Subjectivity in gradable adjectives: The case of *tall* and *heavy*

Steven Verheyen[a,b*], Sabrina Dewil[b], and Paul Egré[a,c]

[a]Institut Jean Nicod, École Normale Supérieure

[b]Faculty of Psychology and Educational Sciences, University of Leuven

[c]Swedish Collegium for Advanced Study

Abstract

We propose an investigation of the ways in which speakers' subjective perspectives are likely to affect the meaning of gradable adjectives like *tall* or *heavy*. We present the results of a study showing that people tend to use themselves as a yardstick when ascribing these adjectives to human figures of variable measurements: subjects' height and weight requirements for applying *tall* and *heavy* are found to be positively correlated with their personal measurements. We draw more general lessons regarding the definition of subjectivity and the ways in which a standard of comparison and a significant deviation of that standard are specified.

Keywords: subjectivity; egocentrism; threshold theory; ideals; categorization; vagueness.

Word count: 11742

Acknowledgements: Part of this research was conducted while SV was a postdoctoral fellow at the Research Foundation – Flanders (FWO). At the time of writing, SV was funded by ANR project *TriLogMean* (ANR-14-CE30-0010), awarded to Paul Egré and Benjamin Spector. PE was funded by the ANR project *TriLogMean* (ANR-14-CE30-0010) and by a fellowship of the Swedish Collegium for Advanced Study. SV and PE also acknowledge grants ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL* for research conducted at the Department of Cognitive Studies of ENS in Paris. We would like to thank Adam Bear, Heather Burnett, Joshua Knobe, Hans Rott, and two anonymous reviewers for useful comments and suggestions on a previous version of this manuscript.

***Address for correspondence**: Steven Verheyen, École Normale Supérieure, PSL Research University, Institut Jean-Nicod, Pavillon Jardin, 29, rue d'Ulm, 75005 Paris, France.

E-mail: steven.verheyen@ens.fr

Subjectivity in gradable adjectives: The case of *tall* and *heavy*

1. Introduction

Gradable adjectives like *tall* and *heavy* are paradigmatic examples of vague words. While they are concerned with quantifiable properties of instances (height and weight, respectively), they fail to draw a fixed and context-independent line between positive instances and negative instances. Various tests have been proposed in the literature to characterize vague words as such. One common diagnosis concerns the *sorites-susceptibility* of vague words (Keefe, 2000; Wright, 1976 a.o.): intuitively, if a man is a positive instance of *tall*, then anyone shorter by a millimeter ought to be a positive instance too, and similarly, if a person is a positive instance of *heavy*, anyone lighter by one milligram ought to remain a positive instance. Step by step, however, this intuition runs counter to the fact that some people are categorized clearly as not tall, and likewise that some are categorized clearly as not heavy.

Another diagnosis that has been discussed concerns the *subjectivity* of vague words (Kamp and Sassoon, 2014; Kennedy, 2013): the idea is that even as the relevant contextual parameters are fixed, in particular the comparison class (Klein, 1980), some room remains for *permissible variation* (Raffman, 2014) or *faultless disagreement* between people (Kölbel, 2004; Wright, 1995). For instance, when asked whether a person measuring 1.80 m is tall or not for a male adult, people will respond differently. While some may require a man to be at least 1.80 m to be called tall, others may require him to be at least 1.90 m. Likewise, when asked whether a person weighing 80 kg counts as heavy or not for a male adult, responses will differ from subject to subject. One linguistic reflect of that subjectivity is exemplified in the find-test discussed by Sæbø and Kennedy among others: 'Anna finds John heavy' is an acceptable sentence, whereas 'Anna finds the number 2 prime' sounds odd and inappropriate (Kennedy, 2013; Sæbø, 2009).

Subjectivity may be understood as a subtle form of context-sensitivity: Anna may represent a male adult weighing 80 kg as heavy if she imagines that person to (typically) measure 1.60 m, whereas Susan may represent the same male adult weighing 80 kg as not heavy if she imagines him to (typically) measure 1.90 m. In such a case, Anna and Susan actually use different standards of comparison or reference points: in order to judge whether someone is heavy nor not, 'male adult' is not enough, but further restrictions like the ones mentioned can be made and remain at the speaker's discretion. Arguably, however, such cases are not genuine cases of faultless disagreement (as opposed to Anna and Susan disagreeing as to whether sardines are tasty; Kölbel, 2004). For Anna and Susan's disagreement to be faultless, it ought to be the case that they can disagree even as *all* contextual parameters have been made explicit and common to them. Practically, however, specifying all contextual parameters is impossible. Even as sex and age and height have been fixed, one could imagine ethnicity, say, to provide a further relevant parameter, which Anna and Susan can set in different ways. In principle, however, we could imagine two persons like Anna and Susan to have the same standards of comparison in mind, but still to differ on where to set their inner thresholds for *tall* or *heavy*. This would happen if Anna and Susan have different subjective appraisals of what counts as a significant difference relative to the same reference point.

Fara (2000, 2008) argues that the predicate *tall* shows a principled relativity to the ascriber's practical interests, even when we specify an explicit comparison class by means of a *for*-phrase as in 'tall for a mountain', or 'tall for a cherry tree'. Unlike the unmodified 'tall', 'tall for a cherry tree' is not self-applicable by a human judge, so in principle, one might argue that the notion of subjective relativity in this case is eliminable (see Stanley, 2003, for an argument along those lines). For Fara, however, 'tall for a cherry tree' is to be analysed as meaning: '*significantly* taller *to me* than is the norm for a cherry tree', where what counts as a significant difference relative to the reference point remains subjective, even if the reference point is not the subject herself in this case (see also Egré, 2016).

Positive forms of gradable adjectives like *tall* are thus predicated relative to a *reference class* (or comparison class, see Klein, 1980; Rips and Turnbull, 1980), relative to a *reference point* (or point of departure, see Sapir, 1944, or standard of comparison, see Kennedy, 2007), but also relative to a *perspective*. Those three parameters are made explicit in the following sentence:

1. I find this tree tall for a cherry tree, compared to the average cherry tree in my mum's garden.

Here 'I find' indicates the perspective, 'for a cherry tree' specifies the reference class, and 'compared to the average cherry tree in my mum's garden' the reference point or standard of comparison. The perspective needs to be distinguished because as hinted above, in principle two persons could have selected the same reference class and reference point for their comparison, but still differ in what they consider to be a significant deviation above or below that point (Egré, 2016; Fara, 2000). Assuming a specified reference class, this means the subjectivity of an agent can influence both the selection of a standard of comparison, and what counts as a significant deviation from the standard.

We entertain the hypothesis that subjectivity is not mere randomness, but rather that the standards of comparison used for the ascription of a vague predicate, like the deviations from those standards that are regarded significant, can to some extent be predicted from properties that are specific to the subject. In this paper, our goal therefore is to provide evidence for such a *systematic* relationship in the case of the adjectives *tall* and *heavy*. For those adjectives, we are interested in the way in which actual ascriptions of those predicates are sensitive to what we call *egocentric indices*. Such indices concern the judge's own standing along the relevant dimensions that underlie the predicates, such as height and weight. That is, we expect taller (heavier) subjects to entertain taller (heavier) standards of comparison and/or to require greater deviations from these standards than shorter (lighter) subjects do.

Obviously, not all vague predicates are predicates for which we may easily detect an influence of egocentric indices, simply because not all vague predicates are necessarily self-applicable by the person using those. However, the fact that such predicates are not self-applicable does not necessarily imply that they do not depend on some egocentric frame of reference.[1] Our main aim in this paper is to show that for a number of predicates that are in principle self-applicable, we do see an effect of egocentricity. In what follows, we start with a brief review of the earlier literature on the manifestation of egocentric references in the ascription of various properties (section 2). We then go on to present a study in which we tested the influence of the egocentric indices height and weight on the ascription of the predicates *tall* and *heavy* (section 3). In section 4, finally, we draw more general lessons regarding the notion of subjectivity in relation to vague predicates.

2. Egocentricity in Comparison: A Brief Review

A number of studies have shown a relativity to human standards of judgments involving gradable adjectives such as *tall*. Suzuki (1970) gives several examples involving out-of-the-blue generic sentences to illustrate the default relativity of *tall* to a human standard of comparison for size. One example inspired by his is the following:

2. Giraffes are tall.
3. ? People are tall.

The difference is predicted if in both sentences human heights are the standards of comparison for *tall*. Giraffes are tall compared to the average human, but generically, people cannot be tall compared to the average human. Hence, in spontaneous speech *tall* will rarely be combined with

[1] Fara (2008) discusses how personal *interests* (as opposed to personal *properties*) may yield subjectivity in predicates that are not self-applicable.

a noun referring to the human body, but instead is used to refer to taller-than-human objects (Goy, 2002; Tribushinina, 2008; Vogel, 2004). This is taken to suggest that the prototypical use of *tall* implicitly refers to the human body as a standard of comparison.

