Ordering subjectivity and the absolute/relative distinction[1]

Stephanie Solt — ZAS Berlin

Abstract. This paper presents the results of an experimental investigation into faultless disagreement effects with the comparative forms of gradable adjectives. It is shown that subjective judgments regarding orderings are possible with a wide range of adjectival predicates, and furthermore that the presence of such subjectivity correlates with the distinction between absolute and relative gradable adjectives. A theory is developed in which both phenomena derive from the formal properties of adjectival measure functions.

Keywords: adjectives, gradability, comparative, measurement, multidimensionality, standard type.

1. Introduction

The data in (1)-(2) illustrate the well-known distinction between **relative** and **absolute** gradable adjectives (Rotstein and Winter, 2004; Kennedy and McNally, 2005; Kennedy, 2007). Members of the relative class such as *tall* and *short* have context- or comparison class-dependent standards, allowing them to occur with *for* phrases, and disallowing composition with endpoint-oriented degree modifiers such as *slightly* and *completely*. Absolute gradable adjectives such as *clean* and *dirty*, by contrast, have scalar maxima or minima as standards, allowing modification with *slightly* and/or *completely*, and resulting in infelicity when modified by a comparison class-setting *for* phrase.

(1) a. Anna is tall / short for an 8-year-old.
      b. ?? The shirt is dirty / clean for a dress shirt.

(2) a. ?? Anna is slightly tall / slightly short / completely tall / completely short.
      b. The shirt is slightly dirty / completely clean.

The data in (3) exemplify another binary subdivision of gradable adjectives, based on the presence or absence of what I will call **ordering subjectivity** (Kennedy, 2013; Bylinina, 2014; McNally and Stojanovic, 2015). A disagreement about which of two individuals is taller or shorter is necessarily factual in nature; only one of the two speakers can have said something correct (3a). But two competent speakers may disagree as to which of two paintings is more beautiful or which of two dishes is tastier, with neither appearing to be at fault (3b,c).

(3) a. A: Anna is taller (shorter) than Zoe.   **factual only**
         B: No, Zoe is taller (shorter) than Anna!

    b. A: The Picasso is more beautiful than the Miró.   **faultless**
       B: No, the Miró is more beautiful.

    c. A: The chili is tastier than the soup!   **faultless**
       B: No, the soup is tastier!

The above difference between *tall/short* on the one hand and *beautiful* on the other corresponds to the distinction drawn by Bierwisch (1989) between **dimensional** and **evaluative** adjectives. *Tasty* exemplifies so-called predicates of personal taste (Lasersohn, 2005), which might be considered a subset of the evaluative class. The presence of faultless disagreement in the comparative is only one of the ways in which these two classes of adjectives diverge in their behavior, others involving the nature of the antonymy relationship, the existence of entailments between comparative and positive forms, and the possibility of embedding under the subjective attitude verb *find* (Bierwisch, 1989; Sæbø, 2009; Kennedy, 2013; Bylinina, 2014).

It should be apparent from the above examples that these two classifications overlap. Specifically, *tall*, which is often cited as the paradigm case of a relative gradable adjective, is also the classic example of a dimensional adjective. This leads to the question of how the other subclasses relate to one another. In particular, there has been no discussion of how absolute gradable adjectives pattern with respect the dimensional/evaluative distinction, and in particular the presence or absence of ordering subjectivity. Can two competent speakers disagree faultlessly as to which of two shirts is cleaner/dirtier? Which of two surfaces is flatter/bumpier? Which of two lines is straighter/more curved? If we are to understand the source of subjective judgments regarding comparative statements, it is important to know first of all <u>which</u> adjectives allow this, but intuitions here seem shaky.

In this paper, I report the results of an experimental study which show that ordering subjectivity is far more common than has been previously recognized, and further that the presence or absence of such subjectivity correlates with the nature of the standard invoked by the adjective in its positive form. After discussing issues faced by existing theories in accounting for these facts, I put forward a theory of adjectival meaning according to which the availability of objective versus subjective readings derives from constraints on the measure functions lexicalized by gradable adjectives of various sorts. I further show that these same factors play a role in determining the nature of the standard for the positive form.

2. Experiment: Faultless Disagreement Paradigm

As seen in examples such as (3), when it comes to dimensional adjectives such as *tall* and evaluative adjectives such as *beautiful*, judgments about comparative statements are clear: the former are necessarily objective or fact based, the latter are subjective. But when we extend our focus to a wider range of adjectives, the picture becomes murkier. The present study aims to establish a firmer empirical basis for theoretical work by using a novel faultless disagreement paradigm to diagnose the presence of ordering subjectivity among a wide range of adjective types.

2.1. Methodology and stimuli

The experiment employed a forced choice task in which participants saw brief dialogues of the form in (4)-(6), and were asked to classify the nature of the disagreement between the two speakers. Two response options were given: "only one can be right; the other must be wrong" and "it's a matter of opinion". The first of these was classified as a judgment of 'fact', the second as a judgment of 'opinion'.

(4)   A: John and Fred look similar but John is taller than Fred.
      B: No, Fred is the taller one of the two.

(5)   A: Look – Tommy's shirt is dirtier than the one his little brother Billy is wearing.
      B: No, Billy's shirt is dirtier than Tommy's.

(6)   A: The necklace Susan is wearing today is uglier than the one she had on yesterday.
      B: No, the one she was wearing yesterday was uglier.

