

Moral adjectives, judge-dependency and holistic multidimensionality

Forthcoming in *Inquiry*

Andrés Soria Ruiz, IFILNOVA, Lisbon. andressoriaruiz@gmail.com

Federico L.G. Faroldi, Research Foundation – Flanders (FWO), Belgium/Lise Meitner Fellow (FWF), Austria. federico.faroldi@ugent.be

Abstract

In recent experimental work, the spectrum-like nature of the phenomenon of ordering subjectivity has been accounted for by recourse to the distinction, within the class of subjective adjectives, between multi-dimensional and judge-dependent ones. One way to cash out judge-dependency is in terms of some kind of experiencer-sensitivity. In this paper, we argue that this approach is insufficient. Applying Solt's (2018) experimental paradigm to moral adjectives (many of which are not experiential) suggests that, within the class of judge-dependent adjectives, one must draw a further distinction between (at least) experiential and evaluative adjectives. This opens up the question of what, if anything, characterises judge-dependency. We propose that judge-dependency is characterised by the notion of holistic multi-dimensionality: a predicate is holistically multidimensional just in case its denotation is composed by various dimensions whose contribution is inseparable.

judge-dependency; subjectivity; gradability; multidimensional adjectives

1. Introduction

Consider the following two dialogues:

(1) a. Carmen is tall.

b. No, she is not.

(2) a. The cake is tasty.

b. No, it is not.

In both cases, speakers appear to be having what is often called a subjective or

‘faultless’ disagreement.¹

It is widely accepted that this phenomenon arises because the positive form of gradable adjectives makes reference to a contextually determined threshold: the predicate ‘is tall [tasty]’ means something equivalent to ‘is at least as tall [tasty] as the accepted threshold’. A conflict arises because speakers do not agree on what the accepted threshold is (Kennedy 2013).

Now consider the contrast between these two dialogues:

(3) a. Carmen is taller than Marieke.

b. No, Marieke is taller.

(4) a. The cake is tastier than the chocolate cookies.

b. No, the chocolate cookies are tastier.

In both of these dialogues, speakers are also having a disagreement. But in this case, only the latter of these disagreement is subjective.

The property in virtue of which the comparative form of an adjective gives rise to subjective disagreement is called ORDERING SUBJECTIVITY (McNally 2017; Bylina 2017; Solt 2018; see also Silk 2019). Ordering subjectivity distinguishes adjectives like ‘good’, ‘beautiful’ or ‘tasty’ from dimensional adjectives like ‘tall’ or ‘wide’.

The purpose of this paper is to explore the distribution and source of this property. Specifically, we address the question of whether and to what extent ordering subjectivity should be equated with experiencer semantics, using moral adjectives as a test case. One salient way to account for ordering subjectivity is to link this property to some kind of experiencer-sensitivity. By looking at moral adjectives, which are ordering subjective but by and large *not* experiencer-sensitive, we reject an account of ordering subjectivity based on experience and propose that ordering subjective predicates are characterised by being HOLISTICALLY MULTIDIMENSIONAL. A predicate is multidimensional just in case it combines different dimensions in its meaning; a multidimensional predicate is holistic just in case the contribution of some of its dimension to its overall meaning is not separable.

To fix some more terminology, let us call a predicate EXPERIENCER-SENSITIVE just in case a disagreement involving that predicate can be accounted for by recourse to a difference in the experience or perception of the disagreeing parties. We can account for the disagreement in (4) by noting that each speaker can have different experiences of the taste of cake and chocolate cookies, resulting in their divergence. But we cannot say a similar thing about (3).

It bears distinguishing experiencer-sensitivity from two closely related properties of adjectives like ‘tasty’. First, whether these adjectives are experiencer-sensitive is different from whether they have a lexically specified *experiencer argument*. Experiencer arguments are prepositional phrases (PPs) headed by ‘to’/‘for’. In early discussions about predicates of personal taste (PPTs), these PPs were assumed to make

¹ See Glanzberg 2007; Kölbel 2003; Lasersohn 2005; MacFarlane 2014; Stephenson 2007; Stojanovic 2007 for classic discussions, as well as the papers in García-Carpintero and Max Kölbel 2008. See Rudolph 2020 as well.

reference to the semantic judge associated with these adjectives (Glanzberg 2007; Stephenson 2007). But more recently, Bylinina (2017) has argued convincingly that, even though these PPs denote experiencers and fill an argument slot for PPTs, predicates that lack this specialized argument slot may still be considered judge-dependent, such as ‘ugly’. We can add something stronger: not only are adjectives like ‘ugly’ judge-dependent, many of these adjectives are presumably experiencer-sensitive as well—even in the absence of a lexicalized experiencer argument.

Secondly, whether these adjectives are experiencer-sensitive is also connected with the question of whether they carry a so-called *acquaintance inference* (Pearson 2013; Ninan 2014; Franzén 2018; Willer & Kennedy 2019). An adjective carries an acquaintance inference just in case an unembedded sentence containing that adjective triggers the inference that the speaker is acquainted with whatever the adjective is predicated of. In most contexts, if a speaker utters ‘sushi is tasty’, this suggests that the speaker has tried sushi. PPTs and aesthetic adjectives are generally thought to carry this inference.² In light of this, it is suggestive to think that the class of experiencer-sensitive adjectives is largely co-extensional with the class of adjectives that carry an acquaintance inference, but these are ultimately separate issues.³

Solt (2018) presents novel experimental data about the distribution of ordering subjectivity.⁴ In the literature, it is often assumed that adjectives are clearly divided between those that are characterized by ordering subjectivity and those that are not. The main empirical discovery of Solt 2018 is that this is not the case: when one considers a sufficiently broad class of gradable adjectives, it turns out that speakers’ intuitions about the distribution of ordering subjectivity are not uniform. Her results show that people agree that adjectives like ‘tall’ lack this property and that adjectives like ‘tasty’ have it. But they show as well that many other adjectives fall in a mixed class, in that there is a large variability in participants’ opinions as to whether disagreement dialogues involving those adjectives in comparative form are subjective. Adjectives in this mixed class include ‘dull’, ‘curved’ or ‘rough’.

Our purpose is to extend this experimental paradigm to moral adjectives, which Solt does not consider. Moral adjectives are interesting *vis-à-vis* Solt’s experimental paradigm for the following reason: moral adjectives arguably involve a judge, but it is unclear to what extent this judge plays the role of an experiencer. Note that moral

² As a reviewer points out, some aesthetic adjectives such as ‘beautiful’ may not carry this inference, or at least they may fail to do so in some contexts. For example, it seems felicitous to say ‘The Hanging Gardens of Babylon were beautiful’ based on historical testimony only. We remain unsure about aesthetic adjectives—*pace* cases based on testimony, the consensus seems to be that they do carry this inference, but exploring this would take us too far afield.

³ The acquaintance inference is also discussed in the literature about aesthetic testimony, where authors debate whether one can gain aesthetic knowledge via testimony (Mothersill 1984: 159–160; Hopkins 2011; Robson 2012). However, see Robson 2015 for an argument that both sides of that debate can accommodate the acquaintance inference. We remain unsure as to whether our observations about the experiencer-sensitivity of aesthetic adjectives would have any bearing on this debate, so we leave this issue for future work.