Similarly, Tribushinina (2008) found that in corpora *tall* is used with the human body as a sort of ruler, both comparatively ('The pot was taller than a man'; 'And they had a tent with them. A small one, shorter than a man') and with the positive form ('Right on the path there was a high cross, as tall as two humans, wrapped with straw and rags'; 'Big juicy berries looked like grapes growing on stems as tall as a man.'). Rips and Turnbull (1980) also point out that when used predicatively, adjectives are verified faster if they exceed not only a standard value within the comparison class but also the anthropomorphic standard ('horses are tall' < 'roses are tall').

However, these cases do not provide an obvious illustration of the effect of subjectivity proper, as much as a way of defaulting the reference point needed to apply the positive form of *tall* to a specifically human standard. In other words, they illustrate some relativity to *human* standards (a form of anthropocentric relativity), more than relativity to *subjective* standards (namely to egocentric indices). There is also ample evidence that people use themselves and their own bodies as reference points for judging properties such as the height, weight, and age of others, however. These observations can be seen as manifestations of *embodied cognition*, the theory that one's cognition (including concepts and categories, language use, and judgments) is shaped by aspects of one's body (Valera, Thompson, and Rosch, 1991). Arguably, this behaviour is also related to the manifest asymmetry people display in comparisons in which they are themselves involved. They compare others to themselves, rather than the other way around (a behaviour sometimes referred to as *egocentricity bias*). People for instance believe that group judgments are better predicated from their own judgments than vice versa (Kunda and Nisbett, 1988) and judge others to be more similar to themselves, than themselves to others (Holyoak and Gordon, 1983). In doing so, they are explicitly considering themselves a reference point.

Egocentricity has been documented in height and weight estimates of the average man and woman. Hinckley and Rethlingshafer (1951) were the first to establish that shorter men estimate the average height of men to be smaller than taller men do. A study by Ward (1967) replicated this finding for women as well: He established a significant correlation between men's height and their judgments of the average man's height and between women's height and their judgments of the average woman's height. In Fillenbaum (1961) men and women were also shown to use their own weight as a reference point when estimating the average weight of members of their own sex. Both Ward (1967) and Fillenbaum (1961) report the relationship weakens or even disappears when subjects estimate the average value of the opposite sex.

Egocentricity has also been established in subjects' estimates of individuals' age, height, and weight. Mintz (1956) found that children's estimates of Peter Pan's age correlated with their own age. The relationship has also been found in adults estimating the age of facial pictures (Vestlund, Langeborg, Sörqvist, and Eriksson, 2009; Voelkle, Ebner, Lindenberger, and Riediger, 2012) and of individuals they just met (Sörqvist, Langeborg, and Eriksson, 2011). This assimilation of estimates of others toward one's own value has also been documented for height and weight estimates of pictured and actual persons (Flin and Shepherd, 1986; Sörqvist, Langeborg, and Eriksson, 2011). It occurs regardless of whether the judgments are made on an objective scale (years, inches, pounds) or on a subjective rating scale (Rethlingshafer and Hinckley, 1963). As was the case for estimates of the average man and woman, egocentricity in estimates for individuals is found to be stronger within than across sex (Flin and Shepherd, 1986). The results in Sörqvist, Langeborg, and Eriksson (2011) suggest the possibility that women assimilate across sex, whereas men do not.

The study that is perhaps most related to the one we will conduct is that by Dunning and Cohen (1992), who asked their participants what height a person should have to be considered tall. Seventy-two percent of their participants provided a minimum value above their own height. The evidence with respect to the egocentricity of the provided minima was mixed. Dunning and Cohen found a significant correlation with own height among the men, but not among the women. In a

replication by Dunning and McElwee (1995) the relationship was absent. Dunning and Cohen themselves called the task they employed 'unnatural' because people are seldom asked to determine what value along a dimension is necessary to deserve a particular label. We will therefore employ a more natural task in which participants have to judge whether a particular label applies or not to items along a continuum, instead of having to represent in the abstract where they would assign a boundary.

Our target will be ascriptions of sentences of the form 'x is tall/ heavy', in which the reference class is made manifest both in our stimuli (female vs. male figures) and in the questions ('Do you find this woman/man tall/heavy?'). The reference point and what counts as a significant deviation from it will each time be left at the speaker's discretion. The question for us is the extent to which judgments of the form 'x is tall/heavy' are likely to be made relative to the utterer and his or her egocentric indices. This is the sense in which this paper is about subjectivity in judgments involving gradable adjectives. Based on the review of the literature above, we expect to observe egocentricity when people judge whether the predicates *tall* and *heavy* apply to stimuli representing members of their own sex. We expect to find less evidence for egocentricity when stimuli representing members of the opposite sex are to be judged. That is, we expect to find the most compelling evidence for egocentricity when utterers can include themselves in the reference class.

3. A Study of Egocentrism and Vagueness

To investigate whether there is evidence for egocentricity in vagueness, we conducted a study in which male and female figures were to be judged as tall or not and heavy or not. For the *tall* judgments, figures of different heights were depicted against a background indicating their height.

For the *heavy* judgments, the figures were body contours corresponding to different body weights. After completing both sets of judgments, participants indicated their own height and weight. Both the *tall* and the *heavy* judgments allow us to investigate egocentricity in the domain of vagueness by relating participants' height to the threshold they employ for distinguishing stimuli that are tall from stimuli that are not and by relating participants' weight to the threshold they employ for distinguishing stimuli that are heavy from stimuli that are not. Since both male and female figures are presented we can compare cases in which the utterers can include themselves in the reference class (same sex judgments) to cases in which they cannot (different sex judgments).

3.1. Participants

We recruited 582 volunteers via social media. The data of 245 participants were not used because they did not meet one or more of four criteria.

The first criterion required participants to have Dutch as L1, the language in which the study was conducted. The second criterion required participants to have Belgian nationality, the country in which the study was conducted. There are regional differences in physical characteristics such as height and weight (Heine, 2008) and the interpretation of gradable adjectives is known to be language- and culture-specific (Reardon and Miller, 2011). Together, these restrictions ensure that we do not mistake any of these differences for egocentricity.

The third criterion required participants to be aged 17-29. People are known to overestimate their height and to underestimate their weight (Ezzati, Martin, Skjold, Vander Hoorn, and Murray, 2006). This is not an issue for this study, however, as what people *believe* to be their own height/weight (rather than their *actual* height/weight) is more likely to influence *tall* and *heavy* judgments. We nevertheless decided to restrict the age range to the most stable period in terms of height and weight. At the age of 17 most people are fully grown and by 29 they have not yet begun to shrink or gain considerable weight. The restriction is

intended to avoid noise due to age-related perception biases (Ezzati et al., 2006), and makes for a more homogeneous participant sample, reducing the influence of nuisance variables.

The fourth criterion entailed an automated quality check to ensure participants' responses were informed by the stimuli's underlying dimensions (height, weight) and were not random (for details about the procedure see Verheyen, Voorspoels, and Storms, 2015). In all likelihood, participants who failed this check did not understand the instructions or did not take the task seriously.

Table 1

Mean Age (in years), Height (in cm), Reported Weight (in kg), Perceived and Ideal Weight (in % of average body size) of the participants. Standard deviations are provided between brackets below the mean values.

Sample	**Age**	**Height**	**Weight**		
	Reported	**Reported**	**Reported**	**Perceived**	**Ideal**
Women	21.88	167.15	63.73	97.24	83.25
(N=183)	(2.55)	(6.75)	(12.03)	(19.21)	(12.32)
Men	21.97	181.44	75.56	88.93	89.38
(N=154)	(2.68)	(6.65)	(11.38)	(16.36)	(12.68)

Table 1 summarizes the characteristics of the participants who met all four criteria, separated according to sex. We report the means and standard deviations (between brackets) for Age (in years), Height (in cm), Reported Weight (in kg), Perceived and Ideal Weight (in % of average body size). While the values for the first three variables were obtained through open questions, the values for the latter two variables reflect participants' choices of their

perceived and ideal weight from among images reflecting various percentages of the average body size for their own sex (see below for details).

The majority of the participants were female (54%). Participants of both sexes experienced a discrepancy between their Perceived Weight and their Ideal Weight. Of the female participants 82% indicated that they would like to be slimmer, resulting in a lower average value for the Ideal Weight than for the Perceived Weight. Among men the average Perceived and Ideal Weight were almost the same. However, only 15% of the men gave identical values for Perceived and Ideal Weight (compared to 10% of women). The male participants either wanted to lose weight (43%) or wanted to gain weight (42%).

3.2. Materials

We studied the predicates *tall* and *heavy* (*groot* and *zwaar* in Dutch) because they are prototypical examples of vague gradable adjectives that one could self-apply. They have the additional advantage that it is feasible to construct meaningful one-dimensional stimuli participants can judge for tallness and heaviness.

