VAGUE JUDGMENT: A PROBABILISTIC ACCOUNT

PAUL EGRÉ

ABSTRACT. This paper explores the idea that vague predicates like "tall", "loud" or "expensive" are applied based on a process of analog magnitude representation, whereby magnitudes are represented with noise. I present a probabilistic account of vague judgment, partly inspired by early remarks from E. Borel on vagueness, and use it to model judgments about borderline cases. The model involves two main components: probabilistic magnitude representation on the one hand, and a notion of subjective criterion. The framework is used to represent judgments of the form "x is clearly tall" vs. "x is tall", as involving a shift of one's criterion, and then to derive observed patterns of acceptance for sentences of the form "x is tall and not tall" / "x is neither tall nor not tall".

1. INTRODUCTION

One way to approach the difference between vague predicates and precise predicates is in relation to distinct measurement methods and scales (see [Fults, 2011]; [van Rooij, 2011b], [van Rooij, 2011a]; [Solt, 2011b], [Solt, 2015]). If, following Stevens, we define measurement as the assignment of numbers to objects according to a rule, we can think of a precise scale as one for which the rule would assign a determinate number to an object (be the scale ordinal, interval or ratio, to consider only those three kinds). Consider a precise predicate like "180cm tall". We predicate it of an object by an operation of so-called extensive measurement, whereby we concatenate the same unit of measurement until we match the height of the individual. In principle – assuming the object's height to stay constant, and the units to be perfectly calibrated–, such a method is deterministic, in that repetition of the method would assign the same object no two different values.

Our mental representation of magnitudes is not as tidy, however. When representing attributes such as height, brightness, loudness, expensiveness, and so on, we appear to rely on a distinct mechanism of analog representation, whereby magnitudes are represented with some noise. We may call digital a measurement method that assigns to every object a unique position on a discrete scale. By contrast, we may call analog a measurement method that assigns an object possibly distinct positions on the scale, though in a way that globally covaries with represented magnitudes.[1] An example of analog measurement is given by the approximate number system, by which we estimate numerosities. A key feature of this system is that cardinalities are represented with some probabilistic error, meaning that the same number of objects will be represented as more or less than it is (see [Feigenson et al., 2004], and [Fults, 2011] for discussion), even though larger cardinalities will on average be estimated as larger than smaller cardinalities.

The idea I want to explore in this paper is that our use of typical vague predicates like "tall", "loud" or "expensive" is to be thought of in relation to such a mechanism of analog magnitude representation. More specifically, I shall investigate a probabilistic model of vague judgment, essentially based on the idea that we mentally represent magnitudes with some approximation. The model I will put forward is itself inspired from early and semi-formal remarks given by Emile Borel concerning the sorites paradox [Borel, 1907]. A brief sketch of the model appears in a recent paper presenting and discussing Borel's original essay (see [Egré and Barberousse, 2014]), but the model deserves further elaboration and discussion. Borel's idea, in a nutshell, is that we should think of vague predicates in relation to a mechanism of approximate measurement, and in relation to the setting of some implicit boundary value, which I propose to call a criterion (following common usage in psychology). On that approach, judgments involving vague predicates involve two sources of individual variation, one that concerns the approximation with which magnitudes are mentally represented, and the other that concerns the positioning of the criterion.

My main goal in this paper is to discuss the interplay between those two sources of variability, and then to extend the model to an account of judgments concerning borderline cases. In the next section, I first present the model more formally, and show how it allows us to derive psychometric curves of the sort familiar in categorization tasks. In section 3, I exploit an analogy between the Borelian model of vague categorization and the model of signal detection to discuss the main commitments of my approach. I argue, in particular, for a distinction between the notions of criterion and standard, and against the suspicion that the notion of a criterion constitutes a commitment to epistemicism. In section 4, I discuss several measure-theoretic assumptions behind the mechanism of magnitude representation postulated by the model. In section 5, finally, I show how the model can account

---

[1]A more common characterization of the analog-digital distinction is in terms of a continuous vs. a discrete system of representations. The model that follows, however, is not committed to the assumption that analog representation is necessarily a mapping to a continuum. In that, the present definition agrees with [Maley, 2011], according to whom "analog representation is representation in which the represented quantity covaries with the representational medium, regardless of whether the representational medium is continuous or discrete." Unlike Maley, I do not assume the covariation is necessarily given by a linear function.

for judgments about borderline cases. I show, in particular, how judgments of clarity and unclarity can be derived from the assumption of a shift of one's criterion. Borderline cases are standardly characterized as cases that are neither clearly P nor clearly not P. The model can fit the pattern observed experimentally for such judgments. Moreover, it can be used to explain the higher assertability of sentences of the form "x is P and not P" (or "x is neither P nor not P") in borderline cases relative to clear cases, and the lower assertability of each conjunct (see [Ripley, 2011], [Alxatib and Pelletier, 2011], [Egré et al., 2013]), provided such judgments are referred to two criteria instead of one. In so doing, the model parallels the strict-tolerant account for vague predicates ([Cobreros et al., 2012]), based on the assumption that the same predicate can be applied relative to a stricter or a laxer criterion.

2. Vague judgment: a Borelian model

One of the first probabilistic accounts of vague categorization was proposed in 1907 by Émile Borel in relation to situations in which some approximate representation of an exact quantity needs to be made ([Borel, 1907]). Borel's insight, generally speaking, is that we should think of vague categorization as a process of approximate measurement. In this section I first present the account and illustrate its workings on a simple example. The present account elaborates on a model of vague judgment put forward in earlier work ([Egré and Barberousse, 2014]), inspired from Borel's mostly informal remarks. Borel did not give a worked out theory, however, nor did he give a proper theory of lexical vagueness. Here I shall therefore leave historical details behind (referring to [Egré and Barberousse, 2014] for those), in order to focus on new aspects of the account sketched in previous work.

The leading idea of this paper is that the application of a vague predicate, such as "tall", or "expensive", implies the mapping of a stimulus onto an inner scale of magnitude representation, and the comparison to some inner criterion. The mapping can be thought of as a process of noisy transduction. Let us consider the adjective "tall". In order to judge whether someone is tall or not (given a context and a comparison class), we need to build a representation of that person's height, and then we need to compare that representation to a distinguished value. The model I am going to discuss postulates essentially three parameters to describe that process. First, we need a scale of measurement for height, setting a basic *measurement unit* to serve as the input of one's representation. Secondly, we need an indication of the *approximation* with which each such unit if subjectively represented. Thirdly, we need what I shall call a *criterion*, that is a threshold value, explicit or implicit, such that anyone beyond that value would in principle count as tall, and anyone beneath would in principle count as not tall.

Two kinds of cases must be distinguished. In one class of situations, I have no numerical information about the person's height: I can see John, for example, and need to decide whether he is tall, given the height that I perceive. In another class of situations, I have numerical information about the person's height: I am told, for example, that John is 178cm tall, and need to decide whether he is tall. One of the main contentions of this

paper is that there is no essential difference between the two kinds of cases. In both cases, I am faced with the task of mapping the stimulus magnitude (whether perceived, or numerically given) to my inner scale, in order to compare it with some inner reference point. In both cases, the mapping is likely to introduce some distortion.

Before justifying that assumption (see Section 4.1), let me illustrate the model on an example. Suppose John is actually $h = 178$ cm tall. I either see him without knowing his exact height, or his height is communicated to me in numerical terms, for example in centimeters. What the model says is that, in either case, John's actual height h is tranduced into some perceived height $f(h)$ on my inner sensory scale. That perceived height $f(h)$ is then compared to an inner threshold β. Let us assume $\beta = 185$ cm on my inner scale for "tall". What is the probability of judging John to be tall? In general, this probability is equal to:

$$Pr(f(h) \geq \beta)$$

What needs to be specified is the form of the function f from actual height to perceived height. What the current model assumes is that each relevant unit of John's actual height is represented with some noise, and that the noisy representation of John's total height is given by the addition of noisy representations of each relevant unit of his height (that assumption will be discussed in Section 4). That is, for each relevant unit value i of John's height, we can associate a random variable X_i that gives the probability of representing each unit i of John's height as more or less than i. For example, if the relevant unit is 1cm and John's height is $h = 178$ cm, then $f(h) = \sum_{i=1}^{i=h} X_i$. The probability of judging John tall if my criterion is $\beta = 185$ cm is therefore $Pr(\sum_{i=1}^{i=178} X_i \geq 185)$. Knowing the expectation and variance of the random variable X_i, we can use the Central Limit Theorem to compute that probability with a good approximation when i is sufficiently large (see [Egré and Barberousse, 2014] for details). In cases where i is small, an exact value can be computed directly.

Let us state this in full generality and then illustrate it. Given a vague predicate P, some object x of a given magnitude $h(x) = k$ relative to some unit i, a criterion value β, and an approximation rule X_i, the probability of judging x to be P can be written as:

$$Pr(\text{``}x \text{ is } P\text{''}|h(x) = k) := Pr(\sum_{i=1}^{i=k} X_i \geq \beta)$$

To make our example concrete, consider a discrete random variable X_i such that X_i can take only three values: 0, 1 or 2, with probabilities 1/4, 1/2 and 1/4 respectively. The variance σ of X_i is 1/2, its mean μ is 1. For each value k of a given height, and assuming the value of the criterion is $\beta = 185$cm, one can compute the probability $Pr(\sum_{i=1}^{i=k} X_i \geq \beta)$. Using the Central Limit Theorem, this value for each k can be approximated by $1 - \Phi(Z)$, with $Z = (\beta - k)/\sigma\sqrt{k}$, and Φ being the cumulative normal distribution. Figure 1 plots each such probability as a function of the height of individuals.[2]

---

[2]In this example and in the examples that follow, I am making the assumption that the random variable X_i is the same for each i, i.e. does not depend on the position of the increment. That assumption is not essential, in particular it can be relaxed in the application of the Central Limit Theorem, provided adequate
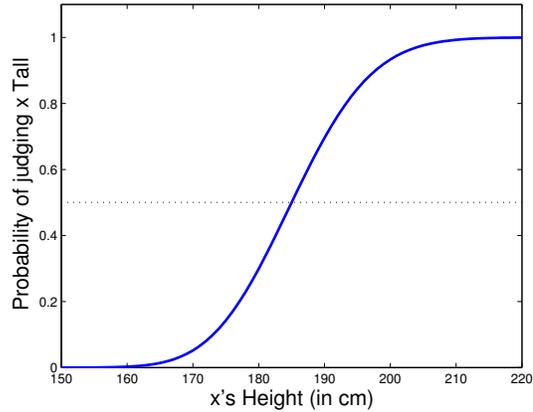
FIGURE 1. Probability of judging x tall (unit=1cm; criterion=185; variance=$\frac{1}{2}$)

We obtain a nice S-shaped curve for the acceptance function attached to judgments of the form "x is tall", centered around 185, for which the corresponding probability is 1/2. In this example, this reflects the fact that the error in representing height is symmetrically distributed around the actual value. If the mean of the random variable had not been of 1, but greater than 1, for example, we would obtain the same kind of curve, but shifted to the left and centered around a lower value. This would represent a particular bias, namely the case of someone overestimating height. If the mean were lesser than 1, this would be a case of height underestimation, shifting the curve to the right.