⁴ See also Smith et al. 2015 for experimental data on taste predicates; as well as Liao, McNally, and Meskin 2016 and Liao and Meskin 2017 for recent empirical work on aesthetic adjectives. See Faroldi & Soria Ruiz 2017 for (theoretical) considerations about the scalar properties of moral adjectives.

disagreements do not necessarily depend on the disagreeing parties having different experiences. This is a reason to think that many moral adjectives are not experiencer-sensitive in the same way that, say, predicates of personal taste are (in addition, moral adjectives hardly carry an acquaintance inference: ‘eating sushi is morally wrong’ does not suggest that the speaker has tried sushi). If moral adjectives turn out to pattern with the clearly subjective class of adjectives, then we should enrich the traditional notion of a semantic judge to make room for a type of judge-dependency that is not experiential.⁵

Before moving on, let us point out that the relation between moral adjectives and experiencer-sensitivity should not be taken to hold without exception. On the one hand, moral adjectives are not the only ordering-subjective adjectives that are not experiencer-sensitive: ‘intelligent’ is a nonmoral, ordering-subjective adjective that does not seem to be experiencer-sensitive (a disagreement about whether, e.g., ‘Einstein was more intelligent than Newton’ does not seem to boil down to a difference in experience or perception); and it may be argued that some aesthetic adjectives are not experiencer-sensitive either. So the enrichment of the notion of judge-dependency proposed here may be called for independently of moral adjectives.⁶ Conversely, some moral adjectives, such as ‘(morally) outrageous’ are probably experiencer-sensitive.

As we will see, our results show indeed that moral adjectives fall in the clearly subjective class. Thus, we defend that the notion of judge-dependency has to be enriched to make room for the type of judge that is involved in the semantics of moral adjectives. Our proposal is that judge-dependent adjectives are characterised by giving rise to orderings that show a particular type of multidimensionality, which we call holistic multidimensionality. Roughly, the idea is the following: traditional accounts of multidimensional adjectives characterize the contribution of each individual dimension as separable (Bylina 2017; Sassoon 2013). Informally, this means that the contribution of each dimension to the overall ordering induced by the adjective can be isolated from the rest. We defend that this fails for judge-dependent adjectives. While it is intuitive to distinguish different dimensions that contribute to the meaning of a judge-dependent adjective, the contribution of at least some dimensions to the overall ordering induced by the predicate cannot be isolated. This is, we defend, what the ‘inherent human element’ of judge-dependent predicates consists in.

This paper is structured as follows: §2 presents Solt’s 2018 study. Subsequently, we present our study, which uses Solt’s paradigm to test a set of moral adjectives together with a sample of the adjectives that she tested (§3). We discuss these results in §4. There, we argue for the need to enrich the notion of judge-dependency in the way just discussed. §5 Concludes.

⁵ As suggested by a reviewer, we recognize that there might be some positions in the literature (cf. e.g. Nussbaum 1990) that, under a plausible construal of ‘judge’ and ‘experience’ would deny the claim that a moral judgment does not require experience. This is potentially quite contentious and requires spelling out carefully what is meant by ‘judge’, ‘judgment’, ‘moral’, and ‘experience’. However, that moral judgments are not necessarily dependent on the experiences of a judge seems plausible under broadly deontologist and consequentialists theories. In any case, that is a thesis on normative ethics and moral psychology that we aim to remain neutral about. Our focus is moral language, not moral judgment.

⁶ We thank a reviewer for pressing this point.

2. Solt's 2018 study

In her 2018, Solt tackles the question of the distribution of ordering subjectivity across gradable adjectives. Let us define ordering subjectivity as follows:

(Ordering subjectivity) A gradable adjective *G* is ORDERING SUBJECTIVE iff whether a sentence containing the comparative form of *G* is true or false is perceived to be a subjective matter.

That definition calls for at least two clarifications: first, what is meant by subjective? And secondly, why is the comparative relevant?

Different authors give different characterizations of subjectivity. For our purposes however, it's easiest to simply follow Solt's characterization. As we will see, Solt presents participants with disagreement dialogues and asks them whether they consider the issue to be a matter of opinion. This will be our criterion as well. If the answer to a certain question or issue is perceived to be a matter of opinion, it is a subjective matter; if it is not, then it is not a subjective matter. To illustrate: whether one can have red wine with tuna is a matter of opinion; whether the sun sets at 4:36:32pm on January 4th, 2019 at coordinates 55.953251,-3.188267 is not. Note that this a binary distinction: an issue either is or is not a matter of opinion.⁷

To the second question there is a short and a long answer. The short answer was given at the outset: the gradable adjectives that give rise to subjective disagreement in the comparative form are a subclass of those that give rise to subjective disagreement in the positive form: 'tall' and 'tasty' belong to the latter class, but only 'tasty' belongs to the former. The longer answer is as follows: in standard semantics for gradable adjectives, whether an adjective is subjective boils down to the question of whether its meaning is determined with the aid of some contextual parameter. Now, the main difference between the positive and the comparative form of a gradable adjective is that the positive form arises from the interaction of two semantic components: the lexical meaning of the adjective—the gradable property it denotes (height, taste or what have you) and, in addition, the feature that any individual's degree of that property has to

⁷ We should not presuppose that speakers are perfectly capable of drawing this distinction: as this recent Pew Research Center poll shows, people are surprisingly bad at distinguishing statements of opinion and statements of fact taken from the news (<https://www.journalism.org/2018/06/18/distinguishing-between-factual-and-opinion-statements-in-the-news/>). Moreover, people appear to consider more factual those statements that lean towards their values and opinions. Solt's and our study abstract away from this confound by presenting participants with items that they could not be antecedently opinionated about. For example, rather than presenting participants with dialogs that compare, say, the taste of cookies to the taste of stew, participants are presented with dialogs containing sentences such as 'this cookie is tastier than that one'. See Kaiser & Rudin 2020 for experimental data pertinent to this.

have for the positive form to be truthfully applied to it.⁸ The comparative, by contrast, arises from a single semantic component, namely the lexical meaning of the adjective.

Thus, if the positive form of an adjective is perceived to be subjective, it could be due to either semantic component: it can be due to speakers determining a different threshold or comparison class for the positive form, or due to the lexical meaning of the adjective being context-sensitive. By contrast, if the comparative form of an adjective is perceived as subjective, it has to be because its lexical meaning is context-sensitive. Therefore, asking whether a disagreement such as (3) and (4) is subjective is a way of investigating whether the ordering induced by the relevant adjective is context-sensitive. That is the question that Solt 2018 tackles.

2.1 Experimental design

To answer that question, Solt tests a set of gradable adjectives, classified along traditional categories:

- (1)RELNUM: Relative-standard with numerical measures:⁹ e.g., ‘tall’, ‘wide’;
- (2)ABS2: Absolute-standard with closed scales on both ends:¹⁰ e.g., ‘empty’, ‘full’;
- (3)ABS1: Absolute-standard with closed scale on one end: e.g., ‘salty’, ‘wet’;
- (4)RELNO: Relative-standard without numerical measures: e.g., ‘light’, ‘hard’;
- (5)EVAL: Evaluative adjectives, among them PPTs (‘tasty’), aesthetic (‘beautiful’) and emotional adjectives (‘happy’).

She presents subjects with disagreement dialogues such as the ones we have just seen, and asks subjects whether the disagreement they just saw is a matter of opinion. Here are a few of her test items:

- (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 lecture we heard last week was more boring than today’s lecture.
b. No, today’s lecture was more boring.

⁸ In degree semantics (Kennedy 2007), this is a specialized degree on a scale, called a THRESHOLD, such that the positive form of the adjective can be truthfully predicated of an object just in case its degree is equal to or above the threshold. In a delineation approach (Barker 2002; Burnett 2016; Klein 1980), the positive form requires introducing a contextually determined COMPARISON CLASS relative to which the extension, anti-extension and extension gap of the relevant adjective are determined.

⁹ Adjectives have numerical measures just in case they admit numerical modifiers, e.g., ‘Carmen is 168 cm tall’.

¹⁰ As Solt notes, ABS2 adjectives may be thought to have numerical measures as well, since they admit of proportional modifiers, e.g., ‘55% full’.

Below each dialogue, she presents participants with a choice between the following two options: ‘only one can be right, the other must be wrong’ and ‘it’s a matter of opinion’.