A total of 35 gradable adjectives were tested, falling into the following five categories:

- Relative gradable adjectives with numerical measures (**RELNUM**): *tall, short, old, new, expensive*

- Relative gradable adjectives without numerical measures (**RELNO**): *sharp, dull, dark, light, hard, soft*

- Absolute gradable adjectives with totally closed scales (**ABSTOT**): *full, empty*

- Absolute gradable adjectives with partially closed scales (**ABSPART**): *wet, dry, straight, curved, rough, smooth, clean, dirty, salty*

- Adjectives which may be classified as evaluative in a broad sense (**EVAL**): *good, bad, beautiful, pretty, ugly, easy, interesting, boring, tasty, fun, intelligent, happy, sad*

Adjectives were assigned to these categories according to judgments reported in the literature as well as linguistic tests. Relative gradable adjectives were identified as those for which both the adjective and its antonym are acceptable in the frame *x is Adj but y is Adj-er*, and for which neither adjective nor antonym allows modification by *slightly*. Absolute gradable adjectives were identified as those for which either adjective or antonym is infelicitous in the above frame and/or can co-occur with *slightly*. An adjective was considered to have a numerical measure if its comparative form can be modified by a measure phrase.[2] The evaluative category includes adjectives of the sort

---

[2]The test was based on the comparative because, as is well known, many adjectives compose with measure phrases in the comparative but do not allow direct measure phrases (e.g. *ten dollars more expensive* vs. **ten dollars expensive*), and as such the comparative provides a better test for the existence of a numerical measurement system. The reasons for the restrictions on direct measure phrases are complex and seemingly idiosyncratic; see Schwarzschild (2005); Sassoon (2010) for discussion.

discussed in the literature under the term 'evaluative' or 'predicates of personal taste'. This is a mixed class, encompassing value, taste and aesthetic judgments, emotion words, and psychological predicates; they are united, and distinguished from the other four categories, in that they do not denote external physical properties. Note finally that the RELNUM category corresponds to the class which in other work has been called dimensional adjectives, a point I will return to below.

The experiment was administered online via Amazon MTurk, with test items split across 4 lists. Each list contained 8-12 test items and 12 fillers. Some adjectives appeared on more than one list, in different item contexts. Fillers were split equally between two types: i) those expected to yield 'opinion' judgments, including vague nominal predicates (e.g. *jerk*), deontic and epistemic modals, statements of likelihood, and moral statements; ii) those expected to yield 'fact' judgments, based on factual statements (example: A: The judge found Frank guilty. B: No, the judge found Frank innocent.). Sample size was 20-25 per list, for a total sample size of $n = 91$. Full stimuli are available at http://www.zas.gwz-berlin.de/fileadmin/mitarbeiter/solt/fault.pdf.

2.2. Results

Results by adjective class and for individual adjectives are displayed in Figure 1. A generalized linear mixed effects model was fitted to the results using the lme4 package (Bates et al., 2014) in R (R Core Team, 2015), with response ('fact' vs. 'opinion') as dependent variable, adjective type as a fixed effect, and random intercept for subject. The effect of adjective class was found to be significant, with post hoc testing via the multcomp package (Hothorn et al., 2008) using Tukey correction for multiple comparison showing all pairwise comparisons to be significant at p<0.001 except RELNO vs. ABSPART (p<0.01) and RELNUM vs. ABSTOT (nonsignificant).

2.3. Discussion and further observations

The first conclusion to be drawn from the present experiment is that ordering subjectivity is widespread, found not only with classic cases of evaluative adjectives such as *beautiful* and *tasty*, as in our original examples, but to some extent for the majority of the adjectives tested.

The second observation that emerges is that in this respect, gradable adjectives divide into not two but three distinct classes. Some allow (almost) exclusively objective or factual readings of the comparative (the RELNUM and ABSTOT classes). Some by contrast are radically subjective in the comparative, eliciting few if any 'fact' judgments (the EVAL class). Finally, some fall in between the two extremes, allowing both objective and subjective readings for the comparative (the RELNO and ABSPART classes); furthermore, among this third group, adjectives range from those that skew towards objective readings to those that skew towards subjective readings.
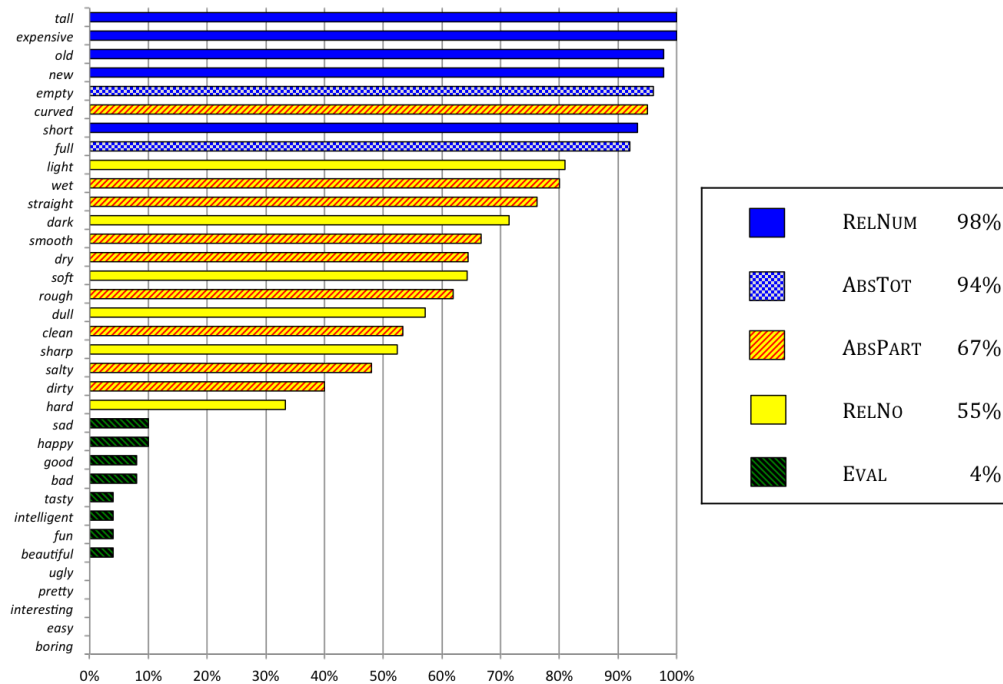
Figure 1: Results of Experiment - Percent 'Fact' Judgments

Looking at the results by adjective class, it seems that we have found a correlation between ordering subjectivity on the one hand and the nature of the standard invoked by the adjective in its positive form on the other. Specifically, purely objective judgments about scalar orderings are restricted to a certain subset of the relative class and the totally closed scale pair *full/empty*. Classic examples of absolute partially closed scale adjectives such as *clean/dirty*, *wet/dry* and *rough/smooth* fall consistently in the mixed class. This is a surprising finding. One of the phenomena under investigation here, the absolute/relative distinction, involves the positive form of the adjective, and has been attributed to the presence or absence of scalar endpoints. The second involves the comparative form, and has to do with the nature of judgments regarding the ordering of individuals along the relevant dimension. There is no obvious *a priori* reason to expect the two to be related.