A key feature of this model is that the smaller the variance of the random variable, the steeper the transition between cases for which the probability of judging x tall asymptotes to 0 and cases for which it asymptotes to 1 (see [Egré and Barberousse, 2014] for more on this, and Figure 2 below for an illustration). Conversely, the larger the variance, the smoother the transition. For all such curves, moreover, we find a value x for which the probability of judging x tall is 1/2. Such a point, in line with the psychometric literature, may be called a point of subjective equality (PSE). On a frequentist interpretation, it represents the value for which, half the time, I would judge x tall, and half the time, I would judge x not tall (assuming a forced choice between "tall" and "not tall" judgments).

The model captures two aspects often considered to be missing in standard theories of vagueness: the first is the smooth transition between the clear(est) cases of the category (those for which the probability of judging tallness is 1) and the clear(est) non-cases of the category (those for which the probability of judging tallness is 0). The second is that it gives a derivation of the S-shape of the associated function: the shape results from the assumption that each unit of a given magnitude is measured with some error or noise.

––––––––––––

convergence conditions are satisfied. For present purposes, however, I confine myself to the assumption of an identical distribution.

Despite this, the model also raises two main questions for discussion. The first concerns what I called the criterion. Isn't the model committed to epistemicism in postulating that we categorize relative to some fixed threshold value? The second concerns the specific assumptions behind the idea that vague categorization depends on a process of magnitude representation. How general is the model? In particular, does it not tie our representations of magnitudes to a specific kind of measurement scale? In the next section, I first clarify the notion of criterion, and then go on to discuss the two parameters behind the notion of approximate magnitude representation.

3. The criterion

The model makes a substantive assumption in postulating that for each vague predicate P, we decide whether an object x satisfies P or not relative to some fixed value (what I have called the criterion, for reasons to be explained below). The model may be given an epistemicist interpretation because of that, but in spirit at least it is very distinct. In this section, I first argue that the criterion should be thought of as a subjective decision rule, distinct from the notion of objective cutoff and from what is commonly referred to as a standard. I rely in particular on the understanding of the notion of criterion in signal detection theory. I show some analogies between the proposed model and the SDT model, but emphasize some important differences between the two. Those differences pertain to the difference between classifying and detecting.

3.1. **Comparison class, standard and criterion.** Epistemicism is the view on which our use of vague predicates tracks sharp properties in the world ([Williamson, 1992]). On a common interpretation of epistemicism, individual judgments of tallness are made relative to a common but unknowable cutoff point. A different view of vagueness, much closer to supervaluationism than to epistemicism, is that each competent speaker in a population, even relative to the same set of objects, may use a distinct implicit cutoff value to judge that someone is tall. Borel's model falls more naturally on the side of that second interpretation (see [Egré and Barberousse, 2014]). On that view, my cutoff value for "tall" and your cutoff value for "tall" may differ without error. Speakers can competently and faultlessly diverge on the extension they assign to a predicate like "tall", as much as they can differ on the interpretation of "rich" or "expensive".[3]

To justify the idea that cutoff values can vary faultlessly between individuals, I need to say more about the way in which this cutoff might be computed. As investigated by Kennedy and van Rooij, vague predicates like "tall", "rich" or "expensive" involve a comparison to some explicit or implicit standard (see [Sapir, 1944], [van Rooij, 2011a], [Kennedy, 2011]). Consider the three following sentences:

(1)   John is tall.

(2)   John is tall for a 7-year old.

_____

[3]See [Wright, 1995] on the idea of an essential connection between vagueness and faultless variability. See [Kennedy, 2013] for a recent discussion.

(3)   John is tall compared to Mary.

A sentence like (1) is only meaningfully asserted by a speaker if the speaker has in mind an implicit *comparison class*, as made explicit in (2), and an implicit *comparison point* or *standard of comparison*, as made explicit in (3). In the case of (3), John and Mary constitute a comparison class, and Mary's height serves as the comparison point or standard relative to which John's height is to be evaluated as tall or not. In a case like (2), a comparison class is explicitly specified, but no explicit reference point within the comparison class is specified. A common view of that standard (in relation to the adjective "tall") is as the average height within the comparison class (see [Bartsch and Vennemann, 1972]) or sometimes as the median height within the comparison class (see [Solt, 2011b]). The idea is that in order to judge whether John is tall for a 7-year old, we need to have some idea of how tall 7-year old children (or boys) are on average.

Now, a distinction I would like to introduce is between the notion of standard of comparison and the notion of subjective criterion. While the term "standard" is often used in linguistics for what I call "criterion", the distinction I am after is not purely terminological (see [Solt, 2011a] for related ideas). I view the standard of comparison as *some salient or typical reference point, implicit or explicit, relative to which the comparison is made*. If we agree to that definition, then, as argued by Kennedy for a sentence like (2), we cannot in general identify the standard with the subjective cutoff value such that being above that value would count as tall. In particular, we can say things like:

(4)   John's height is taller than the average height of a 7-year old, but I would not say
        he is tall for a 7-year old.

Presumably (when it comes to think of the median as the standard) one can also say:

(5)   John is (a tiny bit) taller than most 7-year olds, but I would not say he is tall for a
        7-year old.

The fact that one can utter such sentences does not mean that the notion of average or median is irrelevant in evaluating whether someone is tall or not. What it means, however, is that one's cutoff for whether John is tall can be located at a value that differs from either of those standards. [Fara, 2000] proposes that "tall" means "has significantly more height than is typical". In this case, we can take the typical height, whatever it is, to serve as the standard or comparison point. Where the cutoff is located, however, depends on what counts as "significantly more". The cutoff may in some cases just coincide with the standard, but as argued by Fara, the determination of that value is more often subjective and normatively grounded.

Let us consider a sentence like (3), in which a comparison point is made explicit. We can expect competent speakers to disagree on whether the difference in height between John and Mary counts as significant or not. Assume Lucy's own height is 160 cm, while Dave's is 185cm, and both have to judge whether John, whose height is 190cm, is tall in comparison to Mary, who is 180cm. To Lucy, the 10cm difference between John and Mary

may count as insignificant (because she views both of them as tall), while to Dave, whose height stands in between, that difference may appear more significant. If you don't like the example with "tall", a better case to show the interest-relativity of vague predicates is given by an adjective like "expensive". Consider Dave and Lucy, each having to judge whether "this car is expensive compared to that car". Assume that one costs 15000 euros, while the other costs 20000 euros. To Lucy, who earns much more money than Dave, that 5000 euros difference may not count as significant. To Dave, who can afford the first car, but not the second, the difference might count as significant.

One may object that Lucy and Dave both expand the comparison class so as to define a new standard of comparison (their respective height, or respective budget). But I think a more tidy account is to maintain the identification of the standard with a distinguished reference point, and to say that what counts as a significant deviation from a given comparison point is not necessarily set in the same way depending on the speaker, and can indeed vary depending on the subject's interests or aspirations. My suggestion therefore is that judgment of whether some item is "tall" or "expensive" involves at least three components: a comparison class, a standard or reference point based on the comparison class, and an appraisal of how much the target deviates from the standard. On my approach, the criterion value is *the significant value, relative to the standard, such that exceeding that value would in principle lead to a judgment of tallness.*

The distinction I am drawing here between standard and criterion can be further motivated by considering absolute gradable adjectives, as opposed to relative gradable adjectives. There are good arguments to assume that "bald" semantically means "having 0 hairs on their head" (see [Kennedy, 2007], [Burnett, 2012]). In this case, "0 hairs" fixes a common and interpersonal standard (and indeed, Kennedy calls absolute gradable adjectives like "bald" or "flat" *maximum standard* adjectives). But actual judgments of baldness pragmatically depend on how close one thinks a person is to that extreme value. In that case too, we can expect a fair amount of intersubjective variability in the setting of the criterion.[4]

Let us return now to a sentence like (1), which leaves implicit its comparison class argument. Imagine a conversational context in which Lucy and Dave are asked whether they think John is tall, without further specification of the objects relative to which the judgments are to be made. We can expect at least three sources of variability in this case: they may select different comparison classes (male adult vs. adult for example); even when they select the same comparison class, they may select different standards of comparison (they may have distinct representations of the average, supposing it is the standard, or just select distinct reference points within the comparison class); and finally, even when the standard and comparison class are the same for two agents, they may place their criterion

---

[4]What I call the standard here may therefore equally be an ideal value, an actual value or an average value, used as a specific comparison point. See [Egré and Cova, 2015, Bear and Knobe, 2016] for more on the distinction between statistical vs. normative comparison points. There can exist multiple standards or comparison points along a single dimension that are active for the same predicate. What I here call the criterion is the personal threshold value determined as a function of those (possibly multiple) comparison points. In [Egré and Cova, 2015], we use the term "standard" for what I here call "criterion".

differently, depending on what they count as a significant deviation from the standard. Most of the time, a difference between two person's criteria may originate from the fact that they use different points of comparison, or that they represent the same intended reference point ("the average height", let us say) differently. But this is not an adequate reason to conflate the two notions. In summary, the reason judgments involving vague predicates can differ between speakers can be viewed as a combination of context-relativity (regarding the comparison class, and the standard) and interest-relativity (regarding the criterion proper).