2.2 Results

The main result of Solt’s experiment is that participants’ perception of ordering subjectivity is largely variable. That is, for a large class of adjectives there is variability across participants as to whether the relevant dialogues are perceived to be subjective. While RELNUM and ABS2 adjectives were perceived as objective and EVAL adjectives were perceived as subjective for most participants, there were a surprisingly high number of adjectives for which participants’ judgments diverged, namely all adjectives in the ABS1 and RELNO classes.¹¹

2.3 Solt’s account

In the remainder of her paper, Solt offers an account of her results, that is, of why there is a class of clearly objective and subjective adjectives, as well as a big mixed class of adjectives. In general terms, her account of why adjectives fall into a clearly objective, mixed and clearly subjective class relies on the availability of ‘precise, quantitative measurement’ (Solt 2018, 67). Adjectives in the clearly objective class (RELNUM and ABS2) admit numerical modifiers, so it is reasonable for speakers to assume that those disagreements can be resolved objectively, by measuring the degree of the relevant property that each *relatum* has, and thus that only one of the speakers in those dialogues can be right. On the other hand, adjectives in the clearly subjective class (EVAL) are such that numerical measures are not available at all. Finally, adjectives in the mixed class (ABS1 and RELNO) can receive objective and subjective interpretations. Presumably, participants who interpreted adjectives in the mixed class subjectively answered that the disagreements were a matter of opinion, whereas speakers who interpreted those same adjectives objectively answered that only one speaker could be right.

The mixed class

Why would participants interpret adjectives in the mixed class sometimes subjectively, sometimes not? Solt offers a few diagnostics. First, some adjectives in the mixed class, such as ‘hard’, are ambiguous between a subjective/qualitative and an objective/quantitative interpretation (Kennedy 2013): whereas the most common interpretation of hardness is in relation to a certain perception (how hard a chair *feels*), there exist scientific measures of hardness for materials.

¹¹ What we mean by there being a large variability in participants’ opinion is that some participants took the relevant dialogues to be subjective disagreements, while some did not. But in both Solt’s experiment and ours, participants were (i) given a binary choice between describing a disagreement as a matter of opinion or fact, and (ii) presented with dialogues containing the relevant adjectives only once. So we cannot conclude whether participants are consistent in their judgments. The variability measure is across, and not within, participants.

Other adjectives are MULTIDIMENSIONAL, which means that their orderings depend on different dimensions or aspects. For example, the degree of dirt on a towel can be broken down into several dimensions: e.g., size of stained area and nastiness. This opens up two very salient aspects for cross-subject variability, namely the exact dimensions that are taken into account and the weight that each dimension receives in the overall dirt-ordering. So a disagreement about what towel is dirtier could be deemed subjective by someone who thinks that speakers engaging in that disagreement had different dimensions of dirtiness and/or were giving different weight to each dimension of dirtiness. Importantly however, it is possible too that some speakers consider that there is an objective measure of dirtiness—for example, how much extraneous material is attached to the towel’s surface. This would allow the needed variability for an adjective like dirty to fall in the mixed class.

Multidimensionality has been the subject of much work in recent literature on gradability. Especially relevant is the work of Sassoon (2013, 2016), (although see as well McNally & Stojanovic 2017). Sassoon uses as a grammatical diagnostic for multidimensional adjectives the admissibility of ‘respect-denoting’ PPs and quantificational phrases, such as ‘with respect to’ or ‘except for’. To see this, contrast a dimensional adjective like ‘tall’ with a multidimensional adjective like ‘healthy’:

(7) a. Carmen is healthy with respect to blood pressure.

b. # Marieke is tall with respect to...

(8) a. In what respects is Carmen healthy?

b. # In what respects is Marieke tall?

(9) a. Carmen is healthy except for her sugar level.

b. # Marieke is tall except for...

Interestingly however, many of the adjectives tested by Solt do not fare too well in these tests. Aside from the dimensional ones, which are predicted to be unacceptable, consider the following examples from the mixed class (Solt 2018, 74):

(10) a. The line was(n’t) straight/curved except for ...

b. The leather was(n’t) smooth/rough except for ...

c. The knife was(n’t) sharp/dull except for ...

d. The soup was(n’t) salty except for ...

For all those adjectives, it is difficult to fill the blank with a dimension of the relevant property. Nonetheless, suggests Solt, even if these adjectives do not clearly pass Sassoon’s tests for multidimensionality, we can distinguish different dimensions or aspects that can play a role in determining an ordering. For instance, when considering a curved road, we can take into account the frequency and (average) sharpness of the curves it has, even though it is very odd to say something like:

- (11) ?? The road was curved with respect to frequency, but not with respect to (average) sharpness.

The fact that we can conceptually distinguish different dimensions that may affect the ordering induced by an adjective even though the adjective may not appear multidimensional under Sassoon's tests, suggest a distinction between two types of multidimensionality:

- (1) QUANTIFICATIONAL MULTIDIMENSIONALITY: the type of multidimensionality detectable via respect-denoting PPs and quantification (e.g., 'healthy').
- (2) CONCEPTUAL MULTIDIMENSIONALITY: not detectable via respect-denoting PPs and quantification, but conceptually distinguishable (e.g., 'curved').

Solt's claim is that many adjectives in the mixed class count as conceptually multidimensional, if not quantificationally so.

The clearly subjective

Adjectives that clearly fall in the subjective class (i.e., those antecedently classed as EVAL), according to Solt, are those for which no numerical interpretation can be given. This correlates with the fact that they involve an 'inherent human element' (2018, 76), or as McNally and Stojanovic (2017, 28) put it, 'the necessary mediation of some sentient individual'. In the literature on PPTs, this inherent human element has usually been cashed out in the introduction of a context-dependent JUDGE parameter.¹²

For our present purposes, what matters is the *role* that the semantic judge can play in determining the ordering of individuals that forms the denotation of an adjective. Solt discusses various possible judge-roles (2018, 76). First and foremost, the judge could play the role of an experiencer. This applies most clearly to PPTs ('tasty'), but arguably to aesthetic adjectives as well ('beautiful'). Secondly, emotional adjectives ('happy') suggest that the judge can also play the role of a perceiver of emotional states or feelings (both in herself as well as in other individuals). Finally, Solt points out that many of these adjectives express value judgments, although this may or may not be an inherent component of the meaning of these adjectives (if it is however, it can also be subsumed under the 'inherent human element' of these adjectives).

Although Solt does not say this, it is intuitive to think that this latter judge-role need not involve experience, and thus that in its *evaluative* sense the 'inherent human element' is markedly different from a subject of experiences. Solt goes on to connect evaluativity to Hare's 1952 view that the meaning of 'good' is, at core, an attitude of *commending*, but she does not say much more about the evaluative role of the inherent human element. The evaluative component of the meaning of subjective adjectives will be crucial to us however, so we return to it in section 4.

¹² The judge parameter can either be incorporated as part of the proposition expressed by sentences containing the relevant expressions—so called CONTEXTUALISM about PPTs, cf. Sæbø 2009; Stojanovic 2007 a.m.o.—or as a parameter in the index of evaluation—the RELATIVIST proposal, cf. Lasersohn 2005; Stephenson 2007 a.m.o.

Setting aside the evaluative component, the judge-as-experiencer and the judge-as-perceiver (of emotions) have in common that the judge is a subject of certain experiences: the judge tastes a certain food; perceives certain arrangement of objects; feels or perceives certain emotions in herself or others. Given this, an intuitive account for why the class of clearly subjective adjectives cannot be given an objective interpretation is that the experiential component introduced by the judge can always affect the overall ordering. Hence, what matters for an ordering of taste is, ultimately, what the judge experienced as having a better taste; what matters for an ordering of ugliness is how the judge evaluates her perception of the relevant arrangement of objects; what matters for an ordering of happiness is what degree of happiness was felt or perceived by the judge; and so on. The judge has the last word regarding the ordering, so to speak.

Importantly, different dimensions may also play a role in determining the extension of these adjectives. Following Solt however, those dimensions do not determine the orderings of the adjectives in EVAL in a direct way, but rather act as factors that contribute and alter the judge's overall assessment (Solt 2018, 74-8). This would block putative objective or quantitative interpretations: for example, taste can presumably be broken into (at least) saltiness and texture, which are potentially measurable properties. However, if the saltiness of a food is inevitably no more than a factor in a judge's subjective experience of a food's taste, then it is difficult to see how saltiness could impact different judge's experiences of that food in a uniform way.