Participants judged the tallness of 17 pink female pictograms and 17 blue male pictograms depicting women and men ranging from 1m35 till 2m15 in steps of 5 cm. The stimuli were depicted against a background indicating their height (see Alxatib and Pelletier, 2011 for similar stimuli). The aspect ratio of the stimuli was kept constant. The upper rows of Figure 1 depict the female and male stimuli at the lower end (1m35), halfway (1m75), and at the upper end of the range (2m15). For female stimuli the accompanying question read '*Do you find this woman tall?*'. For male stimuli the question read '*Do you find this man tall?*'.

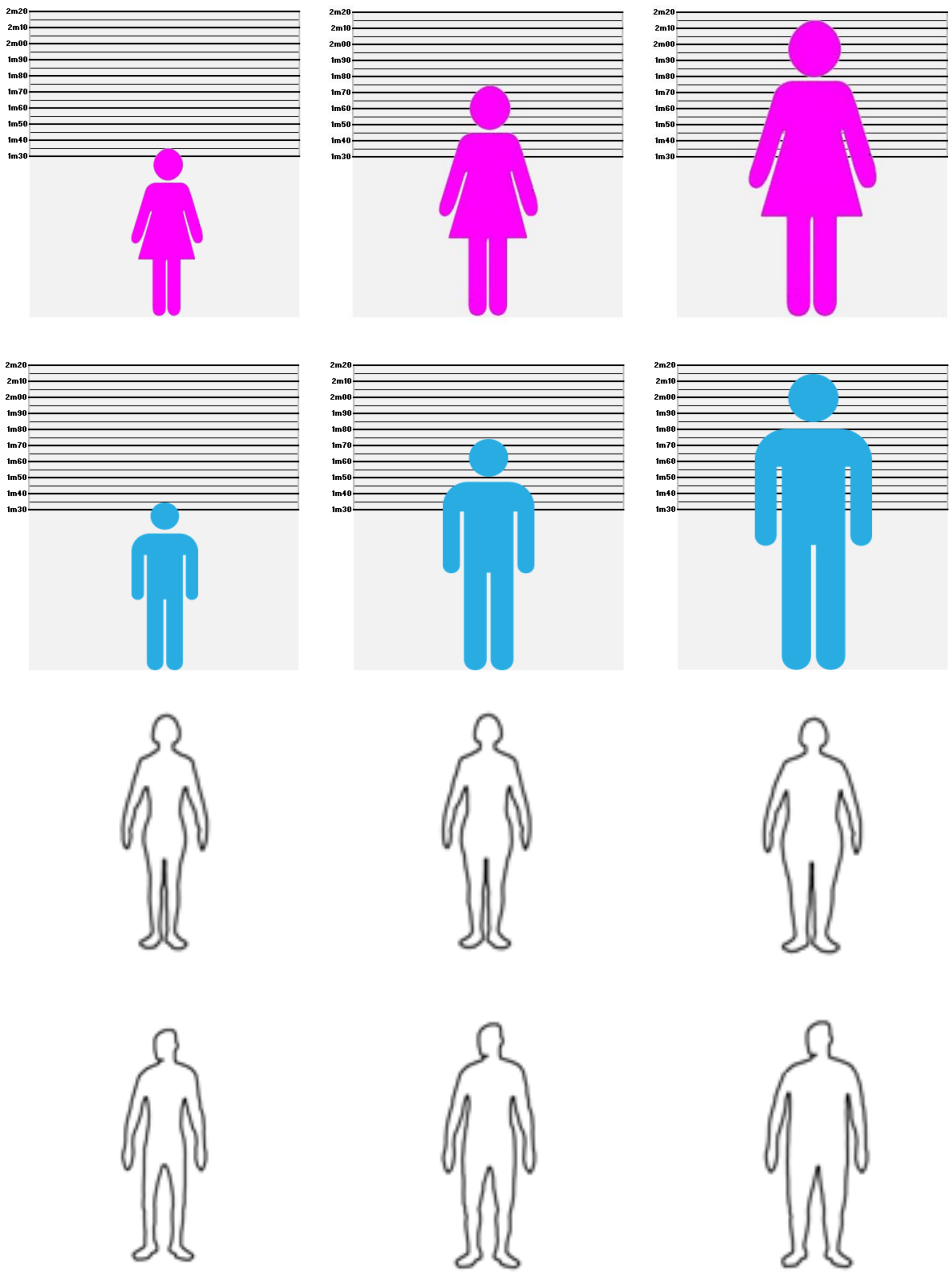

Figure 1. Examples of stimuli that were judged for tallness (upper two rows) and heaviness (lower two rows). The stimuli on the left are located at the lower end of the range of presented stimuli. The stimuli in the middle are located halfway and the stimuli on the right are located at the upper end of the range.

Participants also judged the heaviness of 17 female and 17 male contour-line drawings representing women and men with body weights ranging from 60% to 140% of the known average in steps of 5%. The stimuli were taken from a body-image assessment tool and are based on known

anthropometric body dimensions of shoulder, chest, waist, hip breadth, thigh breadth, and upper leg breadth (Gardner, Jappe, and Gardner, 2009). The lower rows of Figure 1 depict the average female and male stimuli (100%) surrounded by the minimal (60%) and maximal (140%) stimuli. For female stimuli the accompanying question read '*Do you find this woman heavy?*'. For male stimuli the question read '*Do you find this man heavy?*'. Note that no indication of the precise weight was included in the *heavy* stimuli, contrary to the specification of height in the stimuli used for *tall*. This was a deliberate choice meant to represent both a situation in which *explicit* information to base one's judgment on is available, and the perhaps more common situation in which the (quantitative) basis for one's judgment needs to be *inferred* from the available perceptual input.

3.3. Procedure

Participants performed the study online through the survey software tool Qualtrics (www.qualtrics.com). The participants decided for 2x17 stimuli whether they were tall or not and for 2x17 stimuli whether they were heavy or not. That is, participants made categorization judgments for stimuli of their own sex and of the opposite sex. All participants completed the tallness judgment task before they completed the heaviness judgment task. Whether the female or the male figures were presented first was randomly determined for every participant and predicate. All 17 stimuli of a specific gender were presented individually in a random order. Upon answering whether they considered a particular stimulus tall or not (for the stimuli differing in height) or heavy or not (for the stimuli differing in weight) the next stimulus was automatically presented. Participants could go back to previous screens to correct mistakes.

After completing the categorization tasks, participants provided a number of demographic variables: sex, age, L1, country of origin, height, and weight. The participants also indicated their perceived and ideal weight on the body-image assessment tool from which we drew the stimuli

for the heavy judgments. To this end the 17 contour-line drawings corresponding to their own sex were shown simultaneously in ascending order.

3.4. Analyses

Earlier we indicated that subjectivity is a diagnostic of vague words. Even if all the relevant contextual parameters are fixed, there remains more than one way to use them in a competent manner (Kölbel, 2002; Raffman, 2014; Wright, 1995). In highly constrained categorization tasks like the ones we employed, where there is only one dimension along which the stimuli differ (height or weight) this subjectivity is expected to show in the extension of the predicates. While some participants may consider many stimuli to be tall/heavy, others might find that these predicates only apply to a few of the stimuli. If there were to be a systematic relationship between one's height (weight) and the extension of the predicate *tall* (*heavy*), this would constitute evidence for egocentricity.

Most accounts of vagueness would expect the response patterns of individual participants to display a so-called Guttman pattern or *monotonicity* (Guttman, 1944): a cut-off point is situated along the dimension underlying the stimuli (height, weight) prior to which the predicate is consistently denied and after which the predicate is consistently applied.[2] Even for stimuli like ours that vary only along one dimension (height or weight), violations of monotonicity are observed, however (Douven, Wenmackers, Jraissati, and Decock, 2016; Verheyen and Egré, 2017). That is, subsequent stimuli receive alternating responses. Moreover, inconsistent answers are frequently produced when these kinds of judgments are repeated (Egré, de Gardelle, and Ripley, 2013; Hersh and Caramazza, 1976). Subjects have also been shown to experience difficulty when deciding membership for one-dimensional borderline stimuli (i.e., produce longer RTs and lower confidence ratings, Brownell and Caramazza, 1978; Hersh and Caramazza, 1976) or do not decide at all, and both apply and deny the predicate (Alxatib and Pelletier, 2011; Egré and Zehr, 2018; Ripley, 2011).

[2] See also Fine (1975), who includes this monotonicity constraint under what he calls *penumbral connections*.

These findings have spurred the development of probabilistic accounts of vagueness in which the decision to apply a predicate to a stimulus or not includes an element of chance (Egré, 2016; Verheyen, Hampton, and Storms, 2010).