3.2. **The criterion in signal detection theory.** I can now be more explicit about the terminological choice concerning the notion of "criterion". Talk of a "criterion" is motivated by the use of that term in signal detection theory (see [McNicol, 2005], and [Egré and Bonnay, 2010] for a presentation pertaining to vagueness; the term is also used in the same sense in [Verheyen et al., 2010]). Signal detection theory is a theory of imperfect discrimination, intended to account for the behavior of subjects having to categorize events as signal or as noise. For example, a subject may have to detect the presence of a pure tone in a background of white noise, or whether some collection of dots on a screen exceeds some threshold fixed by the experimenter (viz. [Smith et al., 2003]). In signal detection theory, the discrimination behavior of each subject is modeled by two parameters, namely the subject's sensitivity (how good her discrimination abilities are) and the subject's attitude toward risk (how prone a subject is to reporting an event as a signal, depending on the cost of misreporting noise for signal, or of missing a signal event for a noise event). The notion of criterion thus refers to the particular *decision rule* taken by a subject in a situation of uncertainty (such as: report a signal when the number of dots exceeds a certain value / when the amplitude of the sound exceeds a certain level).
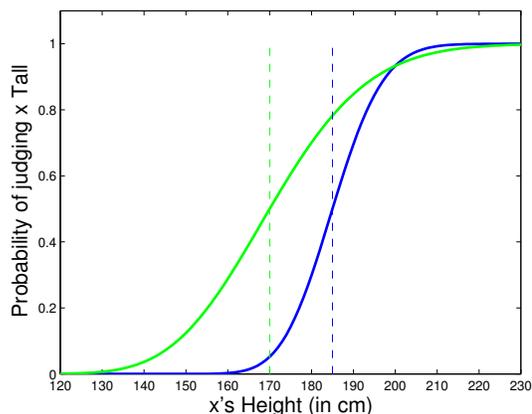

FIGURE 2. Distinct criteria, with distinct sensitivity

There is a broader analogy in this regard between the present model of vague classification for a "tall" vs. "not tall" judgment and the model of signal detection in a standard

Yes-No task. We have just seen that the criterion plays a role analogous to the *decision rule* in SDT. In the model we proposed for judgments about "tallness", the error with which each unit of measurement is represented plays a role analogous to the notion of noise in signal detection theory. By manipulating the variance of our random variable, we thus manipulate a notion of *sensitivity*.

Figure 2 below depicts two judgmental curves for "tall", in which we simultaneously manipulated both parameters, namely the variance of the random variable used for height estimation and the criterion value. The blue curve represents the theoretical acceptance curve of a subject with a high-positioned criterion (someone who would judge "tall" mostly individuals taller than 185cm, such as Dave in the above discussion) and the green curve that of a subject with a lower criterion (someone who would judge "tall" mostly individuals taller than 170cm, such as Lucy in the previous discussion). Furthermore, the blue curve is steeper than the green curve, representing the fact that the imprecision of the green subject is higher than that of the blue subject in representing height.

3.3. **Classification vs. Detection.** There are differences between the SDT model and my proposed model worth pointing out. In a detection task, there is a transparent difference to the experimenter between a mere noise event and a signal event (the experimenter knows when a pure tone event happens, or when the number of dots on a screen exceeds a given value). In a classification task (like judging someone to be "tall") there is no transparent property of "tallness" that we can assume to be available to a privileged subject.

This is an important difference between the SDT model and my model. Both models are about categorization behavior, and both involve a notion of imperfect discrimination, but in the SDT case, a subject with perfect sensitivity would have no choice regarding her decision rule, for the difference between signal and noise is factually determined (and controlled by the experimenter). The situation is intuitively different in a case of classification such as deciding whether someone is tall or not. In that case, the intuition is that even a subject with perfect sensitivity (able to identify heights with absolute exactness) would remain free to set her criterion at different locations *faultlessly*. As hinted above, it remains legitimate to interpret the proposed model of categorization in a way that brings it closer to epistemicism. In that case, we may think that "tall" tracks a determinate property, and that subjects with perfect sensitivity ought to all align their criterion to the objective cutoff for tallness. As I argued earlier, however, such a view appears very unappealing for predicates that are more obviously interest-relative than "tall", like "expensive", and even more so if all vague predicates incorporate the same element of interest-relativity and subjectivity.

A related difference between the signal detection model and our model of categorization concerns the derivation of the psychometric curves of the type represented in Figures 1 and 2. A curve like the one represented in Figure 1 or 2 represents the probability of judging someone to be tall, given the observation of his height x. The SDT-counterpart would be the probability of judging that there is a signal, given distinct values of the subject's evidence variable (representable, for instance, as a function of the amplitude of the pure tone). In SDT, this probability is a posterior probability of reporting a signal given some

value of the evidence variable. This posterior probability is standardly derived from the criterion value and from two prior probability distributions, namely the prior probability of observing specific values of the evidence variable given a signal, and the prior probability of observing various values of the evidence variable given a mere noise. In our model for the classification of someone as tall vs. not tall, there is no obvious analogue to those two prior distributions. We could introduce them if we thought of the difference between "being tall" and "being not tall" as an objective difference such as the one between signal and noise. On an epistemicist approach, this would be a natural thing to do. But once we leave out the idea that "tall" would track a determinate, subject-independent property, such a move becomes much less natural.

A different way of phrasing the same idea is to say that, on an epistemicist conception, it would be quite natural to derive our model of categorization from the exact same assumptions that underly the signal detection model. Indeed, the epistemicist view of vagueness is one on which vague classification is fundamentally a phenomenon of imperfect detection. But from a nonepistemicist perspective, classification is not wholly reducible to detection, and we can no longer rely on direct analogues to the notions of signal and noise.

3.4. **Fixed vs. probabilistic criterion.** Opponents of epistemicism about vagueness should welcome the idea that the criterion can vary faultlessly between subjects, but they may still find unpalatable the assumption that vague judgments are made relative to a crisp, albeit subject-relative, criterion.

A first response to this is that the assumption of a fixed criterion is in a sense an artifact of the theory: the criterion is only meaningful when coupled to the mechanism of approximate magnitude estimation. The point is that in combination with that mechanism, we can already remove a number of the unappealing features of a 'cutoff' theory of vagueness (see also [Egré and Barberousse, 2014] for more arguments). The situation is comparable to what we have in signal detection theory: in SDT the noise is supposed to be globally in the perception-and-decision system, hence the assumption that the noise originates merely in discrimination, and not in the setting of a criterion, is an idealization. What matters is that the combination of noisy discrimination with a fixed criterion is sufficient to account for observed behavior.

That first response is very instrumentalist in spirit, however. A better response is that the criterion could indeed be handled probabilistically, compatibly with the model. Currently, the model assumes that vagueness originates only in the process of analog magnitude estimation, but one may object that that process cannot fully account for vagueness. Arguably, even a subject with perfect discrimination is not able to keep a constant decision rule, as Borel himself admitted in his account of vagueness. Currently, a prediction the model makes is that a subject having to judge whether "x is tall", when his or her inner criterion is set at 178 cm on his or her inner scale, should show the same behavior as when asked to judge whether "x is taller than 178cm", assuming the subject's discrimination to be constant. If the observed pattern of judgments differed systematically between those two cases, and in particular if responses in the former case were associated with a flatter response curve, this would be a reason to handle the criterion probabilistically. A difficulty

with that objection is that the mention of an explicit standard may interact with the way the process of magnitude estimation is carried out. In particular, the granularity in the approximation could change from one kind of case to the other (more on granularity below). For the objection to be operative, one should make sure to use sufficiently different standards to control for such implicit changes of the approximation mechanism.

Nevertheless, I grant the force of the objection, and I agree that handling the criterion probabilistically may not only afford an additional degree of freedom in the model, but may be more explanatory of vagueness than a model in which the criterion is not handled probabilistically. Another motivation to handle the location of the criterion probabilistically is when uncertainty about the position of the criterion of other agents is itself a relevant parameter in judging with vague predicates. This uncertainty is essential for thinking about how we interpret vague utterances in communicative contexts (see [Kyburg, 2000, MacFarlane, 2010, Frazee and Beaver, 2010, Lassiter and Goodman, 2014, Lassiter and Goodman, 2015]). I do not address this issue in this paper, but will return to it briefly in the concluding section of this paper. In what follows, however, I will continue to work with the assumption of a non-probabilistic criterion, firstly because the model is simpler, and secondly because I will argue that there are cases of vagueness in which the assumption of a fixed criterion is not the problem (see section 4.1).

4. Magnitude Representation

Besides the criterion, the present model of vague categorization contains two other parameters in need of clarification. Those two parameters, which are the input unit value and the approximation rule for that unit, together serve to represent the way in which an objective magnitude is subjectively represented, and to compute whether an item stands above or below the criterion value. The assumption of those two parameters raises three main issues, however, which I address in this section. All three issues pertain to fundamental issues about measurement and in particular to what it means to represent "how much" of a property is satisfied.

4.1. **Numbers as input.** The first issue concerns the seeming difference between judgments based on numerical information and judgments made without that information. Consider the predicate "tall' again. Prima facie the model may appear to account for a limited class of situations, in which we have to judge whether someone is tall or not by first building an explicit estimate of his of her exact height. But as pointed out in Section 2, there are obviously situations in which we are given exact measurements for someone, and still have to judge whether the person is tall. Arguably, by being told that John is 178cm tall if I do not see him, I do not necessarily gain facility in judging whether John is tall, for I still have to mentally represent that height and compare it to some innerly represented standard. It can be easier, of course, in case I settle for a specific numerical value as my criterion, but that strategy is not always available. Or consider the vague adjective "expensive". Most of the time, when you have to judge whether something is expensive, you know the price of the item in whatever currency is being used. You don't have first to build an estimate of the exact price of the item, and then to issue a decision

about whether that item is expensive. For such cases, our model of categorization might appear inadequate.