In sum: Solt's main result is that there is great variability in people's assessment of ordering subjectivity across adjectives, ranging from adjectives that clearly lack it (RELNUM and ABS2) to adjectives that clearly have it (EVAL) through a rather big mixed class (RELNO and ABS1). The mixed adjectives can be given subjective and objective interpretations, which is why participants diverge in their judgments. This, in its turn, is most likely due to either a qualitative/quantitative ambiguity in their meaning or their (quantificational or conceptual) multidimensionality. On the other hand, adjectives in the clearly subjective class are broadly judge-dependent. This means that their orderings are determined relative to a judge, broadly construed as a subject of experiences (and possibly of value judgments) that ultimately determine the ordering that corresponds to the relevant adjective.

3. Our Study

Given Solt's conclusions, it would be suggestive to see how moral adjectives fare in this experimental paradigm. As we said in the introduction, the main reason for this is that, even though moral adjectives arguably involve a judge, it hardly falls under the judge-roles that we just discussed—that is, the semantic judge of a moral adjective need not, in general, be an experiencer or feel or perceive any particular emotion.¹³ The notion of the judge as evaluator, briefly discussed by Solt, does seem more appropriate for moral

¹³ Another, potentially motivating thought is that one might predict at least some people to have strongly objective intuitions about morality. After all, a disagreement about what course of action was morally more correct might seem *prima facie* more objective than a disagreement about what movie was more fun. We anticipate that this prediction was not borne out. See Stojanovic 2019 for the distinction between disagreement about morality and about taste.

adjectives, so inquiring whether moral adjectives are ordering-subjective may be thought of as an indirect way of assessing whether the judge-as-evaluator has anything to do with subjectivity. Thus, in this experimental study we set out to test how moral adjectives fare with respect to Solt's experimental paradigm. To do that, we took a sample of adjectives from Solt's experiment, added an approximately equal number of moral adjectives, and tested them under Solt's disagreement paradigm.

3.1 Participants

Participants were 40 native speakers of English, recruited via Amazon Mechanical Turk (MTurk). They were paid \$0.50 for their participation (the task took approx. 5 minutes). Recruiting was limited to MTurk workers with US IP addresses. No participant was excluded.

3.2 Materials

Test items were based on 24 adjectives. From these, 13 were sampled from Solt's study:

- (1)RELNUM: 'tall', 'expensive';
- (2)ABS2: 'empty', 'full';
- (3)ABS1: 'salty', 'wet';
- (4)RELNO: 'light', 'hard';
- (5)EVAL: 'tasty', 'ugly', 'happy', 'intelligent'.

These were supplemented with 11 moral adjectives, classified along the following axes:

- (1)THIN moral adjectives: 'moral', 'ethical', 'virtuous'
- (2)THICK moral adjectives: 'coward',¹⁴ 'generous', 'loyal', 'honest'
- (3)NORM/VALUE: 'important', 'justified', 'rational', 'valuable'

THIN moral adjectives are those that carry exclusively evaluative meaning; THICK moral adjectives, by contrast, have both an evaluative and a descriptive component. While 'moral' is simply a term of evaluation, it is often thought that the meaning of 'generous', for instance, can be factored out into a description (giving beyond what is expected) and an evaluation of whoever satisfies the descriptive component (giving beyond what is expected is good).¹⁵ The class called NORM/VALUE is a heterogeneous class of all-purpose value and normative adjectives: 'important' is usually applied to things that

¹⁴ 'Coward' is both an adjective and a noun. Perhaps more common is the adjective 'cowardly', as a reviewer notes.

¹⁵ For the distinction between thin and thick adjectives, see Cepollaro and Stojanovic 2016; Kyle 2013; Roberts 2013; Väyrynen 2013, 2017.

range from subjective to objective value; ‘justified’ or ‘rational’ are used indistinctly for moral, practical or epistemic praise; and ‘valuable’ is an umbrella evaluative adjective that can be used to evaluate all kinds of things in all kinds of ways. It bears mentioning that no particular pre-test or criterion was applied in order to select these categories of moral adjectives, nor to select the particular examples that were used. The purpose of choosing both THIN and THICK adjectives, on the one hand, as well as the adjectives that we label NORM/VALUE, on the other, was simply to offer a varied sample of adjectives.

For each adjective, a disagreement dialogue was created. For the set of adjectives taken from Solt’s study, the same dialogues were used; for the class of moral adjectives, new dialogues were devised. Here is a sample of the new dialogues:

(12) a. Bill’s friends are more loyal than Mary’s.

b. No, Mary’s friends are more loyal.

(13) a. Ann’s work is more valuable than Jim’s.

b. I disagree, Jim’s work is more valuable.

(14) a. It is more important to follow Mary’s than Bill’s advice.

b. I disagree, Bill’s advice is more important.

Participants were presented with 11 test items and 13 control items, as well as 12 filler dialogues split between factual (a: ‘Sharks are mammals’; b: ‘No they are not!’) and subjective disagreements (A: ‘This restaurant has wonderful service, I love it’; b: ‘No, it’s awful’). See the Appendix for the full list of critical items.

3.3 Procedure

Participants were presented with the following set of instructions:

This study is about disagreements between people. Sometimes when two people disagree, only one of them can be right, and the other must be wrong. For example, in this short dialogue, Speaker A and Speaker B can’t both be right, because Rosa can’t have been born in both July and April.

- Speaker A: Rosa was born in July.
- Speaker B: No, Rosa was born in April.

But sometimes when people disagree, there is no right or wrong answer — it’s just a matter of opinion. Here’s an example:

- Speaker A: Susan looks a lot like her sister.
- Speaker B: No, they don’t look alike at all!

In this HIT, you will see a series of short dialogues between two speakers, A and B. Your task is to say whether there is a right or wrong answer, or whether it's a matter of opinion. Please answer based on your intuitions; do not think too long about each question. Do not proceed with this experiment if you are not a native English speaker.

Participants were then presented with dialogues like (12)-(14) above, and told to choose among the following two options:

What do you think?

1. Only one can be right, the other must be wrong.
2. It's a matter of opinion.

Answering 1 was classified as a FACT answer; answering 2 was classified as OPINION. At the end, participants were asked for their country, age, biological sex and were given the opportunity to comment.

3.4 Results

The proportion of FACT choices per Adjective is illustrated in **Figure 1**; the proportion of FACT choices per Adjective Class is illustrated in **Figure 2**. All the analyses can be consulted at: https://osf.io/5q6gc/?view_only=9d9fe894fd454b09a3caf719c3409914.