There is, however, a clear confound here, namely the presence or absence of a system of numerical measurements. It is not relative gradable adjectives as a whole whose comparative forms elicited exclusively objective interpretations, but more specifically those with corresponding units of measure; those without such units, such as *light/dark* and *soft/hard*, patterned with the absolute partially closed scale group. Furthermore, the totally closed scale pair *full/empty* might also be aligned to the class of adjectives with measurement units, in that degrees of fullness or emptiness can be quantified in percentages (e.g. *90% full*). Perhaps it is the availability of a numerical measurement system itself, and not the scale structure of the adjective or the type of standard it has, that is correlated with the possibility of ordering subjectivity. This is a very reasonable idea. In order to assign a numerical degree of 'adjective-ness' to an entity, it must be possible to objec-

tively measure the degree to which that entity has the property in question. Quite plausibly, the same possibility of objective measurement is also responsible for the objective nature of judgments regarding orderings. We might then suspect the apparent correlation with standard type to be an artifact of the experimental design, in that absolute gradable adjectives with measurement units were not included among the stimuli.

Such an explanation, though, is problematic, for a simple reason: putting aside adjectives such as *full/empty* that allow proportional measures, there <u>are</u> virtually no absolute gradable adjectives with numerical measures. Commonly cited examples of the absolute class, such as those in the ABSPART group tested here, lack standard units. Conversely, Schwarzschild (2005) discusses 49 adjectives that allow measure phrases at least in the comparative. Of these, only 4 pattern (in my judgment, using the above-described tests) as absolute: *early/late* (which Schwarzschild argues are covert comparatives), and *intelligent/stupid* (whose status regarding both standard type and numerical measurability is unclear). This is a surprising pattern, which to my knowledge has not been previously observed, and which certainly merits some sort of explanation. The puzzle we are faced with can then be restated as follows: what is the reason for the correlation between objectivity – encompassing both the existence of measurement units and the absence of ordering subjectivity – and the type of standard invoked by the adjective. It is this puzzle that I address in the remainder of the paper.

3. Previous accounts

In this section I briefly review relevant research in three areas relating to ordering subjectivity and the absolute/relative distinction, noting issues they face in accounting for the experimental findings, and highlighting proposals that will form the basis for the account developed below.

3.1. Gradability

The now standard semantic approach to gradability is the degree-based framework of Cresswell (1977); Kennedy (1997); Heim (2000) and others, according to which gradable adjectives lexicalize measure functions that map individuals to degrees on a scale (7).[3] The comparative can then be analyzed as expressing a relation between the degrees assigned to two individuals, as in (8):

(7) $[\![\text{dirty}]\!] = \lambda d\lambda x.\mu_{DIRTY}(x) \succeq d$

(8) a. Tommy's shirt is dirtier than Billy's.

    b. $\mu_{DIRTY}(\text{Tommy's shirt}) \succ \mu_{DIRTY}(\text{Billy's shirt})$

---

[3]For concreteness, here and below I adopt the version of the degree-based framework according to which gradable adjectives denote degree relations (Heim, 2000). If desired, the analysis could be restated on the alternate approach on which they have denotations instead as measure functions (Kennedy, 2007).

An unspoken assumption that seems to underlie representations of this form is that for any given adjective *Adj*, there is a unique measure function μ_{Adj}.[4] This is reasonable for adjectives such as *tall*, but problematic for those such as *beautiful* and even *dirty*, for which competent speakers can disagree faultlessly about the relative ordering of two individuals. If we adopt a degree-based framework, it seems we must somehow allow different mappings for different speakers.

The problem that arises with representations such as (8b) derives in part from the tendency to equate dimensions of measurement with the scales that track them. Recently, there have been proposals that do away with such a one-to-one correspondence, allowing instead a one-to-many relation between dimensions and scales. For example, Solt (2016b) argues that quantity dimensions – including even cardinality – can be tracked by multiple scales that differ in their structures, in particular in the strength of the ordering relation $\succ$. In the adjectival domain, Kennedy (2013) proposes that gradable adjectives have core meanings as specifications of dimensions, which can then be mapped to different sorts of scalar predicates. Finally, Sassoon (2010) proposes that the measure functions lexicalized by gradable adjectives are not fully specified but rather indexed to contexts or worlds, with two contexts/worlds potentially differing in the degrees that are assigned to individuals even when the physical properties of objects remain the same. Below, this will be taken as one of the building blocks for an account of ordering subjectivity.

3.2. Subjectivity

There is a sizable body of research on adjectival subjectivity or 'judge-dependence' (see Lasersohn 2005; Stephenson 2007 and other work cited below), focusing in particular on predicates of personal taste such as *tasty* and *fun*, which provide the best-known examples of faultless disagreement: when two speakers disagree as to whether or not a dish is tasty or an experience is fun, the disagreement appears to be a real one, yet neither speaker seems to have said something incorrect or false. Recently it has been observed that subjectivity of this sort is found with a much wider range of adjectives beyond obvious personal taste predicates, in particular occurring also with the positive forms of vague gradable adjectives such as *tall*: two speakers may disagree faultlessly as to whether an individual is tall, perhaps because they have in mind different thresholds or standards (Kennedy, 2013; Bylinina, 2014). These same authors have also observed that the subjective interpretation of comparative forms – that is, ordering subjectivity – is a more restricted phenomenon, occurring with only some types of gradable adjectives (cf. (3)).

