives it a significant advantage over Waste when operating in a changing environment. This is because Recycle will update to match trends of changes, allowing it to intelligently adapt. This claim is motivated by the fact that a system like Recycle has already been realized in artificial intelligence.

A group of engineers attempted to build soccer-playing robots that could distinguish between the field, goal, ball, and other players.[23] There were several challenges that needed to be faced, particularly in colour classification (a primary method of distinguishing the listed features). One method of colour classification was to provide a predefined subdivision of colour space that was calibrated to the lighting conditions of an arena. However, algorithms that relied on such static colour classifications quickly ran into difficulties when the lighting changed (Heinemann et al., 2007). This would be an example of a system like Waste – one that doesn't update as it classifies cases, relying on an unchanging environment in which the same inputs should always map to the same judgments. In a changing environment, such systems behave poorly.

A Recycle-like algorithm, on the other hand, provided robots with the ability to adapt to changing lighting conditions.[24] The algorithm was named "The Automatic Color Training Algorithm", or the ACT algorithm for short (Heinemann et al., 2007). The basic task of ACT was to create a fast and automatically training (and retraining) look-up table that dynamically mapped colours in the environment (even as they changed) to different colour classes. ACT could be described in the following way. Given some arbitrarily chosen number of colour classes, the algorithm finds that incoming light inputs cluster because certain input frequencies (the colour of the field, the opposing players shirts, etc.) are more common than others.[25] The algorithm applies a deviation rule to each cluster, determining the acceptable values for each colour class. But in order to track changes in lighting, the algorithm continually recalculates the class to incorporate incoming colour values. So, if the incoming values gradually change over time (due to a fade in lighting, for example), the corresponding colour class continually updates to reflect that change. This is a system like Recycle where information about incoming cases is assimilated and used to update the classification mechanism; the ACT algorithm takes some cases and applies a rule (in this case a deviation rule) to determine the category, and importantly, it allows the category to adapt by updating the case list.

[23] See RoboCup (www.robocup.org). It is an international research initiative targeted to combine technology from artificial intelligence and robotics. The ultimate goal is to create soccer-playing humanoid robots that can eventually play at a competitive level with human players.

[24] Natural lighting in particular changes much more frequently than controlled indoor lighting.

[25] The way ACT does this is by calculating the mean value.

In a similar area of artificial intelligence, researchers have experimented on humanoid robots that played the grounded colour naming game (Bleys et al., 2009). The goal of the game is to have a population of artificial agents coordinate and develop a colour lexicon that is sufficiently shared to allow for successful communication. Bleys et al. (2009) found that robots which invent and coordinate their colour categories from scratch using their individual perceptions achieve the highest amount of communicative success. Such robots create colour categories using ideas from prototype theory; each colour category has a representative member and a standard one-nearest neighbour algorithm is used to classify a given object according to which representative it is closest to. Over the course of a series of games, prototypes shift relative to the success of the category in a game. Interestingly, the researchers found that the resulting colour ontologies of these robots reflect the environment in which they were developed, and moreover, these closely match those of the English language.

These examples are instances of our generalized construction of Recycle and adaptive lists. We too used the idea from prototype theory to represent an infinite number of cases by encoding a few paradigm cases and a rule. This way of modeling categorization has received a great deal of positive attention throughout the latter half of the 20th century.[26] One reason for its success is that it postulates cognitive systems that satisfy a general principle of category formation: ones that exhibit *cognitive economy*. The principle of cognitive economy states that an organism's categories ought to provide maximum information with the least cognitive effort (Rosch, 1999). Organisms that satisfy this principle better than others have an obvious advantage – they require fewer resources.

Prototype theory, broadly construed, is a plausible account of categorization behaviour.[27] We have seen that implementations of it, combined with ways of updating prototypes, have shown to be fruitful ways to design minds that are responsive to changing environments. Prototypes that update are taken to be a good thing. Hampton, for example, has said that, 'The boundaries [of our concepts] remain fluid for good reasons. When the world changes, or we discover new facts about it, our concepts can adapt to the change while their identity is still tracked.' (Hampton, 2007, p.377) Since we have evidence that our concepts update and we nevertheless seem to be able to track their identities, it is at least *plausible* then, that we are implementations of Recycle-like systems.