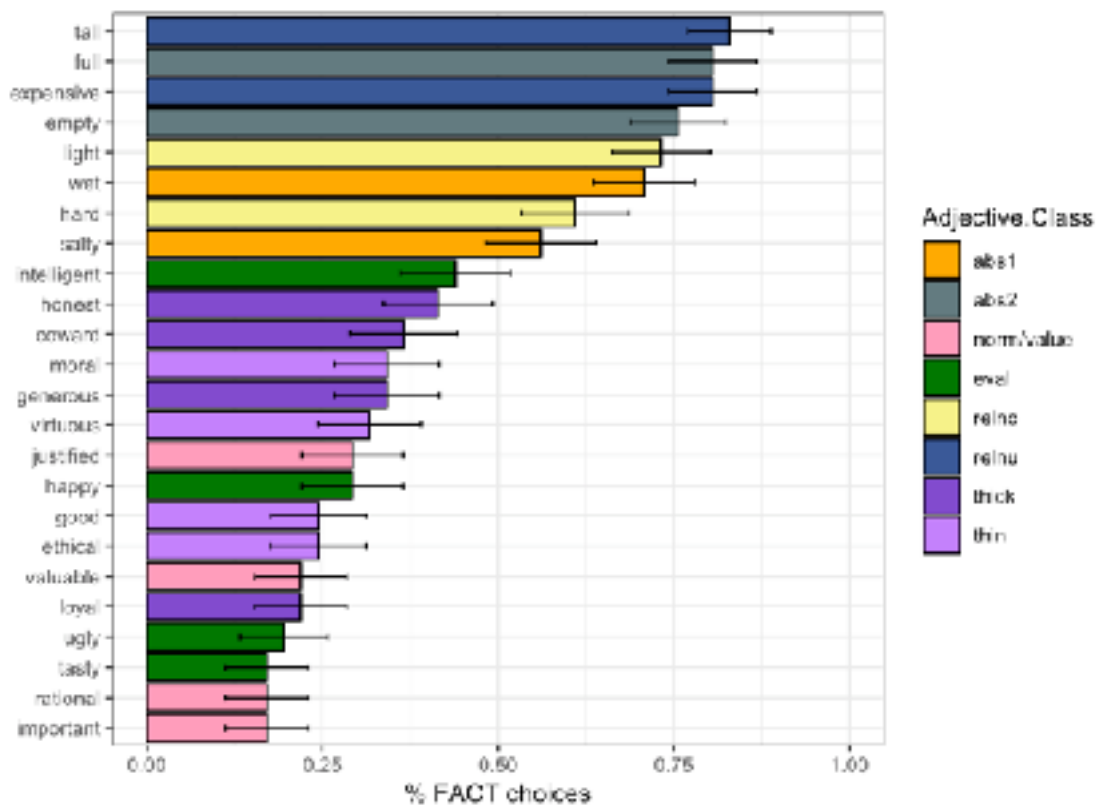


Figure 1. FACT choices per adjective.

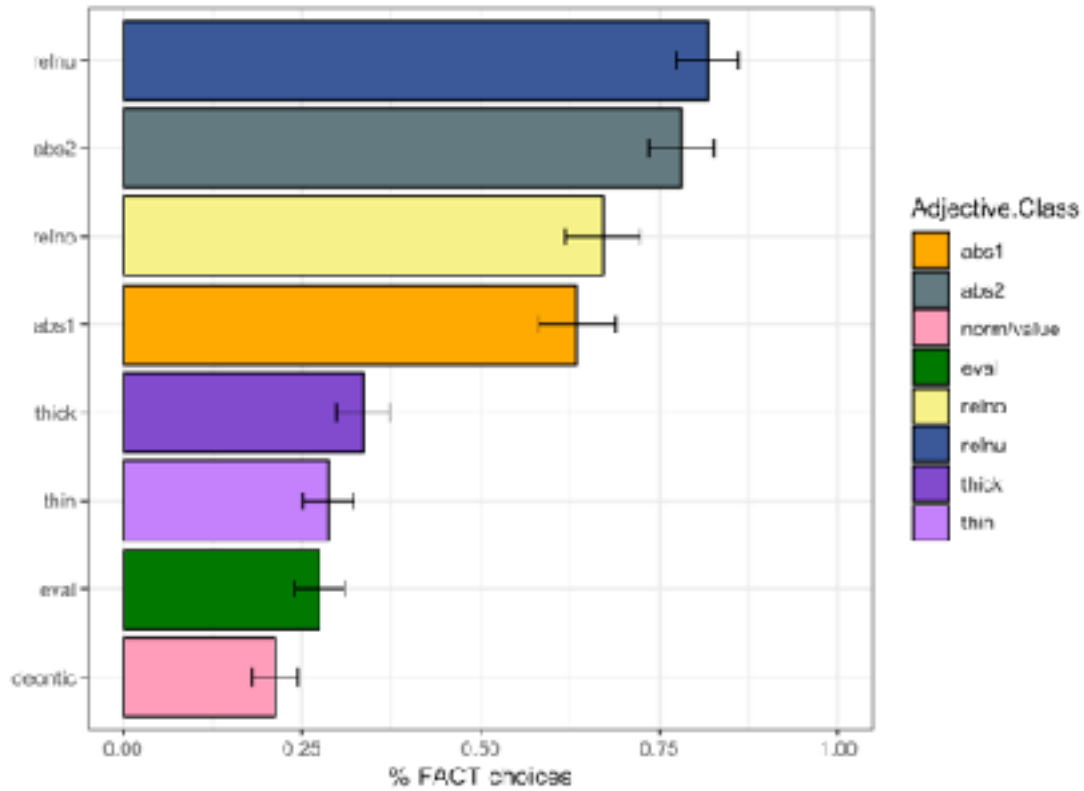


Figure 2. FACT choices per adjective class

The reference level for this omnibus model (that is, the baseline in treatment contrast) was the class RELNUM. The z-scores and p-values reported are those calculated by the lme4 package by a Wald III test.

The results of this omnibus model indicate that all classes are significantly different from RELNUM, with the exception of ABS2. Model output is provided in **Table 1**. These results roughly replicate Solt’s original findings, modulo the new classes added in our experiment.¹⁶

	RelNum (intercept)
Abs2	$z = -0.62, p = 0.53$
Abs1	$z = -2.82, p < .01$
RelNo	$z = -2.31, p < .05$
Eval	$z = -8.13, p < .001$
N/V	$z = -8.958, p < .001$
Thin	$z = -7.95, p < .001$
Thick	$z = -7.3, p < .001$

Table 1. Results of Omnibus model

¹⁶ We have also reproduced the posthoc pairwise comparisons done by Solt (2018) (ABS1 vs. EVAL: $z = -610, p < .001$, ABS1 vs. RELNO: $z = .5, p = .5$, EVAL vs. RELNO: $z = -5.8, p < .001$).

In order to address our main question (i.e., where do moral adjectives fall in the subjectivity spectrum), we fit a second set of models where we compared each of the classes included in Solt 2018 to the new sets of adjectives (THIN, THICK AND NORM/VALUE). These models were constructed in the same way as before (fixed and random structure) differing only in how the baseline was encoded.

- ABS1 (reference level) is significantly different from NORM/VALUE ($z = -7.240$, $p < .001$), THICK ($z = -5.3$, $p < .001$) & THIN ($z = -6.054$, $p < .001$).
- RELNO (reference level) is significantly different from NORM/VALUE ($z = -7.71$, $p < .001$), THICK ($z = -5.9$, $p < .001$) & THIN ($z = -6.42$, $p < .001$).
- EVAL (reference level) is not significantly different from NORM/VALUE ($z = -1.78$, $p = .07$), THICK ($z = 1.71$, $p = .08$) & THIN ($z = .35$, $p = .072$).¹⁷

4. Breaking down judge-dependency

Our results can be summarized in two claims: on the one hand, we replicated Solt's result for the classes that she tested, which is good. On the other hand, the adjectives that we tested (which we may simply call *moral adjectives*) were found to pattern with the adjectives in the EVAL class. Our claim in this section is that including moral adjectives into the EVAL class ought to alter any characterization of this class of adjectives that relies exclusively on an experiential notion of judge-dependency. In particular, we propose that the experiential notion of judge-dependency, which is by and large unsuitable for moral adjectives, be substituted by a notion according to which judge-dependent predicates are *holistically multidimensional*.

4.1 The insufficiency of experience

Our results strongly suggest including moral adjectives in the EVAL class. *Prima facie*, this is supported by Solt's general characterization of the EVAL class as those adjectives whose orderings are determined, or mediated by some 'inherent human element'. So much seems true of moral statements—that their truth inherently depends on a human element; and that if that human element changes, their truth value can change. However, recall that Solt characterized the 'inherent human element' mainly as a subject of experiences (for taste and aesthetic judgment) or as a perceiver of emotions (for emotional predicates). The problem is that, whereas taste, aesthetic or emotional disagreements may very well boil down to a divergence in experience or perception, that does not apply straightforwardly to moral disagreement.

Consider the contrast between dialogues (15)-(17) and (18):

(15) a. The cake is tastier than the chocolate cookies.

¹⁷ A posthoc analysis showed the class of Norm/Value to be significantly different from both the Thin and Thick classes (vs. Thick: $p < .001$; vs. Thin: $p = .03$), but it did not show the latter classes to be significantly different from each other ($p = .19$).