In order to meet the presence of violations of monotonicity in our data, we analysed it using a probabilistic formalization of the threshold theory (Verheyen, Hampton, and Storms, 2010). The threshold theory (Hampton, 1998, 2007) equates subjectivity with the use of different thresholds (see also Egré, 2016). Extension differences arise because individuals employ different thresholds. Each individual is believed to place her own threshold on the dimension along which the items differ (height, weight).[3] This threshold functions as a deterministic value that rigorously separates the stimuli to which the predicate applies from the stimuli to which it does not. The predicates *tall* and *heavy* in principle only apply to the stimuli that surpass the threshold in height and weight, respectively. Verheyen, Hampton, and Storms (2010) provided a probabilistic formulation of the theory by positing that the decision to apply the predicate to the stimulus is informed by the *distance* between the stimulus and the threshold along the dimension. Greater distances make for more certain outcomes, while smaller distances make for uncertain outcomes. A predicate is thus decidedly applied to stimuli that clearly surpass the threshold and decidedly denied of stimuli that clearly fall short of the threshold. The predicate is just as likely to be applied as to be denied for stimuli that are indiscernible from the threshold.

Formally, the probability of applying the predicate is represented by a logistic function of the difference between β_s and θ_i, where β_s represents the position of stimulus *s* on the underlying dimension (its height or weight) and θ_i represents the threshold individual *i* employs for categorization:

[3] In Verheyen, Hampton, and Storms (2010), and in Egré (2016), the threshold is called a *criterion*, following standard terminology in signal detection theory. In this paper, we prefer to talk only of *threshold*, to avoid any confusion between the notion of criterion in that sense, and the notion of criterion understood as a respect of comparison.

$$\Pr(Y_{is} = 1) = \frac{e^{\beta_s - \theta_i}}{1 + e^{\beta_s - \theta_i}}. \quad (1)$$

Both β_s and θ_i are free parameters that are estimated from the categorization data. They should therefore be interpreted in light of the task that is performed. The underlying dimension is thus psychological in nature. It does not necessarily reflect the stimuli's physical properties but is scaled with respect to the individuals' judgments. For instance, stimuli at the ends of the dimension will be positioned closer to each other than in the middle of the dimension, because only few participants will distinguish them in their tallness/heaviness judgments. Similarly, θ_i is the point of subjective equality: the position on the underlying dimension for which individual *i* is indecisive as to whether the predicate should apply or not. In the probabilistic formulation of the threshold theory the threshold thus no longer acts as a sharp boundary. For the purposes of our study, this allows one to estimate the thresholds in the presence of violations of monotonicity. The use of categorization thresholds also allows one to study egocentricity in vagueness without having to commit that it operates on either standards or deviations. We hypothesized that the higher one's own measurements, the higher the standard of comparison and/or the significant deviation one would entertain. Since the standard and the deviation jointly determine the threshold that separates negative instances from positive ones, we can evaluate the egocentricity hypothesis by relating participants' egocentricity indices to their thresholds. We do not know of a model that allows categorization data to be decomposed in standards and deviations. We therefore prefer to employ a model that allows one to estimate their compound, the categorization threshold, rather than to use a model that assumes fixed standards and subjective deviations (or vice versa).

The free parameters of the model in Equation (1) were estimated using WinBUGS (Lunn, Thomas, Best, and Spiegelhalter, 2000) according to the details and code provided in Verheyen, Voorspoels, and Storms (2015). For every analysis three chains were run of 10,000 iterations each,

with a burn-in sample of 4,000. Separate analyses were conducted on the data for male and female stimuli, because different requirements for *tall* and *heavy* might apply to men and women since they differ in their average height and weight (Biernat, Manis, and Nelson, 1991). The data of male and female respondents were also analysed separately, to allow for a comparison of judgments in which participants could and could not include themselves in the reference class. This resulted in eight analyses (2 predicates x 2 sexes stimuli x 2 sexes participants).

3.5. Results

3.5.1. Signatures of Subjectivity

Figure 2 depicts the proportion of participants applying the predicates *tall* and *heavy* as a function of the height and weight of the employed stimuli. It clearly shows variation or disagreement: especially for the stimuli in the middle of the employed range, the categorization proportions take on values between 0 and 1, indicating that some participants considered these stimuli tall (heavy) whereas others did not.

The categorization proportions have a distinctive shape. They start off at 0, indicating that all participants refuse to apply the predicate *tall* (*heavy*) to the stimuli at the lower end of the height (weight) range. About half-way the range, the categorization proportions start to increase, rising eventually to categorization proportions near 1, indicating that all participants agree to apply the predicate *tall* (*heavy*) to the stimuli at the upper end of the height (weight) range. The transition occurs gradually, resulting in S-shaped curves rather than discrete threshold functions that abruptly 'jump' from 0 to 1. These S-shaped curves are very consistently observed in the use of vague predicates (e.g. Egré et al., 2013; Hampton, 1998,

2007; Verheyen et al., 2010). They signal that in otherwise identical conditions, the participants use different thresholds to distinguish non-tall (non-heavy) persons from tall (heavy) ones.

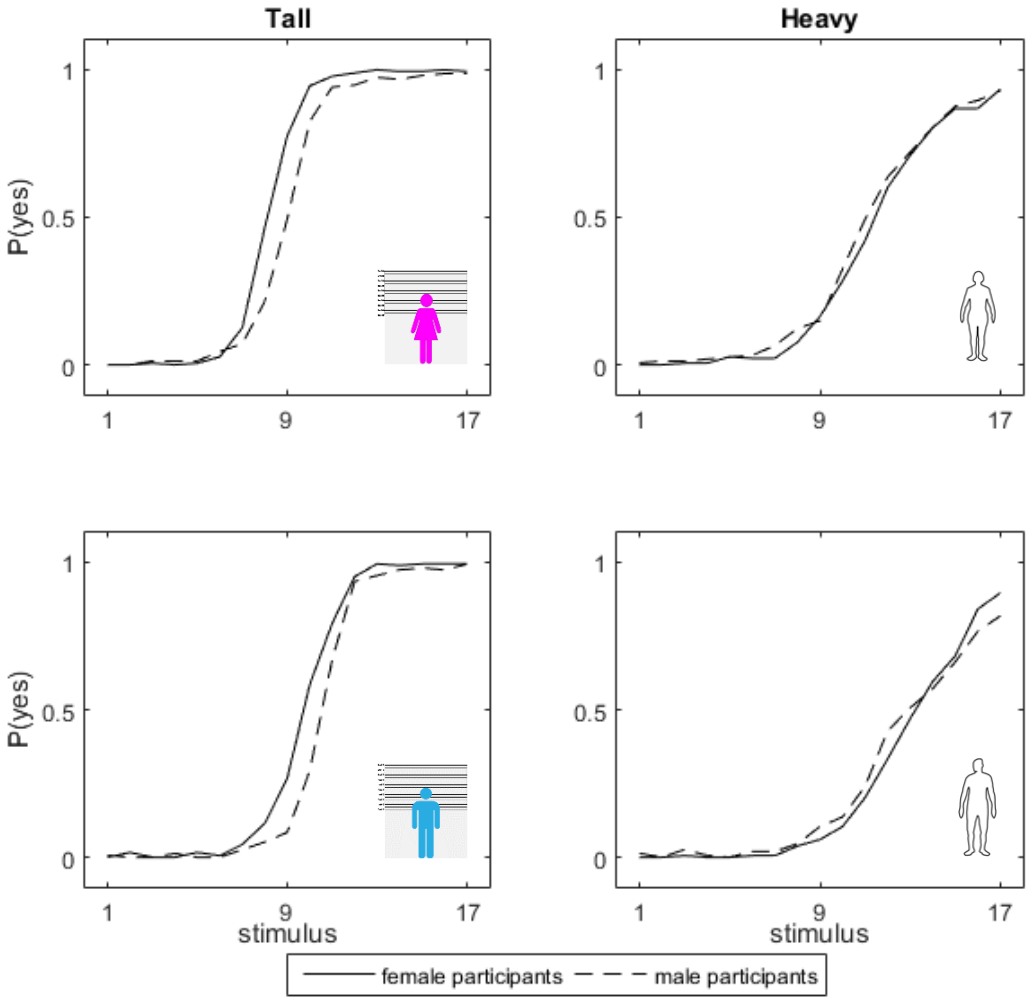

Figure 2. Proportion of female (solid) and male (dotted) participants applying the predicate *tall* (left panels) and the predicate *heavy* (right panels) to stimuli of increasing height/weight. The upper panels pertain to the female stimuli. The lower panels pertain to the male stimuli.

The S-shaped curves in Figure 2 are not just the result of imposing differently located sharp threshold functions on top of each other. We determined the percentage of participants who violated monotonicity in their judgments. When a participant's judgments displayed more

than one response change, we counted it as a violation of monotonicity. The results in Table 2 indicate that even when participants had the means to be consistent because the stimuli they judged contained a specific indication of height (*tall* stimuli), about 9% of the participants violate monotonicity. When no such indication was present and participants had to infer the stimuli's weight (*heavy* stimuli) in order to judge heaviness, this percentage increases to 38%. These violations also show in that the curves in Figure 2 are not all monotonically increasing. It is thus not the case that each individual participant uses a deterministic threshold function to distinguish non-tall (non-heavy) persons from tall (heavy one) ones. Practically, these violations support the use of a probabilistic model to establish the participants' thresholds, as for a considerable proportion of the participants these thresholds do not present themselves as sharp cut-off points.