[26] One obvious reason why it received all the attention it did was because it overthrew the classical view that concepts or categories encode the necessary and sufficient conditions for their application. See Margolis and Laurence (1999) for further reasons why prototypical theories received widespread attention. See Rosch and Mervis (1975) for the influence of Wittgenstein in this regard.

[27] One difficulty of the view is that it is not compositional (Fodor and Lepore, 1996), which has been disputed (Hampton and Jönsson, 2009; Hampton, 2007). We can safely ignore this debate, since we are not concerned with the thesis that *concepts*, the constituents of compositional thoughts, are prototypes. Even if that thesis turns out to be false, prototype theory stands as a thesis of categorization behaviour.

4.3 Why the Sorites Paradox is Compelling

A vague concept, like the ones listed in the introduction, can be used to run a sorites paradox. First, we select a member of some concept as a *base case* (typically an exemplar of the concept). Then, we proceed in step-wise fashion, using an inclusion principle that says small differences can be neglected, to derive an unacceptable conclusion (this is called *the inductive step*). Here is an example of a sorites paradox using the concept MIDDLE AGED.

P1 45 years of age is middle aged (Base Case)
P2 If n years of age is middle aged, then $n + 1$ years of age is middle aged (Inductive Step)
C 100 years of age is middle aged

Presumably, the conclusion of this sorites is false, so one of the premises must be denied. Which one?

P1 looks obviously true – if there is any age that is middle aged, surely 45 is a clear example. Nonetheless, suppose we deny P1, which entails that 45 years of age is *not* middle aged. This consequence is just as bad as (if not worse than) the conclusion, which says that 100 years of age *is* middle aged. So the denial of P1 leads to a consequence that is as bad or worse than the conclusion which forced us to deny a premise. On the presumption that the conclusion is unpalatable, we should not be left with something even more difficult to swallow.

The other option is to deny P2. P2 certainly has intuitive appeal; it seems that for any arbitrary case n, if n is a member of the projection of a sorites concept, then its closest negligibly different neighbour, $n+1$, is also a member. This is particularly obvious when we change our measure of middle age from the scale of *years* to scales like: *months, weeks, days, hours, minutes, seconds,* etc. – surely (it seems) one second can't make the difference to being middle aged! So, if the inclusion principle uses some scale that fails to capture the intuition that a small difference can be neglected (as might be the case for *decades*), a more fine-grained scale can be selected to then run the sorites. So note that the more fine-grained a scale is, the more acceptable we find the inductive step to be. However, despite its intuitive appeal, the simplest solution to the paradox, given the strength of P1, is to deny P2.

Consider how Recycle would respond to the sorites paradox. When it checks P1, it turns out true. If we ask it to find a counterexample for P2, it will fail. It could never find a counterexample because whenever it tested for some case n, its classificatory dispositions would update accordingly. When the grain of updating matches or exceeds the grain of testing, then $n + 1$ would already be in the projection by the next test. So since Recycle implements adaptive lists and these can update whenever a case near the boundary is checked, we have an explanation as to why no counterexample can be found to P2. To find a counterexample would be to violate the Bounded Precision Theorem. Note that the sense of "could not" is not a metaphysical impossibility, but one based

on contingent facts of Recycle's design. A design we motivated from evidence of our own.

The proposed explanation for why it *seems* impossible for us to deny the inductive step is based on two facts. One, in order to deny a generalization, we tend to require one or more counterexamples to the claim. Second, our attempts to find counterexamples would be thwarted if we were Recycle-like systems, since finding a counterexample amounts to finding a boundary to maximal precision (which goes against the Bounded Precision Theorem). The strength of this inference to the best explanation is heightened by the mutually supporting claims that we are Recycle-like systems and the explanation for why the sorties paradox is compelling.

One might suspect that Recycle's sorites susceptibility may not be comparable to ours. After all, we don't seem to need to test cases to realize that a counterexample to the sorites won't (or can't) be found. Moreover, once we've recognized a vague concept, whether in the form of the sorites paradox or not, we don't seem to need to test cases to know that some other vague concept will be sorites susceptible.

I agree that humans have a high capacity for generalizing from relatively few samples, a feature we have left out of the model entirely. Such a capacity allows us to be good at detecting which concepts are going to be sorites susceptible without looking at their particular cases, and it also means that given a particular concept, we don't need to check *every* case to know that we are unable to find a counterexample. But how could we make a generalization from *no* sample data at all? We would have nothing to generalize from.