This worry is easily dispelled. The model would indeed be of very narrow use if it were meant to apply only to situations in which no numerical information is available to the person who categorizes. In effect, the model is also applicable to situations in which exact measurements are available as input. Consider the case of price. I see that the price of a tray of peaches is 5 euros, and I have to decide whether that's expensive. My evaluation of "5 euros for a pack of 12 peaches" as expensive or not is based on the value or utility I attach to "5 euros", both relative to the average tray of peaches and relative to how much that represents for my budget. Some mechanism must be such as to map the exact numerical input on a subjective scale of relative utility whose structure is not given to me transparently.[5] Likewise when I am told that John is 178cm tall, without seeing him. In order to judge whether John is tall – say in comparison to other English male adults –, I am also faced with the task of mapping out that absolute value on a scale of relative magnitudes, in order to determine whether that height counts as tall.

On the present view, therefore, there is no fundamental difference between judging that someone is tall based on merely seeing him and judging that he is tall based on getting precise numerical information about his height. Both situations imply mapping out the perceived or described height onto a subjective scale of measurement. This conversion mechanism, whether for perceived magnitudes or for reported magnitudes, is what we model in terms of the random variable.[6]

As discussed above, one may still object that in cases in which the magnitude of an object is communicated in digital terms to someone, the main uncertainty of the subject is not about the input, but essentially about the location of his or her criterion. On that alternative view, when told that a tray of peaches costs 5 euros, my uncertainty would be merely about which value to count as "expensive". We could propose a dual model, consequently, and decide to set the uncertainty not on the stimulus, but on the criterion instead. Again, I have no quarrel with the suggestion to handle the location of the criterion in probabilistic terms. Ultimately, however, I think a reductive analysis of vagueness must say something about cases in which there is no uncertainty about the standard or comparison point, but in which some vagueness remains. Consider a situation in which I know the average tray of peaches to cost 4.9 euros, and in which this turns out to be my inner criterion. I have to decide whether a 5 euro tray is expensive or not. The issue here is how I represent the distance between 4.9 and 5, assuming 5 to be the standard.

---

[5]See [Stevens, 1966] for a similar argument concerning the perceived value of money.

[6]Further support for the idea that even digital numbers are analogically represented can be found in the *distance effect* evidenced by [Moyer and Landauer, 1967], who found that the time required to compare Arabic digits as well as the accuracy in comparison are inversely proportional to the numerical distance between the digits. This effect, widely replicated including with number words (see [Dehaene, 1997], [Feigenson et al., 2004]), indicates that precise quantities too are transduced with noise. In Moyer and Landauer's original experiment, the mapping operates without even the need for unit specification. It is very likely that it remains operative when digital numbers are communicated in relation to specific quantities (such as prices, heights, etc).

Even if my criterion is 4.9 euros, that is, even when the criterion value coincides with the standard, 5 euros may not automatically be judged expensive, assuming 4.9 is evaluated as "close enough" to the criterion. But to say that it is "close enough" here means that 5 and 4.9 can be innerly represented as sufficiently similar. The main mechanism about vagueness here is not the shifting criterion, but rather, the fact that the distance to the criterion is best represented as a random variable (see [Borel, 1907], [Egré and Barberousse, 2014]).

4.2. **Granularity.** The second issue concerns the choice of the scale and of the granularity relative to which the conversion operates. In our model, we assumed that each centimeter is represented with some approximation. But why not assume instead that each decimeter is counted with some approximation, or conversely, that each millimeter is counted with some approximation? In other words, how can we assume that magnitude estimation is done at the level of the centimeter, rather than at a more coarse-grained or more fine-grained level? More generally, what can we assume to be the basic input unit?

First of all, the assumption that discrete increments of magnitudes are mentally added up could be jettisoned in a model in which continuous variables are used instead of discrete variables. If we take the model at face value, however, we can nevertheless make sense of it by treating granularity as a parameter. We may be relatively sensitive, in judging about people's tallness, to differences of a centimeter, and much less about differences of a millimeter. Conversely, differences by decimeters only may be too coarse-grained to correctly fit an ordinary speaker's judgments about heights. More generally, we know that our judgments about tallness are fundamentally relative to a comparison class ([Klein, 1980]). Our judgments concerning whether a building is tall are likely based on a different metric from our judgments whether a human male adult is tall. When judging of a person's height, we may be sensitive to a difference of a centimeter. For a building, the minimal difference is more likely to be a storey. This implies that when we describe analog magnitude estimation as the conversion of objective units of measurements into subjective units, the relevant input unit at the psychological level need not be the same as the one in which the stimulus is standardly measured (for instance if we plot judgments of tallness for buildings against their height in meters, and if it turns out that storeys are better units relative to which magnitudes are subjectively represented).

The model is therefore neutral in principle regarding the choice of the input scale and of the basic input unit. Furthermore, the choice of a granularity is testable. That is, we can compare the predictions made by the model when the granularity varies. Consider our toy model for illustration. Keeping the imprecision constant, we can compare the probabilities of judging a person x tall for the same unbiased random variable ($\mu = 1$, $\sigma = \sqrt{1/2}$), when the granularity is 1cm, and the criterion 180cm, and when the granularity is 10cm, and the criterion 18dm. Each time, the probability of judging x tall is $1/2$ at the respective criterion value, but the function with the coarsest granularity is much flatter, and asymptotes much less rapidly. The top figure in Figure 3 below represents the associated psychometric curves.[7] Intuitively, the green curve is implausibly flat when referred to judgments of

_____

[7]For the green curves, we computed the values at each multiple of a dm, and interpolated the other points.
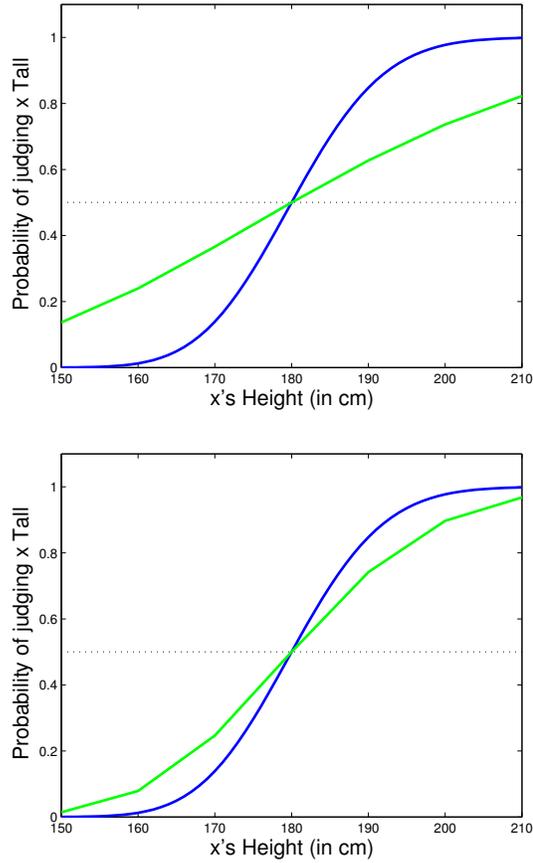
FIGURE 3. Changing the granularity: imprecision is the same for the two plots in the top figure; precision is increased in the bottom one for the green curve.

tallness for ordinary male adults (say), since even a height of 210cm does not guarantee a judgment of tallness. However, we changed the granularity without correcting for the variance, which is a very implausible move. More likely, the choice of the granularity will occasion a change in the imprecision, and a coarser granularity should go with a higher precision.[8] The bottom figure represents the same curve, but increasing the precision ($\sigma = \sqrt{1/8}$). As can be seen, the two plots come much closer.

---

[8]The claim may seem counterintuitive, but by this I mean that the coarser the granularity, the more reliable our judgments will be that a magnitude should be assigned to this or to that number on the corresponding scale. A subject asked to estimate the height of a building with a margin of error of 10 meters is less likely to make errors than one asked to evaluate the height of that building with a margin of 1 meter.

Ultimately, the choice of the relevant granularity may also be dictated by how realist one thinks the model ought to be from a computational viewpoint. A coarser granularity obviously means a lighter computational load. Whichever granularity is chosen, however, we see that the judgmental patterns predicted by the model remain fundamentally the same (that is, the S-shape of the curve, and the gradual transition from clear cases to clear non-cases).

4.3. **Additive measurement.** A third issue concerns the measurement-theoretic assumptions behind our model. To derive our psychometric curves, what I assumed is that some psychophysical process is taking place, which parallels a process of additive measurement. Consider "tall": to get a precise measure of someone's height, we concatenate units (such as centimeters) so as to match the person's height. Similarly, I assumed that the subjective representation of height is done by concatenating the noisy approximations of units of a person's height (whatever they might be).[9]

The model may appear implausibly restrictive relative to the gamut of vague predicates. For adjectives like "tall", "old", or "expensive", we can rely on ratio scales of height, age or price, in order to set measurement units, relative to which psychometric curves like the one shown in Figures 1-3 can be plotted. But there are cases of vague predicates for which not even an interval scale appears to be available. An example discussed by [Bueno and Colyvan, 2012] in their account of vagueness is the predicate "being a religion".[10] They write (p. 26) that:

> 'Such predicates do not have a natural numerical ordering from cases where they apply, through borderline cases, to cases where they do not apply'

Indeed, the respects relevant to judge whether a set of activities is a religion are not easy to map on a numerical scale. The reason for that, however, is that "being a religion" is obviously a multidimensional predicate, for which no obvious dimension of comparison comes to the fore.

A possible answer is that we should not conclude from the absence of an *obvious* numerical ordering for such predicates to the absence of *any* level of numerical representation.