Table 2

Percentage of participants violating monotonicity

Participants	***Tall* stimuli**		***Heavy* stimuli**	
	Female	**Male**	**Female**	**Male**
Female	.08	.07	.34	.33
Male	.10	.12	.44	.44

The data not only reveal the vague/subjective nature of the predicates *tall* and *heavy*, but suggest a number of other insights as well. We briefly go into these in order to avoid any misunderstanding, but do not expand on them as they do not constitute the main topic of this paper.

First, the curves for *tall* are steeper than the curves for *heavy*, and while the former asymptote to 1, the latter do not. This is NOT an indication that the predicate *tall* is less vague semantically than the predicate *heavy* is; rather it reflects the stimulus sampling.[4] While the employed stimulus range for *tall* includes all of the borderline region, the range of *heavy* stimuli does not. Unlike for *tall*, the upper end stimuli for *heavy* were not judged heavy by all participants. This gives the impression of a narrow borderline region for *tall* opposed to a wide one for *heavy*, but the reverse effect could have been obtained if we had sampled more extensively from the borderline-tall region and less extensively from the borderline-heavy region. While our stimulus sampling occurred on the basis of known distributions of heights (for *tall*) and a widely used body-image assessment tool (for *heavy*), it might have worked against our test of egocentricity. The correlations we report in the next section between the thresholds people used and their egocentric indices might have been reduced, because of a restricted threshold range. The correlations we report might increase if we obtained more diverse thresholds, by sampling more extensively from the borderline region for *tall* and by including more extreme stimuli for *heavy*.

Second, the participants judged both male and female stimuli. The curves for male stimuli (lower panels) are shifted to the right compared to the curves for the female stimuli (upper panels) indicating that the height (weight) requirements for men to be called tall (heavy) are higher than those for women. The contention that this reflects a 'correction' for the fact that men are on average taller and heavier than women holds for the *tall* stimuli but not the *heavy* stimuli. Both the male and female *tall* stimuli varied on an absolute scale (height in m) between 1m35 and 2m15. The higher requirement for *tall* may then reflect the fact that the average height for men surpasses that of women. The *heavy* stimuli, on the other hand, varied along a relative scale (% of average body weight). The fact that we observe a threshold

[4] See Burnett (2016) for a review of the linguistic tests that would show that *tall* and *heavy* fall in the same class of relative gradable adjectives.

difference for male and female stimuli here may reflect different beauty standards for women than for men. While 80% of the participants consider a woman with 125% of the average body weight heavy, only 58% of the same participants consider a man with the same percentage of average body weight heavy. The difference also suggests that the participants employed the comparison class 'women' for the female stimuli and the comparison class 'men' for the male stimuli. If the participants had used 'humans' as the comparison class, one would not expect a difference for the male and female stimuli.

A final observation pertains to the fact that for *tall*, but not for *heavy*, the curve for the male participants is shifted to the right with respect to the curve for the female participants. That is, for the same stimuli (regardless of whether these are male or female) the male participants impose a higher height requirement for *tall* than the female participants do. We may take this as an indication of egocentricity already, namely as an expression of the bias imposed by bodily characteristics determined by sex differences.[5]

3.5.2. Signatures of Egocentricity

Below we report the correlation between the posterior mean of the categorization thresholds θ_i and the corresponding height or weight (reported, perceived, idealized) of the individuals making the *tall* and *heavy* judgments, respectively. The results are presented in two tables. Table 3 pertains to judgments made toward stimuli of one's own sex. Table 4 pertains to judgments made toward stimuli of the opposite sex. The evidence in favour of a positive relationship between categorization thresholds and egocentric indices is reflected in the corresponding Bayes Factor

[5] An anonymous reviewer pointed out that this finding could also be attributed to (sociological) gender rather than (biological) sex. The shift in the curve might be an expression of individuals' socialization as men or women (e.g., the tendency in heterosexual couples for the man to be taller than the woman) rather than of an objective, physical difference in height (see Goffman, 1977).

(BF). For ease of discussion we will use the verbal labels *moderate* (3 < BF < 10), *strong* (10 < BF < 30), *very strong* (30-100), and *extreme* (BF > 100) (Wetzels and Wagenmakers, 2012).

Table 3

Correlations between thresholds and egocentric indices, within-sex ratings

Participants	Stimuli	Height	Weight		
		Reported	Reported	Perceived	Ideal
Female	Female	.21**	.22**	.21*	.24***
Male	Male	.31****	.05	.13	.12

Note. All tests one-tailed, for positive correlation

** BF_{+o} > 3 (moderate), ** BF_{+o} > 10 (strong), *** BF_{+o} > 30 (very strong), **** BF_{+o} > 100 (extreme)*

Female participants who judge whether female stimuli are tall or not and heavy or not show evidence of egocentricity in that the thresholds they employ in their judgments are related to their height and weight. Thus, the taller the respondent is, the greater the height she requires to call a female figure tall. On average, the threshold or point of subjective equality for *tall* is located 3.48 cm (*SD* = 6.99) above one's own height. Similarly, female respondents used a higher threshold for calling a female figure heavy, the greater their own weight. The latter relationship holds regardless of whether weight is reported in kilograms or whether one indicates one's perceived or idealized weight on a body-image assessment tool. The strongest support for the relationship is found based on the ideal weight. The threshold for *heavy* is located about 16.01% (*SD*=20.38) above women's perceived body weight and 30.00% (*SD*=14.82) above their ideal body weight. In conclusion:

Women judging whether predicates apply to figures of a reference class they can identify with do so relative to themselves.

Male participants show evidence of egocentricity when judging male figures for height, but not for weight. The correlation between men's height and the thresholds they used for categorizing male figures as tall or not, strongly supported the egocentricity hypothesis. The male respondents on average positioned their threshold 2.10 cm above their own height (*SD*=6.79). The correlations between the threshold for *heavy* and the various weight measures did not support egocentricity, however. In an attempt to explain why this might be the case, we carried out a post-hoc analysis in which we computed these correlations for men who indicated they wanted to gain weight and for men who indicated they wanted to lose weight separately. Unlike the female participants, who expressed an almost uniform desire to lose weight, the sample of male participants comprised two distinct subgroups (see section 3.1). By conducting an analysis for the sample as a whole, we may therefore have obtained a result that is not representative for any of the subgroups. The results of the separate analyses suggest that this might have been the case for the correlation with the ideal weight since we observe an increase in its magnitude in the separate analyses. Among those men who expressed a desire to lose weight (N=65) we established a correlation of the categorization threshold for *heavy* of -.04 with reported weight, of .20 with perceived weight, and of .21 with ideal weight. These men established the threshold for *heavy* on average 23.26% (SD=17.11) above their perceived body weight and 35.91% (SD=15.24) above their ideal body weight. Among men who expressed a desire to gain weight (N=66) these correlations measured .01, .05, and .19, respectively. The latter group established their threshold for *heavy* on average 41.62% (SD=15.26) above their perceived body weight and 27.69% (SD=14.79) above their ideal body weight. Note that the relative position of Ideal and Perceived are reversed in this group compared to the female sample and the other male subgroup whose Ideal was slimmer than Perceived. While the pattern for the two male subgroups is thought provoking, the correlations only provide anecdotal evidence in favor of the egocentricity hypothesis (BF < 3). The pattern does suggest that there is merit in

performing separate analyses for three subgroups of participants (Perceived=Ideal; Perceived < Ideal; Perceived > Ideal) provided sufficiently large sample sizes can be obtained.[6]

Table 4

Correlations between thresholds and egocentric indices, across-sex ratings

Participants	Target	Height	Weight		
		Reported	Reported	Perceived	Ideal
Female	Male	.21**	.07	.14	.18*
Male	Female	.12	.10	.18	.24**

Note. All tests one-tailed, for positive correlation

** BF_{+o} > 3 (moderate), ** BF_{+o} > 10 (strong), *** BF_{+o} > 30 (very strong), **** BF_{+o} > 100 (extreme)*

The correlations for between-sex ratings in Table 4 paint a much less clear picture than the within-sex ratings in Table 3 do. For instance, while in Table 3 the threshold for *tall* was always predictable from the respondents' heights, in Table 4 this appears only to hold for female participants judging the tallness of men, but not for male participants judging the tallness of women. And while in Table 3 reported, perceived, and ideal weight always showed a similar relationship with the categorization threshold for *heavy*, they do not in Table 4, despite considerable correlations among the various weight

[6] Since the female sample is almost entirely comprised of participants of the latter subgroup, excluding the other subgroups from the analyses does not affect the results. This composition of the female participant sample also points to the difficulty of obtaining sufficiently large subsamples.

measures. Only the ideal weight consistently shows a relationship with the categorization threshold for *heavy*, suggesting that the slenderer one's ideal, the more to the left of the weight dimension one's threshold for calling individuals of the other sex is positioned. This relationship is also supported in one of the male subgroups: the correlation between the threshold for *heavy* and ideal is .31 among men who want to lose weight (BF=6.45, one-tailed, N=66) and .22 among men who want to gain weight (BF=1.36, one-tailed, N=65).