- b. No, the chocolate cookies are tastier.
- (16) a. Those sneakers are uglier than the sneakers you tried on earlier.
- b. No, the earlier sneakers were uglier.
- (17) a. This cat is happier than that dog.
- b. No, the dog is the happier one of them.
- (18) a. Bill leads a more ethical lifestyle than Amy.
- b. No, Amy does.

In all these disagreements, an OPINION answer (i.e., thinking that the issue is a matter of opinion) could be driven by considering that each speaker is a different judge with a different—and valid—opinion about the matter. Moreover, this may be rooted on speakers having different experiences, perceptions or feelings about the matter: in (15), speakers have both tried both foods and their tastes diverge; in (16) they have different aesthetic appreciation of the sneakers; and in (17) they have a different ‘emotional appraisal’ of these pets.

However, none of this applies straightforwardly to (18). To have a moral disagreement, it is not necessary to have different experiences or perceptions. Speakers in (18) need not have experienced in any way how Bill and Amy behave, nor do they need to see what emotions they appreciate in their behavior, nor what emotions Bill and Amy’s behavior causes in them. They are judging the matter differently, and people take them to be entitled to their divergent opinion (as witness our results). But it is not clear what exactly could their divergence be about.¹⁸

But if moral adjectives are judge-dependent, as they seem to be, then we have either of two options: first, we can simply assume that two broad notions of a judge play a role in the semantics of these adjectives. One is the usual notion according to which a judge is a subject of experiences (including feelings and perceptions), and the other one is whatever notion is required to account for moral talk. This might have been what Solt has in mind—recall that, besides understanding the judge in terms of experiences or perception, she recognises that the judge can play a different role as evaluator. However, this idea is not developed further and she also entertains the possibility that the evaluative component isn’t inherently part of the meaning of ordering subjective

¹⁸ Perhaps this applies to (16) as well: after all, it isn’t clear that two people looking at the same two pairs of sneakers could have a different perception of them. Consider, in this respect, the contrast with PPTs: in (15), the experience upon which the evaluation rests can differ from speaker to speaker. You and I may perceive the taste of sugar differently, and that can interfere with our perception of the taste of chocolate in different ways, let’s say. But in the case of an aesthetic evaluation, the perceptual experience upon which the evaluation rests is shared. In (16), there is a clear sense in which people are seeing the same thing. In light of this, there is some reason to drop the assumption that aesthetic adjectives belong in the experiencer-sensitive class, and perhaps we should consider them closer to moral adjectives than to PPTs and emotional predicates (this is in line with the possibility, mentioned in the Introduction, that some aesthetic adjectives may not carry an acquaintance inference). We leave this for future consideration.

adjectives. Besides this, a more general issue with this sort of pluralism is that one may be left wondering what do the various notions have in common, that is, what ties the ‘inherent human element’ together. So rather than assume a pluralistic view, we will finish this paper by formulating and arguing for a hypothesis about what characterizes judge-dependency. This hypothesis is aimed to cover both the usual members of the EVAL class as well as its putative new members.

4.2 Holistic multidimensionality

Recall Solt’s distinction between QUANTIFICATIONALLY and CONCEPTUALLY MULTIDIMENSIONAL predicates. Quantificationally multidimensional predicates are those whose dimensions can be quantified over and accessed via dimension-accessing operators (such as ‘with respect to’). These include adjective like ‘healthy’ or ‘sick’. Conceptually multidimensional predicates do not have dimensions that can be quantified over and are accessible via dimension-accessing operators. However, we can distinguish, on a conceptual level, different dimensions that contribute to their meaning. These include some of the adjectives in Solt’s mixed class, e.g., ‘dirty’ or ‘curved’.

Our proposal is that judge-dependent predicates are characterised by a third type of multidimensionality, which we call HOLISTIC MULTIDIMENSIONALITY. The intuitive idea is very similar to what Solt says about the ‘inherent human element’ in judge-dependent predicates: even if we can distinguish different dimensions contributing to an overall ordering, these dimensions are at most a *factor* that a judge can take into account. But the judge has the last word in determining an ordering.

But what does ‘having the last word’ mean? We propose that the sense in which the judge has the last word can be cashed out by turning to the notion of FACTOR SEPARABILITY from value theory.

An important topic in value theory is how to think of the different factors that can contribute to a value function (Oddie 2001a; Oddie 2001b). In this literature, it is useful to consider whether, given a certain value function P and factors X, Y that contributes to P , X is separable from Y . Where X and Y are sets of individuals (predicate extensions), factor X is (strongly) separable from factor Y iff

$$\forall x_1, x_2 \in X \text{ and } \forall y_1, y_2 \in Y, (x_1, y_1) >_P (x_2, y_1) \leftrightarrow (x_1, y_2) >_P (x_2, y_2)$$

Intuitively, this means that, if a certain P -relation between different elements of factor X obtains holding factor Y constant, that same relation obtains for whatever other constant value we assign to factor Y .

Suppose, for instance, that we are considering the size of a city. Intuitively, we can factor out city size in surface area and population. Now, if these factors are separable, then, if it is the case that city a is larger than city b and they have the same population, then it remains true that a is larger than b under any other uniform assignment of population. Suppose that city a has more surface area than city b . And suppose that in one case a and b have 10,000 inhabitants and in another case they have 1,000,000 inhabitants. The basic idea behind separability is that, if population is separable from surface area, then a is larger than b in the former case just in case a is

larger than b in the latter case. In other words, holding fixed the areas of a and b there is no amount of population that a and b could have such that a would not be larger than b .

We believe that the same holds of quantificationally and conceptually multidimensional predicates, but not of judge-dependent predicates. The former two types of multidimensional predicates are such that all of their dimensions are separable under any interpretation. Judge-dependent predicates, by contrast, are such that this fails: under at least some interpretations, at least some of the dimensions that we can intuitively think contribute to their meaning are not separable in this technical sense. This is, we submit, what the ‘inherent human element’ in judge-dependent predicates boils down to.

We will not attempt to defend a formal characterisation of separability for multidimensional predicates. Instead, we will finish this paper by considering, first, how separability intuitively holds for quantificationally and conceptually multidimensional adjectives. Reflection on how the intuitively distinguishable dimensions that contribute to these adjectives’ meaning behave suggests that they are separable, but it might be possible to come up with other dimensions that are not. On the other hand, we will show that, for judge-dependent predicates, separability clearly fails. We will do this by considering examples of how an overall relation that holds between a pair of objects in virtue of a uniform assignment of degrees of a certain contributing dimension can cease to hold when we shift that assignment (uniformly of course) in the appropriate way.

We already gave an example of a quantificationally multidimensional adjective whose dimensions are separable, namely ‘large’ (note that constructions such as ‘large with respect to size but not population’, ‘large in all respects’, etc., are acceptable). Let’s consider two examples of conceptually multidimensional adjectives, ‘curved’ and ‘dirty’.

First, consider ‘curved’: suppose that two dimensions of curved are frequency and (average) curvature or sharpness of curves. Road a is more curved than road b , and they have the same average curvature. Then, according to the separability thesis, there is no degree of curvature such that a and b could both have it and a not be more curved than b (the same goes for frequency). This seems intuitively true: regardless of how sharp the curves in each road are, if given a certain constant degree of curvature road a is more curved than road b , then however we change a and b ’s degree of average curvature, a will still be more curved than b .

Secondly, suppose that the dimensions of ‘dirty’ are stain nastiness and stained surface. Now, suppose that shirt a is dirtier than shirt b , and they have the same area covered by stains, say 3%. Now, if nastiness is separable from stained surface, then for any other possible degree of stained surface, as long as a and b have that same degree, a will be dirtier than b . That is, even if 90% of both shirts is stained, a will still be dirtier than b (in virtue of the fact that the stains in a are nastier, of course). Again, that seems intuitively correct.