---

[9][Stevens, 1957] calls "prothetic" those continua on which "discrimination is mediated by an additive or prothetic process at the physiological level". The model I propose basically handles vague predicates as prothetic in that general sense. Whether the sort of process here envisaged has any psychological reality is certainly not obvious. In particular, the model shares a number of assumptions with what is known as the Rasch model of categorization (see in particular [Verheyen et al., 2010]). The Rasch model, as far as I can see, does not make specific assumptions about underlying processes. A comparison between both models lies beyond the scope of this paper.

[10]See also [Sassoon, 2010] for a discussion of the typology of gradable adjectives in relation to measurement scales. Sassoon argues that several adjectives, among them negative adjectives such as "short", do not appear to come with a ratio scale, but only come with an interval scale instead. However, her discussion of the typology of ratio scales mostly concerns the lexicalization of ratio phrases, and is not about psychophysics proper. While Sassoon expresses skepticism toward Stevens' hypothesis of the universality of ratio scales for prothetic predicates, she is nevertheless cautious not to reject it (see the discussion in 3.6 of her paper). For example, she admits that an adjective such as "happy", however multidimensional the predicate might be, could be used with a ratio scale in mind.

For multidimensional predicates, we are generally able to give numerical ratings on some ordinal scale. For example, we would be able to rank activities on a scale from 1 to 7, with 1 counting as "very clearly not a religion" to 7 for "very clearly a religion". This may appear to miss the objection, however, for the gist of our model is the assumption that each unit of measurement along an interval scale is represented with some approximation. Postulating such units only makes sense if we can assume a meaningful notion of difference between the numbers assigned at each level.

A bolder answer, however, may be that we are only able to rank items on an ordinal scale based on an ability to internally represent magnitudes relative to interval scales. [Stevens, 1966] surveys various experiments that can be used to support the hypothesis. In one cited experiment, by S.I. Perloe, the "prestige" of various professional activities was measured by both direct magnitude estimation and on a 7-point rating scale, showing a logarithmic relation between the two scales, and showing participants to locate more than 100 activities along a 30-level magnitude scale.[11] What this suggests is that participants were able to map out the prestige of various activities to a specific interval scale, which may in some cases be a ratio scale. Conceivably, one's categorical judgments of whether an activity comes out as "prestigious" or not may be parasitic on the position of a criterion along such a scale, despite the fact that the notion of prestige is obviously multidimensional (it depends on various cues, such as salary, level of education, and so on).

The same kind of analysis is applicable to a predicate like "being a religion". When pressed to assess whether an activity is more of a religion than another, we likely look out for the satisfaction of various features and for the extent to which they are satisfied. To use one of Bueno and Colyvan's examples, Brazilian soccer may count as somewhat close to a religion because of the devotion shown by soccer fans. One cue for devotion is how much people are likely to pray, how much they accomplish certain rituals, and so on. The claim is that for every multidimensional predicate, we might be able to extract quantifiable features, allowing one to build a representation of how much of a property is satisfied. If so, then the model can be adapted in principle to accommodate categorization with all kinds of vague predicates.

5. Borderline cases and judgments of clarity

Having clarified the main ingredients of the model, I shall now discuss the way in which it might help to characterize borderline cases of a vague predicate. Borderline cases of a vague predicate have been given several characterizations in the literature. A common characterization we encounter (see [Black, 1937]) is in terms of *instability*: borderline cases of a vague predicate are cases for which our judgments are unstable. More explicitly, borderline cases of a vague predicate are cases of which we feel attracted both toward applying the predicate and toward denying the predicate (viz. [Peirce, 1902], [Schiffer, 2003]). Another characterization is as cases that are neither determinately P nor determinately not P, or – under a certain analysis of "determinately" – as cases that are neither clearly P

---

[11]See also [Krantz et al., 1971]:141, who write (admittedly with some caution) "it may be possible to obtain orderings of intervals from a properly designed rating scale."

nor clearly not P (viz. [Williamson, 1994]). I will refer to it as the characterization of borderline cases in terms of *unclarity* (either way).

In this section I would like to show that a probabilistic model of vague categorization sheds an interesting light on both characterizations. Moreover, I will show how it can be related to the pragmatic characterization, put forward in [Cobreros et al., 2012], as cases of which we can deny (strictly) both that they are P and that they are not P, but also assert (tolerantly) that they are "P and not P" (see particularly [Ripley, 2011], [Alxatib and Pelletier, 2011] and [Egré et al., 2013] on the acceptability of such borderline contradictions). In particular, the current model gives a good account of the Hump Effect reported in [Egré et al., 2013], namely the fact that participants can assent more to "P and not P" than to either conjunct in borderline cases ([Alxatib and Pelletier, 2011]).

An important caveat at this point is that the account to be given of the Hump Effect does not depend in an essential way on the specific assumptions made concerning the function f leading from a physical magnitude h to its sensory transduction $f(h)$. The assumptions of the previous section show how to derive psychometric curves, but the mechanism which I will present in this section is applicable to psychometric functions generally, irrespective of their particular derivation, provided the functions are S-shaped, and provided a notion of criterion is available. This point is important to bear in mind, especially to those who would find the assumptions made on f in the previous paragraph too restrictive.

5.1. **Instability.** Let us consider the instability perspective on borderline cases first. In our model, the natural counterpart to the instability of judgment for a vague predicate is what we called the point of subjective equality, namely the value of the stimulus for which the probability of judging someone to be tall is $1/2$. In most of the cases I described in the previous section, that value turns out to correspond to the criterion value, but remember that in general it need not. For example, in case we have a random variable that is not symmetrically distributed around the actual magnitude of the stimulus, but in which heights would be overestimated, the point of subjective equality is reached for a smaller value than the criterion value. In principle, however, what the model predicts is that an item x for which the probability of judging Px is 0.5 is a case of maximum judgmental instability. Hence, if anything is to be defined as a borderline case of a vague predicate, it should include at least those cases for which, having set the three parameters of our model, the probability of judging P is $1/2$.

Empirical studies of vagueness support this approach. In an experimental study on color categorization, run with V. de Gardelle and D. Ripley (see [Egré et al., 2013]), participants were asked to agree or disagree with the description of squares in the yellow-orange (respectively blue-green) region relative to the sentences "the square is (not) yellow" or "the square is (not) orange" (respectively "the square is (not) blue", "the square is (not) green"), we found the usual S-shaped psychometric curves. Additionally, the inflection point (or point of subjective equality) in those curves turned out to correspond to the region where acceptance of the conjunction "the square is yellow and not yellow" is also highest (see [Egré et al., 2013], Figure 14, and Figure 6 below).

In allowing us to derive psychometric curves, a probabilistic model of vague judgment thus gives us a straightforward articulation of the behavioral characterization of borderline cases as cases of maximum instability. When characterizing borderline cases as cases for which judgments are unstable, however, we said that such cases should at least *include* the point of subjective equality in the psychometric curves, but presumably they will include more points. Depending on how items vary in a comparison class, borderline cases of a vague predicate generally constitute a more or less extended region. Typically, a case for which one's probability of judging P to apply would be between 0.55 and 0.45, hence sufficiently close to 0.5, would likely count as borderline cases, and show the same overall instability. My suggestion here is that we may make a more interesting use of our model to account for the idea that borderline cases of a vague predicate form an extended region, provided we focus on the characterization of borderline cases as cases that are unclear either way.

5.2. **"Clearly" and criterion-shifting.** If we take literally the suggestion that borderline cases of a vague predicate are cases that are neither clearly P, nor clearly not P, we ought to think of the expression "clearly P" as a new predicate, distinct from P, for which one's criterion is likely to be different from one's criterion for P. Similarly, a "clearly not P" judgment ought to correspond to yet another criterion, distinct from the other two. Here again, we can draw on the analogy with signal detection theory to clarify this idea of using distinct criteria. Various methods can be used in signal detection to measure a subject's sensitivity, including the typical Yes-No task on the one hand, but also the so-called rating task. A rating task consists in giving a participant more options than "Signal" vs. "Noise" in order to report her experience. For example, a three-point rating task might give the options between "Definitely a signal", "Definitely a noise" and "Uncertain either way" (see [McNicol, 2005]). With three such points, two criteria are then postulated (see for example [Smith et al., 2003] for a two-criterion theory of uncertainty responses based on SDT).

In our framework, by analogy, we can think of borderline cases as cases that a subject would classify relative to two distinct criteria, a high criterion (for "clearly tall") and a low criterion (for "clearly not tall"). Figure 4 gives an illustration of three theoretical acceptance curves. For ordinary Yes/No judgments of the form "x is tall" (blue curve), I have assumed a baseline criterion of 180cm. To represent judgments of clarity the form "x is clearly tall", I have assumed that they are issued relative to a higher criterion, here set at 190cm. Conversely, I have assumed that clarity judgments of the form "x is clearly not tall" are issued relative to a lower criterion, set at 170cm. The amplitude of the shift, namely 10cm, is set arbitrarily here. A bigger gap between the high and low criteria, in particular, can be checked to lower the point of intersection between the "clearly tall" curve and the "clearly not tall" one. What is not arbitrary is the idea that "clearly + P", for a gradable adjective P like "tall", shifts the baseline criterion for P toward a greater height value, whereas "clearly + not P" will shift the baseline criterion for P toward a lower value. So the assumption is that a gradable property comes with a polarity, which negation inverts. Finally, the assumption that one's criterion for "clearly tall" and one's criterion for "clearly not tall" are shifted *symmetrically* around the criterion for "tall" is
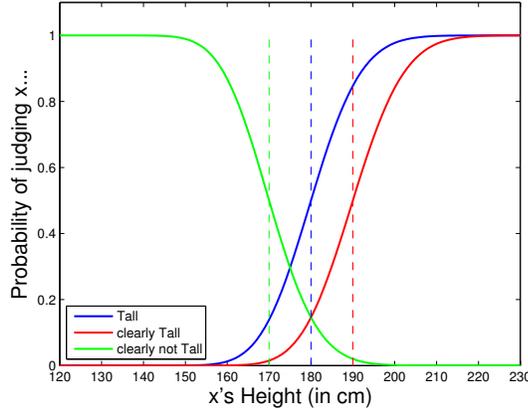
Figure 4. Three criteria; the dashed line represent the position of the criterion (clearly not tall=170cm; tall = 180cm; clearly tall = 190cm).

not mandated in principle. But it seems a natural assumption to make: in the present case the effect of the symmetry is that the point of intersection of the red curve and of the green curve is located at 180cm, which in this case is the criterion for "tall", and the point of subjective equality. One consequence of the symmetry assumption we get is that the peak for judgments of "borderline tall" will be aligned with the PSE.