There are two leading semantic approaches to subjectivity. The relativist analysis (Lasersohn, 2005) includes a judge parameter to the index of interpretation, along with the usual time and world

---

[4]As stated, this is not entirely correct, in that certain adjectives are obviously ambiguous as to dimension; *long*, for instance, has both spatial and temporal interpretations, and *light* can reference either weight or brightness. What seems largely assumed, however, is that once the dimension is fixed, so too is the scale and measure function.

parameters (9a). The contextualist approach (Stojanovic, 2007; Sæbø, 2009), by contrast, assumes that predicates of this sort feature an additional judge or experiencer argument (9b). Elaborations on and combinations of these two approaches are found in Stephenson (2007), Pearson (2013), and Bylinina (2014), among others.

(9) a. $[\![\text{tasty}]\!]^{w,t,j} = \lambda x.x$ tastes good to j in w at t
    b. $[\![\text{tasty}]\!]^{w,t} = \lambda y \lambda x.x$ tastes good to y in w at t

Importantly, in the form shown here, neither of these approaches accounts for ordering subjectivity. Although *tasty* is gradable, the above analyses localize subjectivity at the level of the positive form, thus providing no explanation for subjective judgments regarding scalar orderings. Minimally, then, it seems that measure functions such as those in (7) must also be indexed or relativized to judges. Nor do we have an explanation for our experimental finding that adjectives exhibiting ordering subjectivity fall into two classes, depending on whether or not they also allow objective readings for the comparative.

A perhaps deeper issue is that for at least some of the adjectives we have considered, ordering subjectivity does not appear to be restricted to differences in perspective between speakers; rather, it seems that a single speaker's judgments with respect to orderings are potentially uncertain or variable. Consider for example two shirts, one which is clean except for a grass stain on the sleeve, while the other is slightly dingy overall. Which one do I consider dirtier, and which cleaner? I think my answer has to be to be 'it depends' – on what type of shirt and how it will be used, on what sort of dirt we are most concerned about, and so forth. The same might be said, for example, regarding which of two roads is bumpier, or which of two fences is straighter. Formally, subjectivity of this sort cannot be modeled in terms of variation in a judge parameter.

In recent work that has explicitly discussed ordering subjectivity, the intuition is put forward that it derives from the multidimensionality of the properties in question (see especially Kennedy, 2013; McNally and Stojanovic, 2015). Whereas the attribution of a predicate such as *tall* is dependent on a single dimension, namely height, that of a predicate such as *healthy* or *beautiful* depends on multiple underlying dimensions. Subjectivity of the sort considered here arises when there is uncertainty or potential disagreement as to the component dimensions and how they should be integrated. For example, two speakers might disagree about which of two individuals is healthier because they disagree about which aspects of health should carry more weight.

The above authors do not attempt to formalize this insight. However, a comprehensive semantic theory of adjectival multidimensionality is developed by Sassoon (2013 and other work), and further extended by Bylinina (2014), who in particular applies it to the subjectivity of evaluative adjectives such as *intelligent*. Below I will build on this further to account for the pattern of results from the present experiment.

3.3. Standard type

The final topic of interest to us here involves the distinction between absolute and relative gradable adjectives. As already alluded to, the leading account of this difference relates the interpretation of the positive form of the adjective to the structure of the scale it lexicalizes (Kennedy and McNally, 2005; Kennedy, 2007). When the scale is closed on one or both ends, the scalar minimum or maximum serves as the threshold or standard of application for the positive form. To be clean, for example, is to have a maximal degree of cleanness, while to be dirty is to have a more-than-minimum degree of dirtiness. When the scale is open on both ends, the standard is instead contextually determined, perhaps with reference to a comparison class. Thus to be tall is to have a degree of height greater than some contextually salient threshold, dependent in some way on the sort of entities under consideration. Kennedy proposes that this pattern derives from a Principle of Interpretive Economy, which calls for maximizing the semantic contribution of conventional linguistic elements such as scalar endpoints.

The scale-based account is elegant and highly explanatory, but leaves open the question of what determines the structure of the scale underlying the adjective's meaning. For example, why is dirtiness analyzed as having a lower-closed scale but height an open scale, given that both intuitively have a zero point? Work in this tradition has also not recognized the connection between standard type and objectivity/subjectivity that was uncovered in the present experiment.

A number of authors have proposed to derive the absolute/relative distinction in a principled manner from more basic factors, and certain of these proposals provide a starting point to addressing the latter issue. Toledo and Sassoon (2011) and Sassoon (2012) relate the difference between the two subclasses to the sort of comparison class they reference: for relative gradable adjectives, the comparison class is an extensional category of individuals, while for absolute gradable adjectives, it is made up of counterparts or temporal stages of the same individual. This derives from a more basic distinction in denotations: relative gradable adjectives such as *tall/short* are individual level predicates (Carlson, 1977), denoting stable, enduring properties, while their absolute counterparts such as *clean/dirty* are stage-level predicates, denoting temporally variable properties. The difference in standard type can then be related to how items in the two sorts of comparison classes are distributed. Counterpart comparison classes typically have a salient maximum and/or minimum element (e.g. the shirt in its state of complete cleanness), which provide the basis for an endpoint standard. Extensional comparison classes do not. Even further, in the case of adjectives such as *tall* that denote stable properties, mapping of comparison class members to the scalar zero does not occur, in that anything whose height we can inquire about has non-zero height. A scalar minimum would thus be a trivial standard, and therefore a relative standard emerges instead.

Interestingly, a very similar insight is expressed in a quite different framework by Lassiter and Goodman (2013) and Qing and Franke (2014), who develop related pragmatic analyses in which the thresholds for gradable adjectives are modeled probabilistically as functions of the assumed prior distribution of entities in the comparison class with respect to the relevant dimension. A

context-dependent relative standard emerges when a normal-like prior is assumed; an absolute standard arises when there is sufficient probability density at one of the scalar endpoints.

Here I will not adopt either of these accounts in the form described above. Rather, the central insight that I will draw on is that to understand the source of an adjective's standard type, it is necessary to consider the the possible values to which members of the domain may be mapped, with particular reference to potential triviality of standards.