So the story of our own sorites susceptibility needs two parts. One is that we need at least a few samples from which we can generalize that no sharp boundaries can be found. (This generalization can then cover concepts that we recognize as being relevantly similar.) The second is that the generalization is never challenged by counterexamples. So, a lack of finding counterexamples provides both the germ and maintenance of the story. It is this key fact that we have sought to explain with our model.

Moreover, the kind of sorites susceptibility we have captured in our model is, as in humans, grain sensitive. When we consider subtracting large increments of money, such as tens of thousands of dollars at a time, then it is easier for us to hold a belief which subtraction makes our woman poor. However, when we consider smaller increments, such as dollars or cents, then it becomes much less intuitive for us to say that there is such a particular amount. The fact that this very human feature of the sorites paradox has an analogue that naturally falls out of the idealized model not only makes it plausible that we are appropriately like it, but that its sorites susceptibility is comparable to ours.

4.3.1 Objection: Forced Marches

One might object that a our model of a sorites susceptible concept can or will reduce it to triviality or some sort of incoherence. The reasoning goes

as follows. If no counterexample can be found, then there is no way to halt the complete expansion (or contraction) of a class by a *forced march*, i.e., an iteration of tests that incrementally move the boundary until all possible cases are inside (or outside) the class. Since, by the Bounded Precision Theorem, no counterexample can be found, there is no way to halt a forced march.

Consider the contrapositive of the main premise in this objection: if a forced march can be halted, then a counterexample can be found for the inductive step. This conditional is plausibly false in the case of humans. If I am marched through a series of cases for middle aged and reach some number like 68, it is natural for me to stop the march and say that something has gone wrong - 68 is not middle aged. Since I denied 68 but included 67, you would be right to ask me whether these serve as the counterexample to the inductive step. But it is perfectly coherent for me to retract my inclusion of 67 and say that I must have gone wrong *somewhere* in the march, but I *don't know where*. Hence a forced march can be halted without producing a counterexample; the objection doesn't hold for humans.

As it stands, forced marches may be possible in principle in our idealized model, but there are some contingent features that make it highly implausible to occur. One is that Recycle's algorithm only allows a concept to become all-inclusive if one assumes that cases remain on a list eternally while it runs through a very large number of precisely ordered iterations of tests. Nothing has been said about how long the life cycle of a list member is (or how many iterations of the algorithm it would require to expand to some size). It is possible, for example, that membership expires without reinforcement, i.e., some cases may be forgotten about in time or might need to be dropped to include other cases. So, if members leave the list with appropriate rapidity, all-inclusion will rarely or never occur.

Furthermore, other concepts might compete for the same members, infinitely frustrating any winner-takes-all situation. For example, adaptive lists such as YOUNG and OLD might be tied to MIDDLE AGED via a rule that says 'no case may be on more than one of these lists at a time'. This rule forbids YOUNG, MIDDLE AGED, and OLD from simultaneously overlapping. In this way the expansion of a concept is limited by the activity of competing concepts (c.f. a suggestion by Williamson (1996): one way to stop a sorites paradox is for it to collide with another sorites paradox in the opposite direction (p.87)). In fact, if such a rule is combined with a method of testing pairs of cases via random sampling, then it becomes astronomically unlikely that just the right order of pairs of cases is selected that force all-inclusion (since pairs from the other direction would push the boundary back). The point is that in a more complex system of concepts, some stability can be reached and maintained by the concepts exerting tension on one another.[28]

[28] One might not find this response entirely compelling. If concepts compete for members, the result might be that we end up with (briefly) immobile boundaries, and a forced march would reveal them. For example, let us consider a scenario as above, where we march our model's MIDDLE AGE class from 45 up to 68, at which point its OLD AGE class kicks in and forces it to say that 68 is not middle aged (although it has just said 'yes' that 67 is middle

In short, we concede that the simplified model is vulnerable to forced marches. The response, however, is that this worry dissolves when the idealization is relaxed and we consider more realistic sorting systems. And more importantly, contrary to the forced march worry, we expect that the Bounded Precision Theorem and its corollaries will hold as we move from the idealization to real sorting systems.

For example, it is plausible that our categorization faculties make use of numerous cognitive and metacognitive processes when we categorize (some of which we may be aware of, others not). A metacognitive process may pick up on a particular trend in the cases we are attending to and make predictions about what cases may or may not lie ahead. Taking such 'sneak peaks' at cases in the distance could spoil the precise ordering of positive cases that are needed to do a forced march. Alternatively, some processes may be equipped with thresholds that require a certain number of tests before firing an answer. Such repeated testing could inadvertently update related concepts, which could feedback on the concept under investigation. The point is that these more sophisticated processes do not help in removing interaction effects, which are at the heart of our results. Rather, they increase their complexity.