To compute those three curves, I have moreover assumed that the subject's sensitivity is constant, and so does not vary depending on whether clarity is involved or not, and also that each curve is computed by the same algorithm, in particular with the same granularity. On the figure, this assumption is reflected by the fact that the blue curve (for "tall") and the red curve (for "clearly tall") have the same slope (and the green curve as a slope that is the negative of the other two). It is not an a priori matter, however, whether this ought to be the case in general. We could imagine that judgments of the form "'x is clearly P'" are made relative to a different granularity than judgments of the form "x is P", for example, and as a result, to get steeper curves for judgments of clarity than for basic judgments (see the discussion of granularity from the previous section).[12]

Under those assumptions, we see that an individual x of height 180cm, for example, is one for which the respective probabilities of judging "x is clearly tall" and "x is clearly not tall" get much lower than the probability of judging "x is tall" simpliciter (we get a probability of 0.5 for the latter, when $\sigma = \sqrt{1/2}$, against a probability of 0.1459 for the former two – at the point where the red and green line intersect, see Figure 4). This prediction is in agreement with our intuition about borderline cases.

5.3. **Deriving judgments of unclarity.** The next question is how to represent the probability of a compound sentence, such as "x is neither clearly tall, nor clearly not tall",

---

[12]I am indebted to James Hampton for bringing this issue of the interaction between clarity judgments and steepness to my attention a few years ago.
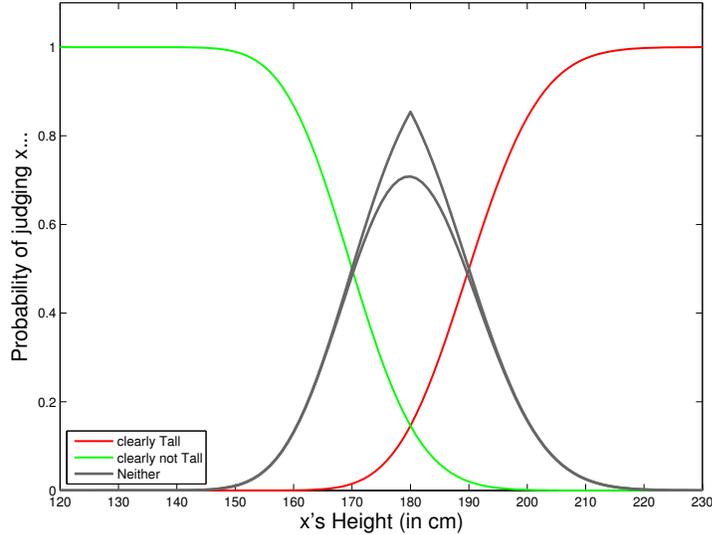
FIGURE 5. Lower and Upper bounds on the probability of judging x "neither clearly Tall nor clearly not Tall" (grey curves), as a function of the probabilities of judging x "clearly Tall" (red curve), and "clearly not Tall" (green curve)

taking it to be the meaning of "x is borderline tall". In all I wrote so far, I dealt with atomic information, except implicitly for the operator "clearly", which I have assumed to shift the criterion used for the predicate P, and except also for negation, namely for "not P", for which I have assumed the probability to be 1 minus the probability attached to "P". In the case of conjunctions, we may not assume in general that the probability of "x is clearly tall" and "x is clearly not tall" are independent. However, we can draw on the following inequalities for arbitrary conjunctions:

$$Pr(A) + Pr(B) - 1 \leq Pr(A \wedge B) \leq min(Pr(A), Pr(B))$$

Figure 5 represents the red and green curves from the previous figure, namely the two opposite judgments of clarity ("x is clearly tall", let us call it A vs. "x is clearly not tall", call it B), and the two grey lines give a representation of $Pr(\neg A) + Pr(\neg B) - 1$ and $min(Pr(\neg A), Pr(\neg B))$ respectively, which set a lower bound and an upper bound for judgments of the form "x is neither clearly tall nor clearly not tall".

Note that for both of them we get a curve whose maximum lies at the intersection point of the two other curves. It can be checked that this relation between the maximum and the intersection holds in general, and does not depend on the distance between the two criteria. What does depend on that distance, obviously, is the height of the curves. The lower the

point where the probability of judging "clearly tall" and "clearly not tall" intersect, the higher the curves are.

Such a representation is not without interest, for it may be used to derive further results about higher-order judgments of unclarity. For instance, we may wonder what it means to be *clearly borderline*. What our model suggests is that the more distant one's criterion for "x is clearly tall" and one's criterion for "x is clearly not tall", the higher the probability of "x is neither clear tall nor clearly not tall" can be. On the other hand, the closer the criteria, the lower the corresponding probability will be. On Figure 5, we see that the probability for "x is clearly tall" reaches much higher than that for "x is neither clearly tall nor not clearly tall". This fact may be used to buttress the intuition, expressed and discussed in [Egré and Bonnay, 2010], of an asymmetry between judgments of clear clarity and judgments of clear unclarity (either way). A further investigation of this issue would take us too far afield, however, and I will not say more about it here.

5.4. **Borderline contradictions and the Hump Effect.** The representation of judgments of borderline status obtained in the previous section bears a connection with yet another characterization of borderline cases, in terms of the relative assertability of borderline contradictions. [Ripley, 2011] and [Alxatib and Pelletier, 2011] have shown that ordinary speakers are much more prone than previsouly thought to use sentences of the form "x is P and not P" or "x is neither P nor not P" to characterize a borderline case of the vague predicate P. For example, Alxatib and Pelletier show that a man of middling height in a specific comparison class is classified as "tall and not tall" to a significant extent by ordinary speakers. On the other hand, the same subjects who classify the man as "tall and not tall" are much less prone to say "he is tall" or to say "he is not tall" separately (see [Alxatib and Pelletier, 2011] for details).
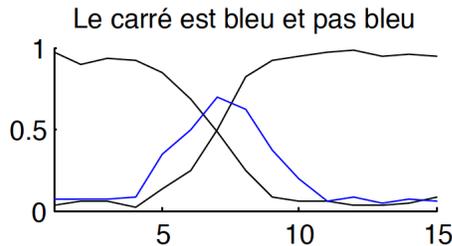

FIGURE 6. The Hump Effect (from [Egré et al., 2013]); the blue curve plots acceptance of sentences of the form "the square is blue and not blue" vs. acceptance of the conjuncts (in black)

In the aforementioned experiment on color categorization, my coauthors and I have sought to measure this effect ([Egré et al., 2013]). We called "hump effect" the existence of a significant difference between the maximum acceptability of borderline contradictions of the form "x is P and not P" and the corresponding acceptability of each conjunct separately. Figure 6 shows the percentage of "agree" responses obtained in that experiment,

with the black line representing the average agreement to "the square is blue" vs. "the square is not blue", and the blue line the average agreement to "the square is blue and not blue", relative to 15 shades at the border between green and blue (x-axis in that Figure). The point where the lines cross in that condition was found to be significantly below the peak of the blue curve.

To answer this problem, consider the theoretical curves we would obtain for judgments of the form "x is tall" and "x is not tall", assuming one's criterion for tall is 180cm. Given our model, the probability α of uttering x is tall, for x of height k cm, is equal to $Pr(\sum_{i=1}^{i=k} X_i \geq 180)$, and the probability of uttering x is not tall is equal to $Pr(\sum_{i=1}^{i=k} X_i < 180)$, that is $1 - \alpha$. Figure 7 plots the two corresponding curves (assuming $\sigma = \sqrt{(1/2)}$), which intersect at 180cm, where the probability of each judgment is 1/2. Now, how could we derive the probability of judging "x is tall and not tall" from those two?

If we refer a judgment of the form "x is tall and not tall", or indeed "x is neither tall nor not tall" to the same baseline criterion for each conjunct, as in shown in Figure 7, then we cannot account for the Hump Effect. For the probability of the conjunction necessarily lies underneath the probability of each conjunct, as show by the shaded area in Figure 7. Arguably, the probability of the conjunction is simply 0 when the same criterion is used for "tall" and for "not tall".

But in agreement with the account of the Hump Effect proposed in [Cobreros et al., 2012], a borderline contradiction of the form "x is tall and not tall" is in fact shorthand for "x is somewhat tall and somewhat not tall", and likewise "x is neither tall nor not tall" is shorthand for "x is neither clearly tall nor clearly not tall". The proposal made in [Cobreros et al., 2012], however, is that we need not use overt modalities, but we can use "tall" or "not tall" with different forces, corresponding to different standards for assertion. In the present setting, the counterpart to this distinction is the idea that we can express different propositions for "x is tall", "x is not tall", or indeed for "x is tall and not tall", depending on where one sets one's criterion for "tall". A sentence of the form "x is tall and not tall" can only be meaningfully uttered if each conjunct is endorsed tolerantly (in the context of the conjunction). To endorse "x is not tall" tolerantly is to not endorse "x is tall" strictly, and to endorse "x is tall" tolerantly is to not endorse "x is not tall" strictly.

Figure 8 plots three curves that differ only on the position of the criterion for "tall". The blue curve corresponds to the baseline criterion. The red curve corresponds to the high criterion, and the green curve to the low criterion. Here I assumed that the lower criterion, for tolerant judgment, is set at 170cm, and the high criterion, for strict judgment, at 190cm. Take someone whose height is 180cm. According to our parameters, the probability of judging that x is tall tolerantly is approximately 0.8541 in this example, while the probability of judging that x is tall strictly is about 0.1459. For each curve, the plot for the corresponding negation is represented as a dotted line of the same color. In agreement with the analysis given in [Cobreros et al., 2012], the probability of judging "not tall" strictly is relative to the low criterion: it is 1 minus the probability of judging "tall" tolerantly.