4. General Discussion

For a range of gradable adjectives expressing sensory attributes (*cold*, *bright*, *loud*) pertaining to one's immediate experience ('this is cold/bright/loud'), it appears obvious that their application should depend on the context or circumstances of the speaker. Thus, the same water can be cold to one hand and warm to another, depending on the adaptation level of each hand. The same light can appear bright to someone and not to another depending on whether it is seen after darkness or similar light, and the same sound can be perceived as loud or not depending on the difference with the sound level of one's perceiving environment (see Helson, 1947; Kahneman and Tversky, 1979). These examples illustrate that subjective judgments are necessarily relative to the prevailing norm. Our results add to that in that they show that the subject making the judgment can be (part of) that norm. They underpin the Protagorean claim that man is the measure of all things, which was discussed exactly with an aim to show the relativity of its application to the subject who judges (see Plato's Theaetetus; Jowett, 1892).

The study we conducted supports the egocentricity hypothesis we sought to investigate, namely that the ascriptions of vague predicates like *tall* and *heavy* to bodily figures relatively to which participants can locate their own measurements, are related to the participants' personal measurements. While the effect was not attested for male participants in the case of *heavy*, it

occurred in all other conditions in which participants had to ascribe those predicates to figures of their own sex. When participants had to ascribe the predicates to figures of the opposite sex, the evidence for egocentricity was less equivocal. These results are in line with the observation in the psychological literature that estimates of personal properties are more strongly related to one's own properties if the estimates are made for a person of the same sex. Together, these results suggest that our personal measurements may inform the norm for what we consider tall or heavy, when we can identify with the comparison class.

In a more general sense, our results refine the notion of subjectivity. We distinguished three aspects of theories of vagueness on which subjectivity could hypothetically operate: the comparison class, the standard of comparison, and the deviation from the standard. To the extent that we can assume that the comparison class was well specified in our study, subjectivity was found to affect the standard of comparison and/or what was considered a significant deviation from that standard. We established a relationship between subjects' personal measurements and the thresholds they used for categorization, which can be regarded a combination of the standard and the deviation. The ability to establish such a systematic relationship goes to show that subjectivity should not be interpreted as a mere stochastic process, but instead can be predicted from subject-specific properties. A challenge for future accounts of vagueness is to allow for the disentanglement of the effects subjectivity may have on the comparison class, the standard, and the deviation, and at the same time identify that part of individuals' response patterns that are due to probabilistic processes, which were clearly apparent in our study as well, in the shape of violations of monotonicity.

Our results also raise a number of questions, which we further clarify in this discussion. A first question concerns the proportion of our participants' judgments that is not explained by their personal measurements (section 4.1). While the correlations we found are reliable, they are not strong. This suggests that other factors are driving people's decisions about *tall* and *heavy* than their personal measurements. The question is whether we can identify those factors, or at least

make reasonable conjectures about them. A second question concerns the influence of ideals. In our study, we asked subjects to report on three variables of interest in the case of *heavy*: they reported their weight in kg and indicated their perceived and ideal weight on a body-image assessment scale. Subjects' categorization thresholds for *heavy* were best predicted from subjects' ideal weight. We may therefore wonder whether subjective ideal representations play a distinctive role in people's ascriptions of vague predicates (section 4.2). Our final questions concern the influence of egocentricity in predicates that do not self-apply and the specific challenges subjectivity poses for researching the vagueness of multidimensional predicates (section 4.3).

4.1. Systematic Subjectivity

We provided evidence for subjectivity in the representations of *tall* and *heavy* by establishing a significant correlation between the personal measurements participants provided and the thresholds these participants used to separate instances to which these predicates apply from those to which they do not apply.[7] While these correlations allow one to predict one's threshold for *tall* (*heavy*) from one's height (weight), these predictions are far from perfect, indicating that other factors than the measurements of one's body influence the position of the threshold. In other work we have identified age (Verheyen, Ameel, and Storms, 2011; Verheyen, Droeshout, and Storms, 2017) as a factor systematically affecting the position of one's categorization threshold. The correlation we established between bodily measurements and categorization cannot be attributed to age however, as we took care to control for this factor.

Height, weight, and age are examples of *personal* characteristics that may systematically affect the threshold one uses for categorization. Characteristics that are *shared* across individuals can

[7] An anonymous reviewer pointed out the possibility of subjectivity in subjects who do not have a metric available to report height/weight. We believe our methodology could still be used to bring subjectivity to light under these circumstances, either by obtaining objective measurements from the subjects (rather than subjective reports) or by having the subjects position themselves on a visual scale like the tool we used to assess Perceived and Ideal weight. Both solutions presume that researcher and subject rely on the same dimension for the application of the term, which needn't be the case (see section 4.3 for further discussion).

also have a systematic impact on the threshold that one employs. The fact that men impose a higher threshold for *tall* than women do (see Figure 2) might be interpreted as evidence for an own-gender perspective, whereby men interpret the question 'Do you find x tall' from a male perspective and women from a female perspective (see also Stukken, Verheyen, and Storms, 2013). There are several other of these *sociolinguistic* variables that might affect the judgments. The range of heights and weights one has experienced, for instance, is bound to influence judgments concerning *tall* and *heavy*. For this reason, we required participants to have the same country of origin. Two factors to consider in future research are the participants' level of education and their socio-economic status, as both have been systematically associated with categorization (see Verheyen & Storms, 2018, and Bourdieu, 1979, respectively) and height and weight (e.g., Davey Smith and Davies, 2016).

The observation that personal and shared characteristics systematically influence categorization, raise the fundamental question whether faultless disagreement should remain a diagnostic of vagueness (see also Verheyen and Storms, 2018). If - as our results suggest - subjectivity is not mere randomness but can to some extent be predicted from the participants' properties, one can imagine identifying additional factors that allow the participants' categorization thresholds to be predicted even better, effectively reducing the unexplained inter-individual variability. The problem then is that one can never be certain that the remaining individual differences truly reflect faultless disagreement. How would one ascertain that *all* relevant participants' properties (both personal and shared) have been taken into account? Rather than to rely on the observation of inter-individual application differences, we therefore propose that more emphasis be placed on intra-individual application differences for the diagnosis of vagueness (Egré, de Gardelle, and Ripley, 2013; Hersh and Caramazza, 1976) as these cannot be explained in terms of egocentric or sociolinguistic variation.

4.2. Ideals

We entertained the hypothesis that participants might rely on egocentric indices to establish a threshold for tallness and heaviness, respectively. The correlations that most strongly supported this hypothesis were those involving participants' ideal rather than reported or perceived weights.[8] Although we cannot infer the contribution that each of these measures might have had in the calculation of the categorization thresholds because the different measurements are all highly correlated with each other, these results do raise the question what role subjective ideal representations play in vagueness.

The possibility that ideal representations might determine the extension of predicates as *tall* and *heavy* resonates with the renewed interest in the role of ideals in the concepts and categories literature. The notion of ideal is understood in different manners in this literature, however. It should not be equated with the interpretation of an extreme, characterized by extreme values on certain dimensions, either true of only a few category members or true of none at all (Barsalou, 1985; Voorspoels, Vanpaemel, and Storms, 2011). The notion of ideals we entertain here is more in line with the culturally determined desirability that has been uncovered in the anthropological literature on concepts and categories (Atran, 1999; Burnett, Medin, Ross, and Blok, 2005) with the exception that we do not consider this desirability to be solely culturally, but also individually determined.[9] Our results support this conjecture in that while we observed a shared tendency to regard slimmer bodies as ideal, we also found considerable inter-individual variability to exist around the average ideal value (see Table 1). This interpretation of ideals in terms of desirability is also found in the philosophical literature (Bear and Knobe, 2016; Egré and Cova, 2015; Fara, 2000). While in Bear and Knobe (2016) it is implicitly assumed that everyone employs the same ideal, we

[8] We did not ask participants to provide ideal height values because we deemed the question meaningless as height, unlike weight, is not under one's control. We later learned this was a mistake: individuals entertain a subjective representation of their ideal weight *and* height (Cash and Jacobi, 1992; Jacobi and Cash, 1994).
[9] Desirability might be conceived of as an ideal in the extreme sense when it is thought of as a reconfiguration of the dimensions that underlie the stimuli. For instance, if we represent the height dimension on a string and pick up the string at the point representing the ideal height, that would effectively collapse the dimension into a new dimension with the ideal as an extreme.

follow Fara (2000) and Egré and Cova (2016) in entertaining the possibility of subjective ideals. In order to establish whether ideal representations have a distinct contribution to the threshold for categorization, one will in future research have to turn to predicates where there is no systematic relationship between the perceived and desired situation (as opposed to weight, where most people indicate that they would like to lose weight).