4. Accounting for Ordering Subjectivity

Bierwisch (1989) proposes that gradability derives from a mental operation of comparing individuals: "there is no degree without comparison and no comparison without degree" (p. 112). If this is indeed so, namely that degrees originate in a speaker-internal psychological process, we might wonder why it is not the case that <u>all</u> comparative statements are subjective. The reason, I will argue, is that for some dimensions we can impose a structure on the domain through which speakers' judgments are necessarily aligned.

We may formalize this using concepts from measurement theory (see Kranz et al. 1971 for a basic introduction to measurement theory and Sassoon 2010 and Lassiter 2011 for linguistic applications). We begin with a domain Dom and a binary relation R on Dom reflecting some dimension of measurement, where R has the properties of a strict weak order (a relation that is irreflexive, asymmetric, transitive and transitive with respect to incomparability). Measure functions and scales are then defined as follows:

(10)   A **measure function** is a function $\mu_{DIM} : Dom \rightarrow S$ from Dom to some scale $S = \langle D, \succ, DIM \rangle$, where
   - DIM is the dimension of measurement
   - D is a set of degrees
   - $\succ$ is an ordering relation on D
   - $\forall a, b \in Dom, \mu_{DIM}(a) \succ \mu_{DIM}(b)$ iff aRb

As discussed above, this formalization has the effect of establishing a one-to-many rather than one-to-one relation between dimensions of measurement and the scales that track them.

We then adopt the previously discussed proposal of Sassoon (2010), according to which gradable adjectives are underspecified, lexicalizing not a single measure function but a family of functions indexed to contexts c. As on her account, I assume that two contexts c and c' may differ in the measures assigned to individuals, even if the physical properties of objects remain the same. Thus the general template for gradable adjective meaning is the following:

(11) $[\![\text{Adj}]\!]^c = \lambda d \lambda x. \mu^c_{DIM}(x) \succeq d$

To put this differently, gradable adjectives on this view lexicalize dimensions rather than particular scales. The presence or absence of ordering subjectivity can then be related to the further constraints on the set of measure functions μ^c_{DIM} encoded by the adjective.

4.1. Sources of objectivity

I propose that objective judgments regarding orderings arise when μ^c_{DIM} is such that it allows a principled, order-preserving mapping to the real numbers. This has the effect of externalizing orderings of individuals, aligning them across speakers and contexts to the fixed order of the number line. Such a mapping can come about in various ways.

a) Additive measure functions. An important class of measure functions is made up of those that are additive with respect to concatenation, meaning that the measure assigned to two individuals concatenated in the relevant way is the sum of their two individual measures. Height is a classic example: the height of two individuals stacked one on top of the other is the sum of their individual heights. Other dimensions with this property include weight, depth, width, length, volume, duration and arguably cost (while items are often cheaper if purchased in quantity, the fact that we recognize this as a discount is an indication that we perceive cost as inherently additive).

Additivity may be encoded via a constraint on the family of measure functions lexicalized by an adjective. We define a concatenation operation $\oplus$ on elements of the domain, and a corresponding addition operation $+$ on degrees on the scale. The constraint is then that in (12), where C is the set of possible contexts; a sample denotation for an adjective satisfying this constraint is (13).

(12)   **Additivity constraint:**
    $\forall a, b \in Dom$ and $c \in C$, $\mu^c_{DIM}(a \oplus b) = \mu^c_{DIM}(a) + \mu^c_{DIM}(b)$

(13) $[\![\text{tall}]\!]^c = \lambda d \lambda x. \mu^c_{HEIGHT}(x) \succeq d$,
      where $\forall c \in C$ and $a, b \in Dom$, $\mu^c_{HEIGHT}(a \oplus b) = \mu^c_{HEIGHT}(a) + \mu^c_{HEIGHT}(b)$

A scalar mapping satisfying additivity is a ratio scale. Such a scale can be given a numerical representation via a further structure-preserving mapping to the real numbers under ordinary addition. This allows the establishment of numerical measures, if some standard object is selected as the basis for a unit of measurement (e.g. an object serving as the meter standard). Furthermore, such numerical representations are unique up to multiplication by a positive constant k; this means in particular that the relation between the numerical values assigned to two individuals in one context cannot be reversed in another context. This invariant (up to multiplication) mapping to the number line forms the basis for objective judgments regarding statements about orderings, i.e., comparatives; this is consistent with our experimental findings for the additive-dimension adjectives *tall*, *short* and *expensive*. Thus additivity yields both aspects of objectivity that were discussed above.

A large number of gradable adjectives fall into this class, including most of those that compose with measure phrases (e.g. *wide/narrow*, *heavy/light*, *long/short*); we predict these too would pattern as objective with respect to orderings. But not all dimensions lexicalized by gradable adjectives are additive in the sense of height. Temperature is a prime example (Lassiter, 2011): the temperature of two bowls of soup poured together is not the sum of their two individual temperatures, but rather somewhere intermediate between them. Among adjectives that lexicalize non-additive dimensions, there are nonetheless at least two other possible sources of numerical mappings.

b) 'Natural' units. There are dimensions for which natural, speaker-external phenomena serve as the basis for numerical measurement systems. Two examples of this are time and temperature. In the case of time, the rotation of the earth and its orbit around the sun provide the basis for the units 'day' and 'year'; subdivision and concatenation of these units yield further units such as 'hour', 'minute', and 'week'. For temperature, the freezing and boiling points of water provide two anchor points on the scale, which can then be divided into equal increments, for instance by equal increases in the level of mercury in a thermometer.

Units derived in this way provide another sort of principled mapping from entities to numbers. This of course provides the basis for measure phrases (e.g. *20 years old*, *5 degrees warmer*); on the account proposed here, it also enables objective judgments about orderings. This is in line with our experimental findings for the adjectives *old* and *new* (temperature adjectives were not included in the experiment, but we would predict similar results for them as well).

c) Context-independent derived measure functions. A final set of adjectives that are objective with respect to orderings are those lexicalizing measure functions that can be built up from measure functions of the above two classes in a context-independent way. The dimension of fullness provides a good example: the degree of fullness of a container (say, a bottle or gas tank) can be expressed as the volume of its contents divided by its capacity, i.e., the volume it is able to hold. A half full tank, for example, is one whose contents have half the volume of its capacity.