5 Conclusion

Cognitive Sortal Mobility is the claim that the mobility of cognitive sortals explains our inability to find sharp boundaries. We argued for the plausibility of this thesis by constructing an idealized model that can have very real, yet very mobile, sharp boundaries. Nevertheless, we proved that such boundaries are not findable beyond some degree of precision.

Throughout our construction, we focused on a simplified implementation of the model, which we called Recycle, and compared it to the alternative, Waste. We noted that Recycle's class was adaptable, which gave it an advantage over Waste's static class. Furthermore, our construction of Recycle also yielded an innocent side-effect: sorites susceptibilty. We also looked at evidence which suggests that we are systems appropriately like the idealized model. The idealized model reflected cognitive features that have been documented in the literature. Moreover, we used it to show how the sorites paradox can be compelling despite the existence of sharp boundaries. It is plausible then, that humans are implementations of the model. So *even if* our sortal boundaries were sharp, we would nevertheless be unable to find them.

Acknowledgements Thanks goes first and foremost to Bernard Molyneux for his extensive feedback on this work, from its conception to the final version. Thanks also to Aldo Antonelli, Adam Sennet, Paul Teller, and an anonymous referee for invaluable comments on drafts.

aged). Has the forced march thereby exposed a boundary? If it has, it is very short lived. By having considered 68 as old aged, the OLD AGE class may expand its boundary as well, requiring the MIDDLE AGE class to retract. So if we were to test the model's MIDDLE AGE class against 67 again, the model would respond in the negative.

For helpful discussions at various points in the development of this paper I am indebted to Lawrence Barsalou, Jeff Schank, Miguel Sebastian, and Matthew Stone.

References

Barker, C. (2002). The dynamics of vagueness. *Linguistics and Philosophy 25*(1), 1–36.

Barsalou, L. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization*, pp. 101–140. Cambridge: Cambridge University Press.

Barsalou, L. (1990). On the indistinguishability of exemplar memory and abstraction in category representation. In *Content and process specificity in the effects of prior experiences*, Volume III, pp. 61–88. Lawrence Erlbaum.

Barsalou, L. (1992). Frames, concepts, and conceptual fields. In A. Lehrer and E. F. Kittay (Eds.), *Frames, Fields, and Contrasts*, pp. 21–74. Hillsdale, NJ: Lawrence Erlbaum Associates.

Barsalou, L. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. *Theories of memory*, 29–101.

Bleys, J., M. Loetzsch, M. Spranger, and L. Steels (2009). The grounded colour naming game. *Proceedings of Roman-09*.

Braitenberg, V. (1984). *Vehicles, experiments in synthetic psychology.* Bradford Book.

Evans, G. (1978). Can there be vague objects? *Analysis 38*(4), 208.

Fine, K. (1975). Vagueness, truth and logic. *Synthese 30*(3), 265–300.

Fodor, J. and E. Lepore (1996). The pet fish and the red herring: Why concepts arent prototypes. *Cognition 58*, 243–76.

Goguen, J. (1969). The logic of inexact concepts. *Synthese 19*, 325–373.

Goldman, A. (1988). *Epistemology and Cognition.* Harvard University Press.

Goldman, A. (2003). Epistemology and the evidential status of introspective reports. In A. Jack and A. Roepstorff (Eds.), *Trusting the Subject?: The Use of Introspective Evidence in Cognitive Science*, Volume 2, pp. 1–16. Imprint Academic.

Graff, D. (2002). Shifting sands: An interest-relative theory of vagueness. *Philosophical Topics 28*(1), 45–82.

Halldén, S. (1949). *The Logic of Nonsense.* Uppsala: Uppsala Universitets Arsskrift.

Hampton, J. (2007). Typicality, graded membership, and vagueness. *Cognitive Science: A Multidisciplinary Journal 31*(3), 355–384.

Hampton, J. and M. L. Jönsson (2009). Typicality and compositionality: The logic of combining vague concepts. In W. H. E. Machery and M. Werning (Eds.), *Handbook on Compositionality.* Oxford: Oxford University Press.