In future work we plan to investigate to what extent our findings pertaining to adjectives for physical characteristics can be generalized to the full range of subjective adjectives. We are first and foremost interested in establishing whether egocentricity also affects the interpretation of predicates that do not self-apply, like *expensive*. Although one generally does not say of oneself that one is expensive (except in the metaphorical sense), what one considers an expensive product might be influenced by one's monthly income or one's savings. Ideals might be involved in the use of these predicates too. A product might be considered expensive in light of one's current income, but affordable in light of one's anticipated income (e.g., once a debt has been paid off or one is promoted). For a range of these predicates we are in the process of assessing how subjectivity in descriptive, normative values vs. subjectivity in prescriptive ideals affects their use (Verheyen and Egré, 2017). The finding that people's culinary taste is a marker of age (Stevens, 1996) and one's taste in arts is a marker of social class and level of education (Bourdieu, 1979) suggests that the framework can also be fruitfully extended to the study of individual differences in the use of predicates of personal taste such as *tasty* (Lasersohn, 2005) and aesthetic adjectives such as *beautiful* (McNally & Stojanovic, 2017).

4.3. Vagueness in Degree and Vagueness in Criteria

Our demonstration of egocentricity involved the predicates *tall* and *heavy*. Both are gradable adjectives meaning that they express a quality that can vary in intensity or degree. By design, we made it so that any variation among the stimuli occurred along an apparent, single dimension. The

tall stimuli varied in height and the *heavy* stimuli varied in weight. As a result, the vagueness of the predicates necessarily manifested itself in the extension of the predicates: any individual categorization differences pertain to the number of instances the predicates are applied to. This type of vagueness is sometimes referred to as extensional vagueness, gradual vagueness, or vagueness in degree (as opposed to intensional vagueness or vagueness in criteria; Alston, 1964; Burks, 1946; Devos, 1995, 2003; Kennedy, 2013; Machina, 1976; Verheyen, Droeshout, and Storms, 2017; Verheyen and Storms, 2013, 2018). Devos (1995, 2003) defines vagueness in criteria as the indeterminacy with respect to (the combination of) the conditions for application of a term. Vagueness in degree, then, is the extent to which a term can be applied given that the conditions have been determined. This is straightforward in our experiments as each set of stimuli only varies along one dimension by design. This constitutes a simplification, however, as one can, for instance, easily imagine that both weight and body shape are independently or concurrently used to establish whether someone is heavy or not. The simplification is a necessary one in this stage of the research agenda. Establishing egocentricity would be even more complicated for multidimensional predicates, for individuals might not only value the dimensions differently, but also might entertain competing ideals on each dimension (picture a boxer working towards a firm body, but who needs to keep his weight down to be able to box in a specific weight class).

Which dimensions are used to determine whether a term applies or not and which corresponding ideals are entertained varies between individuals and is largely determined by one's interests and goals (Verheyen, Voorspoels, and Storms, 2015). Based on the psychological literature on self-attribution we expect to find even more evidence for the involvement of ideals when vagueness in criteria is in play. Arguably vagueness in criteria is less of an issue for predicates like *tall* and *heavy*, coming with a limited number of rather concrete dimensions, than for more psychological predicates based on a larger of number of relevant dimensions. Participants have been shown to idealize more on the latter than on the former (Dunning, Meyerowitz, and Holzberg, 1989; Hayes and Dunning, 1997).

5. Conclusion

Gradable adjectives are subjective: the comparison class, the standard of comparison, and what counts as a significant deviation from the standard may differ depending on the subject using the adjective. In this paper we looked at two paradigmatic examples of gradable adjectives, namely heavy and tall, in order to shed light on the phenomenon of subjectivity. What we found is that the use of both adjectives is affected in a systematic way by a form of egocentric reference. In particular, the subject's own height and weight systematically impact the interpretations of tall and heavy, in that the employed categorization threshold (standard + deviation) for tall will be higher, the taller the subject is. A similar relationship holds for heavy and the subject's weight. We found that correlation not just between participants' actual measurements and their threshold, but established an even more robust one between their representation of ideal values and their threshold. From this, we may conclude than when two persons disagree about whether someone is tall, or heavy, part of their disagreement originates from the fact that they judge relative to their own personal characteristics. The phenomenon presumably generalizes over a wide class of gradable adjectives (including those that a subject cannot self-apply such as expensive) and characteristics (such as level of education, social-economic status, age).

References

Alston, W. P. 1964. *Philosophy of Language*. Englewood Cliffs: Prentice-Hall.

Alxatib, S., and Pelletier, F. J. 2011. The psychology of vagueness: Borderline cases and contradictions. *Mind and Language, 26*, 287-326.

Atran, S. 1999. Itzay Maya folkbiological taxonomy. In D. L. Medin and S. Atran (eds.), *Folkbiology* (pp. 119-204). Cambridge, MA: MIT press.

Barsalou, L. W. 1985. Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 629-654.

Bear, A. and Knobe, J. 2016. Normality: Part descriptive, part prescriptive. *Cognition,* http://doi.org/10.1016/j.cognition.2016.10.024..

Biernat, M., Manis, M., and Nelson, T. E. 1991. Stereotypes and standards of judgment. *Journal of Personality and Social Psychology, 60*, 485-499.

Bourdieu, P. 1979: *La Distinction: Critique social du jugement*. Paris: Editions de minuit.

Brownell, H. H., and Caramazza, A. 1978. Categorizing with overlapping categories*. Memory & Cognition, 6*, 481-490.

Burks, A. W. 1946. Empiricism and vagueness. *The Journal of Philosophy, 43*, 477-486.

Burnett, H. 2016. *Gradability in Natural Language: Logical and Grammatical Foundations*. Oxford, UK: Oxford University Press.

Burnett, R. C., Medin, D. L., Ross, N. O., and Blok, S. V. 2005. Ideal is typical. *Canadian Journal of Experimental Psychology, 59*, 3-10.

Cash, T. F., and Jacobi, L. 1992. Looks aren't everything (to everybody): The strength of ideals of physical appearance. *Journal of Social Behavior and Personality, 7*, 621-630.

Davey Smith, G. and Davies, N. M. 2016. Can genetic evidence help us understand why height and weight relate to social position? *BMJ, 352*, doi: https://doi.org/10.1136/bmj.i1224.

Devos, F. 1995. Still fuzzy after all these years. A linguistic evaluation of the fuzzy set approach to semantic vagueness. *Quaderni di Semantica, 16*, 47-82.

Devos, F. 2003. Semantic vagueness and lexical polyvalence. *Studia Linguistica, 57*, 121-141.

Douven, I., Wenmackers, S., Jraissati, Y., and Decock, L. 2016. Measuring graded membership : The case of color. *Cognitive Science, In press*.

Dunning, D., and Cohen, G. L. 1992. Egocentric definitions of traits and abilities in social judgment. *Journal of Personality and Social Psychology, 63*, 341-355.

Dunning, D., and McElwee, R. O. 1995. Idiosyncratic trait definitions: Implications for self-description and social judgment. *Journal of Personality and Social Psychology, 68*, 936-946.

Dunning, D., Meyerowitz, J. A., and Holzberg, A. D. 1989. Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57*, 1082-1090.

Egré, P. 2016. Vague judgment: A probabilistic account. *Synthese,* DOI 10.1007/s11229-016-1092-2.

Egré, P., and Cova, F. (2015). Moral asymmetries and the semantics of *many*. *Semantics and Pragmatics, 8*, 1-45.

Egré, P., de Gardelle, V., and Ripley, D. 2013. Vagueness and order effects in color categorization. *Journal of Logic, Language and Information, 22*, 391-420.

Egré, P., and Zehr, 2018. Are gaps preferred to gluts? A closer look at borderline contradictions. In G. W. Sassoon, L. McNally, and E. Castroviejo (eds.), *Gradability, scale structure and vagueness: Experimental perspectives*. Springer. Manuscript accepted for publication.

Ezzati, M., Martin, H., Skjold, S., Vander Hoorn, S., and Murray, C. J. L. 2006. Trends in national and state-level obesity in the USA after correction for self-report bias: analysis of health surveys. *Journal of the Royal Society of Medicine, 99*, 250-257.

Fara, D. G. 2000. Shifting sands: An interest-relative theory of vagueness. *Philosophical Topics, 28*, 45-81. Originally published under the name 'Delia Graff'.

Fara, D. G. 2008. Profiling interest relativity. *Analysis*, *68*, 326-335.

Fillenbaum, S. 1961. How fat is fat? Some consequences of similarity between judge and judged object. *Journal of Psychology, 52*, 133-136.

Fine, K. 1975. Vagueness, truth and logic. *Synthese, 30*, 265-300.

Flin, R.H., and Shepherd, J. W. 1986. Tall stories: Eyewitnesses' ability to estimate height and weight characteristics. *Human Learning: Journal of Practical Research and Applications, 5*, 29-38.