This class can be defined as adjectives that satisfy the constraint in (14); as an example, the corresponding lexical entry for the adjective *full* is given in (15):

(14) **Context independent derived function constraint:**
$$\forall c \in C, \ \mu_{DIM}^c(x) = f(\mu_{DIM_1}^c(x), \mu_{DIM_2}^c(x), \ldots \mu_{DIM_n}^c(x)),$$
where $\mu_{DIM_1}^c, \ \mu_{DIM_2}^c, \ldots \mu_{DIM_n}^c$ are objective measure functions

(15) $[\![full]\!]^c = \lambda d \lambda x . \mu_{FULLNESS}^c(x) \succeq d,$

where $\forall c \in C$ and $x \in Dom, \mu_{FULLNESS}^c(x) = \dfrac{\mu_{VOLUME}^c(content(x))}{\mu_{VOLUME}^c(capacity(x))}$

As defined here, fullness is not strictly speaking additive: two half full glasses when placed to-gether do not produce a full glass, though their combined contents are the same as that of a single

full glass. But because degrees of fullness can be derived via a context-independent function of two additive measure functions, a principled mapping to numbers can nonetheless be derived. Thus we may have measure expressions (e.g. *20 % full*), and we correctly predict objective judgments regarding comparative statements (cf. the experimental findings for *full/empty*). Other dimensions in this class might be purity (defined as volume of impurities relative to total volume) and speed (distance traveled divided by duration). As we would expect on the account developed here, the corresponding adjectives allow numerical measures (*90% pure*, *5 kilometers per hour faster/slower*); we predict them also to elicit purely factual judgments of their comparative forms.

4.2. Sources of subjectivity

Having discussed factors that give rise to objective judgments regarding orderings, let us turn now to the nature of the measure functions lexicalized by adjectives whose comparative forms can be interpreted subjectively. Recall that the experiment reported in Section 2 identified two subclasses of such adjectives: those that also allow objective readings for the comparative, and those that do not. Here I propose that this difference corresponds to two distinct sources of ordering subjectivity. This proposal is similar to ones made by Kennedy (2013) and Bylinina (2014), but I will attempt use the experimental findings to shed new light on the relationship between the two factors.

d) Multidimensionality. In Section 3 I discussed the insight that adjectives exhibiting ordering subjectivity are multidimensional. Underspecification in or uncertainty about the component dimensions and how they should be integrated results in the potential for disagreement as to orderings. Take for example the pair *clean/dirty*. Intuitively, the degree of cleanness or dirtiness of an object is a function of the amount and type of dirt on it, perhaps in proportion to its size. But which sorts of dirt (broadly construed) we are concerned with, and how different sorts are weighted relative to one another, are matters of potential disagreement, and there does not seem to be a principled correct choice. On one way of making this more specific, shirt a might work out to be dirtier than shirt b, while on another equally valid choice, the reverse relation might obtain.

To formalize this, I build on Sassoon (2013) and Bylinina (2014) in assuming that adjectives of this sort are associated in each context c with a set of component dimensions $DIM_1^c, DIM_2^c, \ldots,$ DIM_n^c, whose measure functions are integrated by some function f^c. We have already seen something similar in the form of the lexical entry for *full*. But in that case, subjectivity did not arise, because both the component dimensions and the manner of their combination were fully specified. Ordering subjectivity arises when this requirement is relaxed, such that one or both of these components becomes context dependent. (16) specifies the form of such functions, and (17) gives a plausible if undoubtedly overly simplistic entry for *dirty* in this form.

(16) **Context-dependent derived function constraint:**
$$\forall c \in C, \mu_{DIM}^c(x) = f^c(\mu_{DIM_1^c}^c(x), \mu_{DIM_2^c}^c(x), \ldots \mu_{DIM_n^c}^c(x))$$

(17) $[\![\text{dirty}]\!]^c = \lambda d\lambda x.\mu^c_{DIRTINESS}(x) \succeq d,$

$\quad$ where $\forall c \in C$ and $x \in Dom,\ \mu^c_{DIRTINESS}(x) = \frac{\sum_{i=1}^{n} k^c_i \cdot \mu^c_{AMOUNT}(dirt^c_i(x))}{\mu^c_{SIZE}(x)}$

Note that the individual dimensions that underlie such entries may themselves be objectively measurable; subjectivity derives from the potential for variation in the choice of these dimensions and how they are combined. This precludes the creation of a numerical measurement system, and also allows for faultless disagreements regarding comparative statements.

As was seen in Section 2, many of the adjectives that allowed subjective interpretations for the comparative form could also be interpreted objectively (examples being *clean*, *dirty*, *smooth*, *rough*, *sharp* and *dull*). I see two possibilities for how this may arise. The first is that the two entities under comparison are so different with respect to the property in question that for any context $c \in C$, the measure function μ^c_{DIM} returns the same ordering. For example, there may be potential for disagreement as to how precisely dirtiness should be measured, but regardless of how we resolve the underspecification that is responsible for this, a shirt covered with oil stains must be evaluated as dirtier than one that is clean except for a small smudge of dirt near the hem. A second possibility is that subjects who gave 'fact' judgments did so in the belief that for the purposes at hand, there was in fact some correct way to measure, that is, some principled choice of measure function among the family μ^c_{DIM}, even if it may not be known (or knowable) to the speaker.

e) Judge dependence. The reader will note that up to this point, there has been no mention of a judge or experiencer parameter as contributing to ordering subjectivity. This is intentional: as noted earlier, the potential for variability in orderings does not seem limited to differences between speakers, but may persist in the judgments of a single speaker. Such variation is captured by (16), which ties the variation to a difference in contexts rather than specifically judges. We might ask, though, if this is sufficient, given the intuition that faultless disagreements – including disagreements regarding comparative statements – derive from the perspectives of different speakers.