Dually the probability of judging "not tall" tolerantly is relative to the high criterion, and it can be defined as 1 minus the probability of judging "tall" strictly.
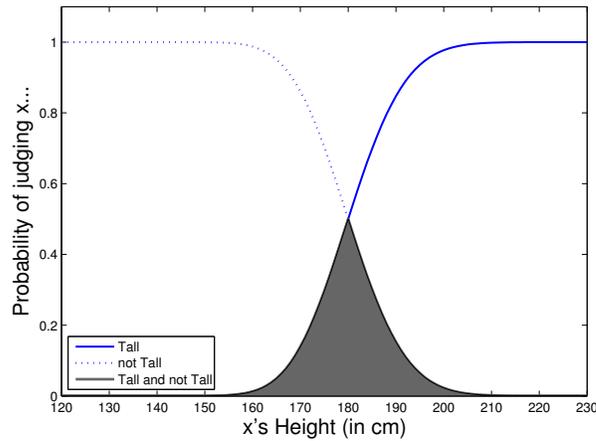
FIGURE 7. Only one criterion: the grey area represents the area between the minimum of the probabilities of "tall" and "not tall" respectively and the sum of those probabilities minus 1 (line $x = 0$); criterion=180cm; the probability of "tall and not tall" is less than the probability of "tall" or "not tall"
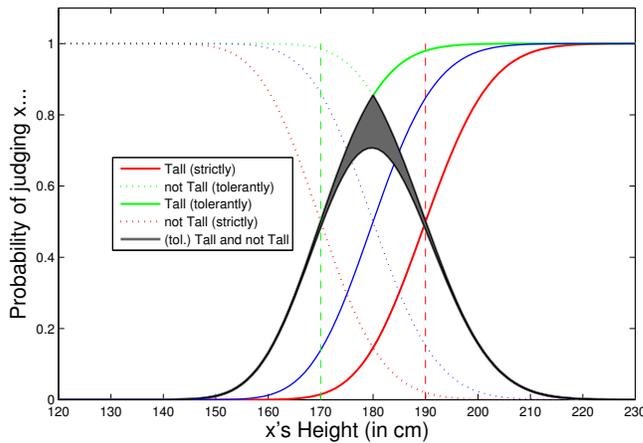

FIGURE 8. Three criteria: baseline, strict and tolerant (blue, red, and green), with only the latter two represented by vertical dashed lines; the dotted lines represent the acceptance curves for the negations; the grey area is defined as previously, but now relative to the probabilities for "tolerantly tall and tolerantly not tall"; that probability lies above the intersection points of the blue curves for "tall" and "not tall"

What Figure 8 shows is that we get a good fit for the Hump Effect observed in Figure 6 by assuming that a judgment of the form "x is tall and not tall" is issued relative to criteria for "tall" and "not tall" that differ from the criteria used for "x is tall" or for "x is not tall" when those sentences are considered or uttered in isolation. The current model predicts that the acceptance curve for "x is tall and not tall" will lie in the shaded area, which indeed lies above the point of intersection of the blue curves (corresponding to a probability of 0.5, like for the actual data reported in Figure 6), and likewise above the point of intersection of the red curves (namely the curves for "x is strictly tall" and "x is strictly not tall").

The present account parallels the account given in [Cobreros et al., 2012] for the Hump Effect. There we say that subjects who accept "x is tall and not tall" but who reject the conjuncts separately do so because they tend to interpret each conjunct strictly, when presented in isolation.[13] In the color case, it may be more accurate to say that subjects who accept "x is blue and not blue" more than "x is blue" or "x is not blue" do so because in the context of the conjunction, they accept each conjunct more tolerantly than when presented with the conjuncts separately. The underlying psychological and pragmatic mechanism is basically the same, however.[14] The main innovation here is that by using probabilities we can account for actual responses in a way that would not be possible if we were to stick to the more algebraic approach used in [Cobreros et al., 2012].[15] Indeed, the approach taken in [Cobreros et al., 2012] is implicitly four-valued, or three-valued when the notion of a classical extension is discarded. With only three or four semantic values, we would get a poor fit of the data. Even though the present account relies on exactly the same

---

[13]For more on the pragmatic explanation of the acceptance of borderline contradictions, I refer to [Alxatib et al., 2013] and [Cobreros et al., 2014].

[14]In the color experiment, the response curves for "x is blue" and "x is not blue" do intersect at a probability of about 0.5, and are roughly symmetrical. It is therefore natural to refer both sentences to the same "baseline" criterion. If on average participants were using distinct criteria for "blue" and "not blue", the curves should intersect at a distinct position, as shown in Figure 8. Hence, we do not need to assume that "blue" and "not blue", when presented in isolation, are evaluated relative to separate *strict* criteria. This is an interesting difference with the setting of [Alxatib and Pelletier, 2011]'s experiment, where some subjects who check True to "x is tall and not tall" check False to "x is tall" and to "x is not tall". To explain the behavior of those subjects, it can't be the case that the separate sentences "x is tall" and "x is not tall" are referred to a common criterion. Instead, we need to assume that the conjuncts, when uttered in isolation, are assessed relative to distinct criteria. The difference may be attributable to the fact that, in the color categorization experiment, the conjunctive sentence and the conjuncts were never presented together on the same screen, but always in separate blocks, whereas in the setting of Alxatib and Pelletier's experiment, they are presented simultaneously. A test of the present model on Alxatib and Pelletier's data based on those assumption confirms that the observed proportions of True answers to the "and" and "neither" descriptions fall, in the case of the man of middling height, in the interval predicted from the observed proportions for the conjuncts. However, the number of items is too limited and the predicted intervals too narrow for that test to be reliable. Finally, the present model does not concern itself with finer differences in acceptance between "x is tall and not tall" vs. "x is neither tall nor not tall". I refer to [Egré and Zehr, 2016] for a study of that question.

[15]My use of "algebraic" vs. "probabilistic" is inspired from [Luce, 1959], who draws a similar opposition in relation to his account of imperfect discrimination.

mechanism as the strict-tolerant framework, the use of probabilities erases the dividing lines that appear to go with the notions of strict vs. tolerant extensions in the original account.

6. Conclusion and perspectives

In order to conclude this paper, let me first summarize our progression. Starting from the idea of vagueness as approximate measurement, I have proposed a model of vague judgment that depends on two main ingredients: a notion of approximate magnitude representation and a notion of individual decision rule or criterion. I have then used this model to cast light on a certain pattern of judgments involving borderline cases.

There are two advantages of this model over many of the extant approaches: one is that it is able to predict a certain amount of faultless variability between competent speakers in their use of vague predicates, based on the idea that competent speakers may use different approximation rules as well as different criteria in their application of the same predicates. The second is not specific to this model, but pertains to the use of probabilities more generally: for each of the vague predicates we have considered, whether simple or complex ("tall", "not tall", "clearly tall", "neither clearly tall nor clearly not tall", "tall and not tall"), we can account for a gradient in their acceptance. In that sense, although the theory relies on a notion of crisp inner criterion, we see that it does not predict sharp judgmental cutoffs (see [Smith, 2008] and [Egré and Barberousse, 2014] for more on the issue of 'jolts' in relation to vagueness).

Several aspects of our model deserve further investigation. One issue to which the last section invites us to think more concerns the link between a probabilistic theory of judgment like the one here proposed and a logic of vagueness proper. The link established in the last section between the present model and the strict-tolerant account of vagueness already give us some hints about this connection. What we see is that we can give the notions of strict vs. tolerant assertion, as defined in [Cobreros et al., 2012], a naturalist understanding, by relating them to the notion of psychophysical criterion. In particular, we have seen in what sense a high-criterion/low-criterion theory of vague judgment delivers better empirical results than a single-criterion theory, when it comes to explaining actual judgments about borderline cases. This naturalist inspiration was part and parcel of the motivations given in [Cobreros et al., 2012] for the notion of tolerant extension for vague predicates, and in the antecedent work by [van Rooij, 2011b] connecting vagueness and the notion of semi-order due to Luce. It is now taken a step further.

A second issue concerns the form of what I have presented as the transduction function f in Section 2, mapping physical magnitudes to sensory magnitudes. As was stressed earlier, while our assumptions on f give us a specific derivation of psychometric curves, the account of the Hump Effect presented in the last section does not fundamentally depend on those assumptions. The proposed account of the Hump Effect remains available for any psychologically plausible derivation of S-shaped psychometric functions, provided the derivation involves some notion of criterion such as the one used in my own theory. This feature of the model is welcome in my opinion, but it begs the question of the psychological

reality of the process of magnitude estimation that I have postulated. The main worry one might have about the present model is the idea that magnitude estimation is performed analytically, namely by the summation of noisy estimates. An alternative approach is to think our perception of heights and other magnitudes, like that of discrete numerosities, is more direct, that is that different magnitudes are directly encoded by distinct Gaussian variables. I leave a comparison between the model of this paper and alternative models of magnitude estimation for further work.