Gardner, R. M., Jappe, L. M., and Gardner, L. 2009. Development and validation of a new figural drawing scale for body-image assessment: The BIAS-BD. *Journal of Clinical Psychology, 65*, 113-122.

Goffman, E. 1977. The arrangement between the sexes. *Theory and Society, 4*, 301-331.

Goy, A. 2002. Grounding Meaning in Visual Knowledge. In K. R. Coventry and P. Olivier (eds.), *Spatial Language: Cognitive and Computational Perspectives* (pp. 121–145). Dordrecht, The Netherlands: Springer.

Guttman, L. 1944. A basis for scaling qualitative data. *American Sociological Review*, 9, 139-150.

Hampton, J. A. 1998. Similarity-based categorization and fuzziness of natural categories. *Cognition, 65*, 137-165.

Hampton, J. A. 2007. Typicality, graded membership, and vagueness. *Cognitive Science, 31*, 355-384.

Hayes, A. F., and Dunning, D. 1997. Conceptual processes and trait ambiguity: Implications for self-peer agreement in personality judgment. *Journal of Personality and Social Psychology, 72*, 664-677.

Heine, S. J. 2008. *Cultural Psychology*. New York, NY : W. W. Norton.

Helson, H. 1947. Adaptation-level as frame of reference for prediction of psychophysical data. *The Americal Journal of Psychology, 60*, 1-29.

Hersh, H. M., and Caramazza, A. 1976. A fuzzy set approach to modifiers and vagueness in natural language. *Journal of Experimental Psychology: General, 105, 254-276*.

Hinckley, E. D., and Rethlingshafer, D. 1951. Value judgments of heights of men by college students. *The Journal of Psychology, 31*, 257-262.

Holyoak, K. L., and Gordon, P. C. 1983. Social reference points. *Journal of Personality and Social Psychology, 44*, 881-887.

Jacobi, L., and Cash, T. F. 1994. In pursuit of the perfect appearance: Discrepancies among self-ideal percepts of multiple physical attributes. *Journal of Applied Social Psychology, 24*, 379-396.

Jowett, M. A. 1892. *The Dialogues of Plato, Volume 1. Translated into English with Analyses and Introductions.* Third edition revised and corrected. London, UK: Oxford University Press.

Kamp, H., and Sassoon, G. 2014. Vagueness. In P. Dekker and M. Aloni (eds.), *The Cambridge handbook of formal semantics* (pp. 389-441). Cambridge, MA: Cambridge University Press.

Kahneman, D., and Tversky, A. 1979. Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263-292.

Keefe, R. 2000. *Theories of Vagueness*. Cambridge, MA: Cambridge University Press.

Kennedy, C. 2007. Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, *30*, 1-45.

Kennedy, C. 2013. Two sources of subjectivity: Qualitative assessment and dimensional uncertainty. *Inquiry, 56*, 258-277.

Klein, E. 1980. A semantics for positive and comparative adjectives. *Linguistics and philosophy*, 4, 1-45.

Kölbel, M. 2004. Faultless disagreement. *Proceedings of the Aristotelian Society*, 104, 53-73.

Kunda, Z., and Nisbett, R. E. 1988. Predicting individual evaluations from group evaluations and vice versa: Different patterns for self and others? *Personality and Social Psychology Bulletin, 14*, 326-334.

Lasersohn, P. 2005. Context dependence, disagreement, and predicates of personal taste. *Linguistics and Philosophy, 28*, 643-686.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. 2000. WinBUGS: A Bayesian modeling framework: Concepts, structure, and extensibility. *Statistics and Computing, 10*, 325-337.

Machina, K. F. 1976. Truth, belief and vagueness. *Journal of Philosophical Logic, 5*, 47-78.

McNally, L., and Isidora, S. 2017. Aesthetic adjectives. In J. Young (ed.), *The Semantics of Aesthetic Judgments* (pp. 17-37). Oxford, UK: Oxford University Press.

Mintz, E. 1956. An example of assimilative projection. *Journal of Abnormal and Social Psychology, 52*, 270-280.

Raffman, D. 2014. *Unruly Words*. Oxford, UK: Oxford University Press.

Reardon, J., and Miller, C. 2012. The effect of response scale type on cross-cultural construct measures. *International Marketing Review, 26*, 24-53.

Rethlingshafer, D., and Hinckley, E. D. 1963. Influence of judges' characteristics upon the adaptation-level. *The American Journal of Psychology, 76*, 116-119.

Ripley, D. 2011. Contradictions at the borders. In R. Nouwen, R. van Rooij, U. Sauerland, and H.-C. Schmitz (eds.), *Vagueness in Communication* (pp. 169–188). Berlin, Heidelberg: Springer.

Rips, L. J., and Turnbull, W. 1980. How big is big? Relative and absolute properties in memory. *Cognition, 8*, 145-174.

Sapir, E. 1944. Grading, a study in semantics. *Philosophy of science*, *11*, 93-116.

Sæbø, K. J. 2009. Judgment ascriptions. *Linguistics and Philosophy,* 32, 327–352.

Sörqvist, P., Langeborg, L., and Eriksson, M. 2011. Women assimilate across gender, men don't: The role of gender to the own-anchor effect in age, height, and weight estimates. *Journal of Applied Social Psychology, 41*, 1733-1748.

Stanley, J. 2003. Context, interest relativity and the sorites. *Analysis*, *63*, 269-281.

Stevens, D. A. 1996. Individual differences in taste perception. *Food Chemistry, 56*, 303-311.

Stukken, L., Verheyen, S., and Storms, G. 2013. Representation and criterion differences between men and women in semantic categorization. In M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth (eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3474-3479). Austin, TX: Cognitive Science Society.

Suzuki, T. 1970. An essay on the anthropomorphic norm. In R. Jakobson, and S. Kawamoto (eds.), *Studies in General and Oriental Linguistics* (pp. 552-556). Tokyo, Japan: TEC.

Tribushinina, E. 2008. *Cognitive Reference Points: Semantics Beyond the Prototypes in Adjectives of Space and Colour.* Unpublished doctoral dissertation, Leiden University, Leiden, The Netherlands.

Varela, F., Thompson, E., and Rosch, E. 1991. *The Embodied Mind*. Cambridge, MA: MIT Press.

Verheyen, S., Ameel, E., and Storms, G. 2011. Overextensions that extend into adolescence: Insights from a threshold model of categorization. In L. Carlson, C. Hölscher, and T. F. Shipley (eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2000-2005). Austin, TX: Cognitive Science Society.

Verheyen, S., Droeshout, E., and Storms, G. 2017. Age-related degree and criteria differences in semantic categorization. *Unpublished manuscript.*

Verheyen, S., and Egré, P. 2017. Typicality and graded membership in dimensional adjectives. *Unpublished manuscript.*

Verheyen, S., Hampton, J. A., and Storms, G. 2010. A probabilistic threshold model: Analyzing semantic categorization data with the Rasch model. *Acta Psychologica, 135*, 216-225.

Verheyen, S., and Storms, G. 2013. A mixture approach to vagueness and ambiguity. *PLoS ONE, 8 (5)*: e63507.

Verheyen, S., and Storms, G. 2018. Education as a source of vagueness in criteria and degree. In G. W. Sassoon, L. McNally, and E. Castroviejo (eds.), *Gradability, scale structure and vagueness: Experimental perspectives*. Springer. Manuscript accepted for publication.

Verheyen, S., Voorspoels, W., and Storms, G. 2015. Inferring choice criteria with mixture IRT models: A demonstration using ad hoc and goal-derived categories. *Judgment and Decision Making, 10*, 97-114.

Vestlund, J., Langeborg, L., Sörqvist, P., and Eriksson, M. 2009. Experts on age estimation. *Scandinavian Journal of Psychology, 50*, 301-307.

Voelkle, M. C., Ebner, N. C., Lindenberger, U., and Riediger, M. 2012. Let me guess how old you are: Effects of age, gender, and facial expression on perceptions of age. *Psychology and Aging, 27*, 265-277.

Vogel, A. (2004). *Swedish Dimensional Adjectives*. Stockholm, Sweden: Almqvist and Wiksell.

Voorspoels, W., Vanpaemel, W., and Storms, G. 2011. A formal ideal-based account of typicality. *Psychonomic Bulletin and Review, 18*, 1006-1014.

Ward, C. D. 1967. Own height, sex, and liking in the judgment of the heights of others. *Journal of Personality, 35*, 381-401.

Wetzels, R., and Wagenmakers, E.-J. 2012. A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin and Review, 19*, 1057-1064.

Wright, C. 1976. Language-mastery and the sorites paradox. In G. Evans and J. MacDowell (eds), *Truth and Meaning* (pp. 223-247). Cambridge, MA: MIT Press.

Wright, C. 1995. The epistemic conception of vagueness. *The Southern Journal of Philosophy*, 33, 133-159.