Let us consider here the adjectives identified as purely subjective in the experiment. Among this group, most denote properties whose ascription depends necessarily on what McNally and Stojanovic (2015) refer to as the mediation of a sentient individual. These include value judgments (*good/bad*), aesthetic judgments (*beautiful/ugly*), taste judgments (*tasty*), experiential properties (*interesting/boring*) and internal states (*happy/sad*). These adjectives do not directly describe properties of objects and events in the world, but rather our perceptions of, judgments about and experience with the objective world, and as such, it seems plausible that their dependence on sentient intermediation be represented in their semantics. I thus follow the existing tradition of work on subjectivity in taking these to involve measure functions parameterized to a judge. Adapting for concreteness the relativist approach, we may represent this as follows:

(18) **Judge dependent measure functions:** $[\![\text{Adj}]\!]^{c;j} = \lambda d\lambda x.\mu^{c;j}_{DIM}(x) \succeq d$

(19) $[\![\text{beautiful}]\!]^{c;j} = \lambda d\lambda x.\mu^{c;j}_{BEAUTY}(x) \succeq d$

Importantly, the underspecification in (18)-(19) cannot be resolved in the same way as that in (16)-(17). The situation might be such that we can fix the dimensions and weights that determine an entity's degree of cleanness or dirtiness, resulting in objectivity with respect to orderings; but degrees of beauty as represented here are inherently dependent on the perspective of a judge, which cannot be eliminated. I thus hypothesize that the distinction we found between mixed predicates that allow both subjective and objective interpretations for their comparative forms and purely subjective predicates is largely that the former denote underspecified, multidimensional measure functions, while the latter are judge dependent. In this I differ from Kennedy (2013), who suggests that all types of adjectival subjectivity – including ordering subjectivity – might be reduced to 'dimensional uncertainty'.

To this point, we might ask whether adjectives of the judge-dependent sort should also be represented explicitly as functions of objective measure functions, and as explicitly multidimensional. Plausibly, this varies by adjective. *Salty*, for example, is experiential, but also allows objective readings, perhaps because it (in contrast e.g. to *tasty*) encodes a judge-dependent function of an objective measure function, salt content. And while properties such as beauty, interest and difficulty clearly depend on multiple facets of the external world for their attribution, not all of the corresponding adjectives pass accepted tests for multidimensionality (e.g. *interesting in every respect* vs. ??*tasty in every respect*). Space considerations do not permit me to go into these issues here; I refer the reader to Solt (2016a) for more in-depth discussion.

5. Accounting for Standard Type

In the previous section we saw that the nature of the measure functions lexicalized by a gradable adjective determines the availability of objective versus subjective interpretations of comparative statements. Here I will argue that the same factors play a role in determining the nature of the standard for the adjective's positive form.

My starting point is Sassoon's insight regarding the role of triviality in ruling out a potential scalar endpoint-based standard. Consider for example the dimension of height. Nothing of which we might reasonably predicate *tall* or *short* can have zero height. While a scale of height might itself include a zero point, nothing in the relevant domain will be mapped by the corresponding measure function μ_{HEIGHT} to that point. Thus a scalar zero point would be a trivial standard, in that it would not yield a meaningful partition of the domain; rather, every individual would be evaluated as tall, and none as short.

I propose this to be the general principle guiding the determination of standard type, as follows:

(20) For an adjective Adj lexicalizing measure functions μ^c_{DIM}, if for every context $c \in C$ there is neither a minimum nor a maximum degree d to which entities $x \in Dom$ may be mapped by μ^c_{DIM}, a relative standard will obtain. Otherwise, an absolute standard will be preferred.

The crucial point is now the following: virtually all dimensions that are additive with respect to concatenation pattern like height in disallowing zero mappings. Nothing whose weight we might ask about can have zero weight; no event whose duration might be at issue can have zero duration; and so forth. In fact, this is a consequence of one standard axiomatization of additive measurement, specifically of the Archimedean condition (Kranz et al., 1971; Lassiter, 2011). We then predict that the corresponding adjectives (e.g. *heavy, long, wide, deep*, and their antonyms) will have relative standards, and this is precisely what we find. Thus the same underlying factor that yielded objectivity with respect to comparison constructions also yields a relative standard.

Interestingly, there are exceptions to the generalization that adjectives lexicalizing additive dimensions disallow zero measures, prime examples being *expensive/cheap*: something can have zero cost without ceasing to be the sort of thing whose cost we might inquire about. Another example might be *likely* (see again Lassiter 2011). Yet these nonetheless have relative standards. Here I would like to propose that additivity alone is sufficient to align their interpretation to that of other additive-dimension adjectives, yielding a relative standard; this might be thought of in terms of evolutionary pressures of the sort discussed by Qing and Franke (2014).

Let us turn now to adjectives whose (families of) measure functions have a different form, in particular those that can be stated as derived measure functions (classes (c) and (d) in Section 4). Whether an absolute or relative standard obtains is dependent on whether there is a scalar minimum and/or maximum point to which relevant individuals can be mapped. Here it is instructive to consider the lexical entries for *full* (15) and *dirty* (17), which are based on division. From their form, it is clear that an individual can have zero degree of dirtiness or fullness without ceasing to exist as something whose dirtiness or fullness might be at issue; this is the case when the numerator is zero but the denominator is not. The mapping for *full* also has a maximum point (when contents=capacity), as presumably does that for *clean*, assuming its scale in each context c to be the reverse of that for *dirty*. In either case, an endpoint-based standard induces a potentially nontrivial partition on the domain, and is therefore to be preferred. Thus again, the form of the measure functions lexicalized by the adjective determines standard type. Many absolute gradable adjectives lexicalize measure functions of this form (e.g. *flat/bumpy, dry/wet, smooth/rough, safe/dangerous*); what other types might exist requires further investigation.