Heinemann, P., F. Sehnke, F. Streichert, and A. Zell (2007). Towards a calibration-free robot: The ACT algorithm for automatic online color train-

ing. In *RoboCup 2006: Robot Soccer World Cup X*, Volume 4434 of *Lecture Notes in Computer Science*, pp. 363–370. Springer Berlin/Heidelberg.

Kamp, H. (1981). The paradox of the heap. In U. Mönnich (Ed.), *Aspects of Philosophical Logic*. Dordrecht Reidel.

Kant, I. (1996). *The critique of pure reason (Werner S. Pluhar, Trans.)*. Hackett Publishing Company.

Keefe, R. (2000). *Theories of Vagueness*. Cambridge University Press.

Kennedy, C. (1999). Projecting the adjective: The syntax and semantics of gradability and comparison. *The number sense 4*(4), 11.

Körner, S. (1960). *The Philosophy of Mathematics*. London: Hutchinson.

Lewis, D. (1979). Scorekeeping in a language game. *Journal of philosophical logic 8*(1), 339–359.

Lewis, D. (1988). Vague identity: Evans misunderstood. *Analysis 48*(3), 128.

Lewis, D. (2001). *On the plurality of worlds*. Wiley-Blackwell.

Ludlow, P. (1989). Implicit comparison classes. *Linguistics and Philosophy 12*(4), 519–533.

Malt, B. (1989). An on-line investigation of prototype and exemplar strategies in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition 15*(4), 539–555.

Margolis, E. and S. Laurence (1999). *Concepts: Core Readings*. MIT Press.

Morreau, M. (2002). What vague objects are like. *Journal of Philosophy 99*, 333–61.

Prinz, J. (1998). Vagueness, language, and ontology. *Electronic Journal of Analytical Philosophy 6*.

Raffman, D. (1994). Vagueness without paradox. *The Philosophical review 103*(1), 41–74.

Raffman, D. (1996). Vagueness and context-relativity. *Philosophical Studies 81*, 175–192.

Raffman, D. (2004). Borderline cases and bivalence. *The Philosophical Review 103*, 41–74.

Rosch, E. (1999). Principles of categorization. In E. Margolis and S. Laurence (Eds.), *Concepts: Core Readings*, pp. 189. MIT Press.

Rosch, E. and C. Mervis (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology 7*, 573–605.

Sainsbury, R. M. (1990). Concepts without boundaries. In R. Keefe and P. Smith (Eds.), *Vagueness: A Reader*. MIT Press.

Shapiro, S. (2003). Vagueness and conversation. In Beall (Ed.), *Liars and Heaps: New Essays on Paradox*, pp. 39–72.

Shapiro, S. (2006). *Vagueness in context*. Oxford University Press, USA.

Smith, E. and D. Medin (1981). *Categories and concepts*. Harvard University Press Cambridge, Mass.

Smith, E. and D. Medin (1999). The exemplar view. In *Concepts: Core Readings*, pp. 207–222. Bradford Book.

Smith, N. (2005). A plea for things that are not quite all there: Or, is there a problem about vague composition and vague existence? *Journal of Philosophy 102*, 381–421.

Sorensen, R. (2001). *Vagueness and Contradiction*. New York: Oxford University Press.

Sorensen, R. (2005). Precis of vagueness and contradiction*. *Philosophy and Phenomenological Research 71*(3), 678–685.

Stanley, J. (2003). Context, interest-relativity, and the sorites. *Analysis 63*, 269–80.

Sturgeon, S. (1993). The gettier problem. *Analysis 53*(3), 156–164.

Swain, S., J. Alexander, and J. Weinberg (2008). The instability of philosophical intuitions: Running hot and cold on truetemp. *Philosophy and Phenomenological Research 76*(1), 138–155.

Tye, M. (1990). Vague objects. *Mind 99*, 535–57.

Tye, M. (1994). Sorites paradoxes and the semantics of vagueness. In J. Tomberlin (Ed.), *Philosophical Perspectives: Logic and Language*. Atascadero, California: Ridgeview.

van Inwagen, P. (1990). *Material Beings*. Ithaca, New York: Cornell University Press.

Williamson, T. (1996). *Vagueness*. Routledge.

Williamson, T. (1997). Precis of vagueness. *Philosophy and Phenomenological Research 57*(4), 921–928.

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

Zemach, E. (1991). Vague objects. *Nous 25*, 323–40.