A third issue which I did not have the space to broach in this paper concerns the problem of communication with vague predicates (see [Kyburg, 2000], [Lawry, 2008], [Frazee and Beaver, 2010], [Lassiter, 2011], [Goodman and Lassiter, 2014], [Lassiter and Goodman, 2014], [Lassiter and Goodman, 2015], [Qing and Franke, 2014]). Two speakers who enter a conversation and make judgments using the word "tall" can be described in our model as two agents who enter the conversation with possibly distinct criteria and distinct magnitude estimation apparatus (see Figure 2). This picture is an invitation to say more about the way in which the prior judgmental profile of an agent is to interact with the judgmental profile of another agent. Here I have focused on the probability for an agent to apply a vague predicate to a given stimulus, in isolation of other speakers. But there is obviously a reciprocal problem, which is to infer which value an item x might have upon hearing "x is tall" or "x is not tall". What our theory gives us is a recipe to compute the probability of judging an item x to be P given that it has a certain psychophysical characteristic $h(x)$. Using Bayes' rule, we can in principle use it to compute the probability that some item has the psychophysical characteristic $h(x)$ given that it is judged P. Similarly, we may exploit the notion of subjective criterion further: depending on the risk associated with miscommunication, an agent who is uncertain about someone's judgmental profile may shift her criterion to a higher or lower position depending on the stakes. Furthermore, one's uncertainty about another agent's criterion is certainly the best motivation to handle the position of the criterion probabilistically (see [Lassiter and Goodman, 2015]). The task of describing communication with vague predicates is beyond the scope of this paper. My hope is that the current model can be put to use in that broader enterprise.

References

[Alxatib et al., 2013] Alxatib, S., Pagin, P., and Sauerland, U. (2013). Acceptable contradictions: Pragmatics or semantics? A reply to Cobreros et al. *Journal of philosophical logic*, 42(4):619–634.

[Alxatib and Pelletier, 2011] Alxatib, S. and Pelletier, F. (2011). The psychology of vagueness: Borderline cases and contradictions. *Mind & Language*, 26(3):287–326.

[Bartsch and Vennemann, 1972] Bartsch, R. and Vennemann, T. (1972). The grammar of relative adjectives and comparison. *Linguistische Berichte*, 20:19–32.

[Bear and Knobe, 2016] Bear, A. and Knobe, J. (2016). Normality: part descriptive, part prescriptive. Manuscript, Yale University, under review.

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

[Borel, 1907] Borel, E. (1907). Un paradoxe économique: le sophisme du tas de blé et les vérités statistiques. *La Revue du Mois*, 4:688–699. English translation by P. Égré and E. Gray ("An economic paradox: the sophism of the heap of wheat and statistical truths"), in *Erkenntnis* 79:5, 1081:1088.

[Bueno and Colyvan, 2012] Bueno, O. and Colyvan, M. (2012). Just what is vagueness? *Ratio*, 25(1):19–33.

[Burnett, 2012] Burnett, H. (2012). *The Grammar of Tolerance: On Vagueness, Context-Sensitivity, and the Origin of Scale Structure*. University of California, Los Angeles. PhD thesis.

[Cobreros et al., 2012] Cobreros, P., Egré, P., Ripley, D., and van Rooij, R. (2012). Tolerant, classical, strict. *The Journal of Philosophical Logic*, pages 1–39.

[Cobreros et al., 2014] Cobreros, P., Egré, P., Ripley, D., and van Rooij, R. (2014). Pragmatic interpretations of vague expressions: Strongest meaning and nonmonotonicity. *Journal of philosophical logic*, pages 1–19.

[Dehaene, 1997] Dehaene, S. (1997). *The Number Sense*. New York:Oxford.

[Egré and Barberousse, 2014] Egré, P. and Barberousse, A. (2014). Borel on the heap. *Erkenntnis*, 79(5):1043–1079.

[Egré and Bonnay, 2010] Egré, P. and Bonnay, D. (2010). Vagueness, uncertainty and degrees of clarity. *Synthese*, 174(1):47–78.

[Egré and Cova, 2015] Egré, P. and Cova, F. (2015). Moral asymmetries and the semantics of 'many'. *Semantics and Pragmatics*.

[Egré et al., 2013] Egré, P., de Gardelle, V., and Ripley, D. (2013). Vagueness and order effects in color categorization. *Journal of Logic, Language and Information*, 22(4):391–420.

[Egré and Zehr, 2016] Egré, P. and Zehr, J. (2016). Are gaps preferred to gluts? a closer look at borderline contradictions. Manuscript, under review.

[Fara, 2000] Fara, D. (2000). Shifting sands: an interest-relative theory of vagueness. *Philosophical Topics*, 28(1):45–81. Originally published under the name "Delia Graff".

[Feigenson et al., 2004] Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. *Trends in cognitive sciences*, 8(7):307–314.

[Frazee and Beaver, 2010] Frazee, J. and Beaver, D. (2010). Vagueness is rational under uncertainty. In *Logic, Language and Meaning*, pages 153–162. Springer.

[Fults, 2011] Fults, S. (2011). Vagueness and scales. In *Vagueness and Language Use*, pages 25–50. Palgrave Macmillan.

[Goodman and Lassiter, 2014] Goodman, N. D. and Lassiter, D. (2014). Probabilistic semantics and pragmatics: Uncertainty in language and thought. *Handbook of Contemporary Semantic Theory*.

[Kennedy, 2007] Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, 30(1):1–45.

[Kennedy, 2011] Kennedy, C. (2011). Vagueness and comparison. In Egré, P. and Klinedinst, N., editors, *Vagueness and Language Use*. Palgrave Macmillan.

[Kennedy, 2013] Kennedy, C. (2013). Two sources of subjectivity: Qualitative assessment and dimensional uncertainty. *Inquiry*, 56(2-3):258–277.

[Klein, 1980] Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics and philosophy*, 4(1):1–45.

[Krantz et al., 1971] Krantz, D., Luce, D., Suppes, P., and Tversky, A. (1971). *Foundations of measurement, Vol. I: Additive and polynomial representations*. Dover.

[Kyburg, 2000] Kyburg, A. (2000). When vague sentences inform: A model of assertability. *Synthese*, 124(2):175–191.

[Lassiter, 2011] Lassiter, D. (2011). Vagueness as probabilistic linguistic knowledge. In Nouwen, R., van Rooij, R., and Schmitz, H.-C., editors, *Vagueness in Communication*, pages 127–150. Springer.

[Lassiter and Goodman, 2014] Lassiter, D. and Goodman, N. D. (2014). Context, scale structure, and statistics in the interpretation of positive-form adjectives. In *Proceedings of SALT*, volume 23, pages 587–610.

[Lassiter and Goodman, 2015] Lassiter, D. and Goodman, N. D. (2015). Adjectival vagueness in a Bayesian model of interpretation. *Synthese*. this issue.

[Lawry, 2008] Lawry, J. (2008). Appropriateness measures: an uncertainty model for vague concepts. *Synthese*, 161(2):255–269.

[Luce, 1959] Luce, R. D. (1959). *Individual Choice Behavior*. Dover. Reedition Dover 2005,.

[MacFarlane, 2010] MacFarlane, J. (2010). Fuzzy epistemicism. In Dietz, R. and Moruzzi, S., editors, *Cuts and Clouds: Vagueness, its Nature and its Logic*, pages 438–463. Oxford University Press.

[Maley, 2011] Maley, C. J. (2011). Analog and digital, continuous and discrete. *Philosophical studies*, 155(1):117–131.

[McNicol, 2005] McNicol, D. (2005). *A primer of signal detection theory*. Psychology Press.

[Moyer and Landauer, 1967] Moyer, R. S. and Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, 215:1519–1520.

[Peirce, 1902] Peirce, C. S. (1902). Vague. In Baldwin, J., editor, *Dictionary of Philosophy and Psychology*, volume 2, page 748. New York: Macmillan.

[Qing and Franke, 2014] Qing, C. and Franke, M. (2014). Gradable adjectives, vagueness, and optimal language use: a speaker-oriented model. In *Proceedings of SALT 24*, pages 23–41.

[Ripley, 2011] Ripley, D. (2011). Contradictions at the border. In Nouwen, R., Schmitz, H.-C., and van Rooij, R., editors, *Vagueness in Communication*, pages 169–188.

[Sapir, 1944] Sapir, E. (1944). Grading, a study in semantics. *Philosophy of science*, 11(2):93–116.

[Sassoon, 2010] Sassoon, G. W. (2010). Measurement theory in linguistics. *Synthese*, 174(1):151–180.

[Schiffer, 2003] Schiffer, S. (2003). *The things we mean*. Clarendon Oxford.

[Smith et al., 2003] Smith, J. D., Shields, W. E., and Washburn, D. A. (2003). The comparative psychology of uncertainty monitoring and metacognition. *Behavioral and brain sciences*, 26(03):317–339.

[Smith, 2008] Smith, N. J. J. (2008). *Vagueness and Degrees of Truth*. Oxford University Press, Oxford.

[Solt, 2011a] Solt, S. (2011a). Notes on the comparison class. In *Vagueness in Communication*, pages 189–206. Berlin, Heidelberg: Springer.

[Solt, 2011b] Solt, S. (2011b). Vagueness in quantity: two case studies from a linguistic perspective. In *Understanding Vagueness*, pages 157–174. College publications.

[Solt, 2015] Solt, S. (2015). On quantification and measurement: the case of 'most' and 'more than half'. forthcoming in *Language*.

[Stevens, 1957] Stevens, S. S. (1957). On the psychophysical law. *Psychological review*, 64(3):153.

[Stevens, 1966] Stevens, S. S. (1966). A metric for the social consensus. *Science*, 151(3710):530–541.

[van Rooij, 2011a] van Rooij, R. (2011a). Implicit vs. explicit comparatives. In Egré, P. and Klinedinst, N., editors, *Vagueness and Language Use*, pages 51–72. Palgrave Macmillan.

[van Rooij, 2011b] van Rooij, R. (2011b). Vagueness and linguistics. In Ronzitti, G., editor, *Vagueness: A guide*, pages 123–170. Springer.

[Verheyen et al., 2010] Verheyen, S., Hampton, J. A., and Storms, G. (2010). A probabilistic threshold model: Analyzing semantic categorization data with the Rasch model. *Acta psychologica*, 135(2):216–225.

[Williamson, 1992] Williamson, T. (1992). Vagueness and ignorance. *Proceedings of the Aristotelian Society*, 66:145–162.

[Williamson, 1994] Williamson, T. (1994). *Vagueness*. Routledge, London.

[Wright, 1995] Wright, C. (1995). The epistemic conception of vagueness. *The Southern journal of philosophy*, 33(S1):133–160.