Note that the above-stated principle governing standard types is independent of the stable versus variable nature of the property in question (contra Toledo and Sassoon 2011). Temperature is a variable property, but as there is for practical purposes no maximum or minimum temperature an object may have, the corresponding adjectives (e.g. *warm, hot*) have relative standards. Conversely, *purity* is arguably an enduring property, yet a substance may at least in principle be completely pure, making an endpoint standard non trivial; we correctly predict *pure/impure* to be absolute.

This brief discussion has not exhausted all types of gradable adjectives (in particularly, nothing has been said about the standards for evaluative adjectives such as *beautiful*). But it should be sufficient to see that objectivity/subjectivity and standard type are connected in a meaningful way.

6. Conclusions

The experiment reported in this paper yielded the surprising finding that a wide range of gradable adjectives display at least some degree of subjectivity in the interpretation of their comparative forms. Such adjectives divided into two groups, those that also allow objective or factual readings for the comparative, and those that do not. Furthermore, the availability of objective versus subjective interpretations was found to correlate with the distinction between relative and absolute gradable adjectives. A theory was put forward in which both of these patterns derive from the formal properties of the measure functions lexicalized by different sorts of adjectives, with an important role played in both cases by functions that are additive with respect to concatenation.

The account outlined here leaves a number of important questions open. Most centrally, we might ask how distinct the three groups identified in the experiment truly are, and whether they might be further subdivided. Among dimensional adjectives such as *tall*, are subjective readings for the comparative absolutely ruled out (as suggested here), or might such readings be allowed in certain contexts (per Kennedy 2013)? Among the purely subjective group, might there be other sources of subjectivity than those discussed here? How exception-free are the generalizations made about standard type? The relation between ordering subjectivity – particularly among the 'mixed' group – and other characteristics of evaluative predicates such as embedding under *find* also merits further study. I leave these as topics for future research.

References

Bates, D. M., M. Mächler, B. Bolker, and S. Walker (2014). Lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. http://CRAN.R- project.org/package=lme4.

Bierwisch, M. (1989). The semantics of gradation. In M. Bierwisch and E. Lang (Eds.), *Dimensional Adjectives*, pp. 71–261. Berlin: Springer Verlag.

Bylinina, L. (2014). *The grammar of standards: Judge-dependence, purpose-relativity, and comparison classes in degree constructions*. Ph. D. thesis, Utrecht University.

Carlson, G. (1977). *Reference to Kinds in English*. Ph. D. thesis, University of Massachusetts.

Cresswell, M. J. (1977). The semantics of degree. In B. H. Partee (Ed.), *Montague Grammar*, pp. 261–292. New York: Academic Press.

Heim, I. (2000). Degree operators and scope. In B. Jackson and T. Matthews (Eds.), *Proceedings of the 10th Semantics and Linguistic Theory Conference*, Ithaca, NY, pp. 40–64. CLC Publications.

Hothorn, T., F. Bretz, and P. Westfal (2008). Simultaneous inference in general parametric models. *Biometrical Journal 50*(3), 346–363.

Kennedy, C. (1997). *Projecting the adjective: The syntax and semantics of gradability and comparison*. Ph. D. thesis, University of California at Santa Cruz.

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: An Interdisciplinary Journal of Philosophy 56*, 258–277.

Kennedy, C. and L. McNally (2005). Scale structure, degree modification and the semantics of gradable predicates. *Language 81*, 345–381.

Kranz, D. H., R. D. Luce, P. Suppes, and A. Tversky (1971). *Additive and Polynomial Representations*, Volume I of *Foundations of Measurement*. New York: Academic Press.

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

Lassiter, D. (2011). *Measurement and modality: the scalar basis of modal semantics*. Ph. D. thesis, New York University.

Lassiter, D. and N. D. Goodman (2013). Context, scale structure, and statistics in the interpretation of positive-form adjectives. In T. Snider (Ed.), *Proceedings of the 23rd Semantics and Linguistic Theory Conference*, pp. 587–610.

McNally, L. and I. Stojanovic (2015). Aesthetic adjectives. In J. Young (Ed.), *The Semantics of Aesthetic Judgment*. Oxford: Oxford University Press.

Pearson, H. (2013). A judge-free semantics for predicates of personal taste. *Journal of Semantics 30*, 103–154.

Qing, C. and M. Franke (2014). Gradable adjectives, vagueness, and optimal language use: A speaker-oriented model. In T. Snider, S. D'Antonio, and M. Weigand (Eds.), *Proceedings of the 24th Semantics and Linguistic Theory Conference*, pp. 23–41.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rotstein, C. and Y. Winter (2004). Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics 12*, 259–288.

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

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

Sassoon, G. (2012). A *slightly* modified economy principle: stable properties have non-stable standards. In E. Cohen (Ed.), *Israel Association of Theoretical Linguistics (IATL) 27*, pp. 163–182. MIT Working Papers in Linguistics.

Sassoon, G. (2013). A typology of multidimensional adjectives. *Journal of Semantics 30*, 335–380.

Schwarzschild, R. (2005). Measure phrases as modifiers of adjectives. *Recherches linguistiques de Vincennes 34*, 207–228.

Solt, S. (2016a). Multidimensionality, subjectivity and scales: experimental evidence. In preparation for E. Castroviejo, L. McNally, and G. Sassoon (Eds.), Gradability, scale structure and vagueness: experimental perspectives. Springer.

Solt, S. (2016b). On measurement and quantification: the case of *most* and *more than half*. *Language 92*(1), 65–100.

Stephenson, T. (2007). Judge dependence, epistemic modals, and predicates of personal taste. *Linguistics and Philosophy 30*, 487–525.

Stojanovic, I. (2007). Talking about taste: Disagreement, implicity arguments, and relative truth. *Linguistics and Philosophy 30*, 691–706.

Toledo, A. and G. W. Sassoon (2011). Absolute vs. relative adjectives - variance within vs. between individuals. In N. Ashton, A. Chereches, and D. Lutz (Eds.), *Proceedings of the 21st Semantics and Linguistic Theory Conference*, pp. 135–154.