Let us turn now to judge-dependent predicates. We will show that separability intuitively fails for some of these adjectives, showing that some of its dimensions, at least under some interpretations, are not separable. We take this argument to be conclusive for each of these adjectives (under the assumption that the dimensions to be distinguished really are dimensions of their meaning). But there is, in addition, a closely

related induction to make, which is that *all* the adjectives in the EVAL class have inseparable dimensions, and are thus holistically multidimensional.

Consider ‘tasty’: suppose that two dimensions of taste in coffee are bitterness and sweetness. Suppose that, without any sugar, coffee *a*, a mellow Colombian coffee, is tastier than *b*, a mighty Portuguese *bica*. The Colombian is fantastically balanced; the *bica* is too bitter. But now, suppose that we start adding the same amount of sugar to both coffees. At some point, it is perfectly intuitive to suppose that the Colombian coffee will be too sweet, while the *bica* will be just right, its bitterness compensated by the sweetness of the sugar. Then, it is no longer the case that *a* is tastier than *b*. Rather, *b* is tastier than *a*.

Next, take an aesthetic adjective, e.g., ‘beautiful’. Suppose that two dimensions of beauty (for cars) are design and color. Suppose, moreover, that design can range from completely old-fashioned to contemporary; and that color can range from classic (black, brown, grey, etc.) to more flashy tones (pink, green, bright red, etc.). Suppose that, for two cars *a*, *b* that have the same (classic) color, *a* is more beautiful than car *b* in virtue of the fact that the design of *a* is more old-fashioned than the design of *b*. But when we change both colours to a flashy color, then suddenly the car with the more contemporary design, that is *b*, turns out to be more beautiful overall.

Finally, let us consider now the case of moral adjectives, starting with a thin one. Suppose, that two dimensions of ‘ethical’ (for action types) are promoting virtue and producing happiness. According to some theories of value aggregation (Bader 2017; Oddie 2001a) it is possible to think that two equally virtuous actions *a*, *b* are such that *a* is more ethical than *b* in virtue of the fact that *a* produces more happiness to the agent than *b*. However, if we remove virtue altogether, then producing happiness turns out to be worse than not producing it (as it is *undeserved* happiness), and so *a* turns out to be worse than *b* in the absence of virtue.¹⁹

Let’s consider now a thick term, ‘generous’. Two dimensions or aspects along which one can be generous are time and money spent (of course, there might be many other possible dimensions of generosity that we are not taking into account). Focusing on these two dimensions, the following example can show that their contribution to an overall ordering of generosity is inseparable: suppose that we organise a fundraiser, to which we invite an idle millionaire as well as a very busy politician. Suppose that the idle millionaire donates 5.000 €, while the politician donates 4.000 €, and they both stay in the event for 5 minutes. Given this, one would say that the millionaire was more generous than the politician. But it seems intuitive to think that, given that the politician has a very busy agenda while the millionaire doesn’t, if they were both to stay, say, an hour, then we would perhaps say that the politician was more generous than the millionaire. This suggests that ‘generous’, at least if time and money are acceptable dimensions that contribute to its meaning, is also holistically multidimensional.

¹⁹ Note that even if one endorsed a monist theory of (intrinsic) value, there would still be a question of what factors are relevant to it. Take a fairly common interpretation of J. S. Mill, for whom the only (intrinsic) value is happiness. Happiness is not simple, but can be obtained via "higher" and "lower" pleasures. While this is only indicative, the paper only puts forward a semantic analysis, without committing to any specific value theory. We thank a reviewer for pointing out this issue.

And finally, consider a member of the NORM/VALUE class, 'justified'. When we assess whether a decision is justified, there are different dimensions to take into account. One dimension of justification is whether the decision leads to action (*commission*) or inaction (*omission*). All things equal, omission is more justified than commission. A different dimension of justification is whether a decision leads to harm or benefit. All things equal, a beneficial decision is more justified than a harmful one. These two dimensions of justification seem inseparable: suppose that two decisions *a* and *b* have equally harmful consequences, but whereas *a* is an omission, *b* is a commission. Normally, *a* is more justified than *b*. For example, assuming that death is harmful, letting someone die is more justified than killing them. But now suppose that two decisions *c* and *d* have equally beneficial outcomes, and again *c* is an omission, while *d* is a commission. In this case, *d* is more justified than *c*. For example, making someone a present is more justified than letting someone else make it for you, all other things being equal. Thus, 'justified' has inseparable dimensions, and is therefore holistically multidimensional.

We submit that all of these examples show that some dimensions involved in judge-dependent predicates are non-separable, and thus that those adjectives are holistically multidimensional. Bear in mind, before finishing, three important points: first, inseparability does not entail that there is no operation of aggregation of different dimensions for a judge-dependent predicate.²⁰ Rather, separability is a constraint on the class of possible aggregation operations that a multidimensional adjective has. If a multidimensional adjective has separable dimensions, then the class of aggregation operations that may contribute to its meaning is constrained in the way just described. If a multidimensional adjective is not separable, its aggregation operation is not constrained in this way, and there might be aggregation operations such that the contributing dimensions are inseparable.

Secondly, whether a multidimensional adjective has separable dimensions is independent of whether the operation that aggregates such dimensions is context-sensitive: adjectives with separable dimensions might still have context-sensitive aggregation functions, so long as these satisfy the constraint of separability. Conversely, holistically multidimensional adjectives may also have context-sensitive and context-insensitive measure functions. Separability constrains the menu of aggregation functions, but not completely. If it did, then separability would entail that the aggregation operation for adjectives with separable dimensions is context-insensitive. Not only is that not the case, as separability and context-sensitivity of aggregation operations are logically independent properties, but in addition, assuming such link would be at odds with one way in which Solt (2018, p. 81-2) for example, makes room for subjective interpretations of quantificational- and conceptually multidimensional adjectives, namely by allowing their aggregation function to be context-sensitive (in addition to their choice of dimensions and respective weights). Solt illustrates this with 'dirty', which—according to us—is separable, but—according to her—is context-

²⁰ Although it does establish that the aggregation, if any, cannot be *additive*, given certain background assumptions.

sensitive at the level of its aggregation operation.²¹ There is no tension here: ‘dirty’ can be separable while at the same time having an aggregation operation that is contextually underspecified, so long as all its admissible values are separable. Again, separability does not entail a context-insensitive aggregation operation; it merely introduces a constraint that limits the choice of this operation to well-behaved, separable functions. It is only by explicitly requiring (some of) the underlying dimensions to be not separable that we can capture holistic multidimensionality.²²

Thirdly, this does not mean that judge-dependent predicates are non-separable *under every interpretation*—there might be available interpretations of these adjectives, and possibly choices of dimensions as well, relative to which these come out as separable. However, the results of our experiment do suggest that those interpretations might not be the most salient or available.

The relationship between separability and subjectivity is the following: since separability greatly restricts the choice of aggregation functions for a multidimensional adjective, adjectives that respect separability have a more constrained meaning, and therefore, will appear much less subjective, than those that admit inseparable dimensions. Conversely, inseparability implies more room for subjective disagreement. Deciding on the aggregation procedure when the dimensions are inseparable, we contend, is the role that a judge plays.²³ The picture of a judge that emerges from this account, contrary to other notions of a semantic judge, is not experiential, and it explains why moral adjectives pattern with the clearly subjective class.

If the most available interpretations of these adjectives give rise to non-separability, then we have some reason to conclude that we have arrived at a simple and novel way of characterizing judge-dependency: non-separability, or as we have called it, **HOLISTIC MULTIDIMENSIONALITY**. This invites the following enrichment of Solt’s classification of multidimensional adjectives:

- (1) **QUANTIFICATIONAL MULTIDIMENSIONALITY**: detectable via respect-denoting PPs and quantification; conceptually distinguishable; separable (e.g., ‘large’, ‘healthy’);
- (2) **CONCEPTUAL MULTIDIMENSIONALITY**: not detectable via respect-denoting PPs and quantification; conceptually distinguishable; separable (e.g., ‘curved’, ‘dirty’).

²¹ We might mention, however, that even though Solt claims that ‘dirty’ is a multidimensional adjective with a context-sensitive aggregation function, the lexical entry that she gives for ‘dirty’ (Solt 2018, p. 82) does not seem underspecified at the level of the aggregation procedure (which combines in a non-context-, non-judge-dependent way sum, multiplication, and division), but only at the level of specifying the values of the variables (i.e., how the amount of dirt and the size of the stains are to be measured).

²² Note that the usual technical independency, cardinality and continuity assumptions are required to avoid a collapse onto separability.

²³ As seems sensible, we keep fixed the dimensional factorisation. In particular, it would be easy to establish non-separability if the dimensions were logically not independent. Cf. Bader 2017.

- (3) **HOLISTIC MULTIDIMENSIONALITY**: not detectable via respect-denoting PPs and quantification; conceptually distinguishable; non-separable (e.g., ‘tasty’, ‘beautiful’, ‘ethical’, ‘generous’...).

We conclude that judge-dependent predicates are not characterised by the fact that disagreeing speakers can have different experiences, perceptions or feelings, as this does not capture the ordering subjectivity of many moral adjectives (and perhaps also of some aesthetic adjectives). By contrast, we propose that adjectives that fall clearly in the subjective class are characterised by being holistically multidimensional.

5. Conclusion

In this paper, we have discussed Solt’s 2018 experimental paradigm to test the ordering subjectivity of adjectives. We noted that Solt did not test for broadly moral adjectives, which were an interesting class *vis-à-vis* her account of why certain adjectives are clearly subject to ordering subjectivity. Our results confirmed that moral adjectives fall in this class as well. This called for an enrichment of the notion of judge-dependency, beyond mere experiencer sensitivity. We have proposed to adopt a notion of judge-dependency as holistic multidimensionality, according to which a predicate is holistically multidimensional just in case there are dimensions that contribute to its meaning whose contribution is non-separable. This is a characterisation of judge-dependency that does not rely on experience and is therefore appropriate to capture the meaning of all ordering-subjective predicates.

Acknowledgements

Special thanks to Mora Maldonado for invaluable help designing the experiment and analysing the data. Thanks to audiences at Logic Now And Then 4 (Brussels) and the 2018 Filolab International Summer School in Granada. Thanks to Isidora Stojanovic, Carla Umbach, Paul Egré, Milica Denic, Amir Anvari and two referees from *Inquiry* for their comments. This work was supported by a European Grant of Obra Social La Caixa, by a postdoctoral fellowship of the Flanders Research Foundation (FWO) and by a Lise Meitner grant of the Austrian Science Fund (FWF), M 25-27 G32.

Disclosure statement

No potential conflict of interest was reported by the author.

References

- Bader, Ralf. (2017). Admissible factorisations and hyperintensional independence. Unpublished manuscript.
- Barker, Chris (2002). The dynamics of vagueness. *Linguistics and Philosophy* 25.1, pp. 1–36.
- Burnett, Heather (2016). *Gradability in Natural Language : Logical and Grammatical*

Foundations. Oxford University Press.

Bylinina, Lisa (2017). Judge-dependence in degree constructions. *Journal of Semantics* 34.2, pp. 291–331.

Cepollaro, Bianca and Isidora Stojanovic (2016). Hybrid Evaluatives: In Defense of a Presuppositional Account. *Grazer Philosophische Studien* 93.3, pp. 458–488.

Faroldi, F. L. G., & Soria Ruiz, A. (2017). The Scale Structure of Moral Adjectives. *Studia Semiotyczne*, 31(2), 161–178.

Franzén, N. (2018). Aesthetic Evaluation and First-Hand Experience. *Australasian Journal of Philosophy*, 96(4), 1–14.

García-Carpintero, Manuel and Max Kölbel (2008). *Relative Truth*. Oxford University Press.

Glanzberg, M (2007). Context, content, and relativism. *Philosophical Studies* 136.1, pp. 1–29.

Hare, R. M. (1952). *The Language of Morals*. Oxford University Press.

Hopkins, R. (2011). How to be a pessimist about aesthetic testimony. *The Journal of Philosophy*, 108(3), 138–157.

Jaeger, T. Florian (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59.4, pp. 434–446.

Kaiser, E., & Rudin, D. (2020). When faultless disagreement is not so faultless: What widely-held opinions can tell us about subjective adjectives. *Proceedings of the Linguistic Society of America*, 5(1), 698–707.

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* 30.1, pp. 1–45.

– (2013). Two Sources of Subjectivity: Qualitative Assessment and Dimensional Uncertainty. *Inquiry* 56.2-3, pp. 258–277.

Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics and Philosophy* 4.1, pp. 1–45.

Kölbel, M (2003). Faultless disagreement. *Proceedings of the Aristotelian Society* 104.1.

Kyle, Brent G. (2013). How Are Thick Terms Evaluative? *Philosophers' Imprint* 13.1.

Lasersohn, Peter (2005). Context dependence, disagreement, and predicates of personal taste. *Linguistics and Philosophy* 28.6, pp. 643–686.

- Liao, S., McNally, L., & Meskin, A. (2016). Aesthetic adjectives lack uniform behavior. *Inquiry*, 59(6), 618–631.
- Liao, S., & Meskin, A. (2017). Aesthetic Adjectives: Experimental Semantics and Context-Sensitivity. *Philosophy and Phenomenological Research*, 94(2), 371–398.
- MacFarlane, John (2014). *Assessment sensitivity : relative truth and its applications*. Oxford University Press.
- McNally, Louise and Isidora Stojanovic (2017). Aesthetic Adjectives. In: *The Semantics of Aesthetic Judgment*. Ed. by J. Young. Oxford University Press.
- Mothersill, M. (1984). *Beauty restored*. Oxford University Press.
- Ninan, D. (2014). Taste predicates and the acquaintance inference. In *Proceedings of SALT 24* (pp. 290–309).
- Nussbaum, M. (1990). *Love's Knowledge. Essays on Philosophy and Literature*. Oxford University Press.
- Oddie, G. (2001). Axiological atomism. *Australasian Journal of Philosophy*, 79(3), 313–332.
- (2001). Recombinant values. *Philosophical Studies*, 106(3), 259–292.
- Pearson, H. (2013). A judge-free semantics for predicates of personal taste. *Journal of Semantics*, 30(1), 103–154.
- Roberts, Debbie (2013). Thick concepts. *Philosophy Compass* 8.8, pp. 677–688.
- Robson, J. (2015). Norms of belief and norms of assertion in aesthetics. *Philosophers' Imprint*, 15(6), 1–19.
- (2012). Aesthetic testimony. *Philosophy Compass*, 7(1), 1–10.
- Rudolph, R. E. (2020). Talking about appearances: the roles of evaluation and experience in disagreement. *Philosophical Studies*, 177(1), 197–217.
- Sæbø, Kjell Johan (2009). Judgment ascriptions. *Linguistics and Philosophy* 32.4, pp. 327–352.
- Sassoon, Galit W. (2013). A typology of multidimensional adjectives. *Journal of Semantics* 30.3, pp. 335–380.
- (2016). Multidimensionality in the grammar of gradability. Unpublished manuscript.
- Silk, A. (2019). Evaluational adjectives. *Philosophy and Phenomenological Research*.
- Smith, E. A., Castroviejo, E., & Mayol, L. (2015). Cross-Linguistic Experimental Evidence Distinguishing the Role of Context in Disputes over Taste and Possibility. In

International and Interdisciplinary Conference on Modeling and Using Context.

Solt, S. (2018). Multidimensionality, subjectivity and scales: experimental evidence. In E. Castroviejo, L. McNally, & G. W. Sassoon (Eds.), *The Semantics of Gradability, Vagueness, and Scale Structure* (pp. 59–91). Springer.

Stephenson, Tamina (2007). Judge dependence, epistemic modals, and predicates of personal taste. *Linguistics and Philosophy* 30.4, pp. 487–525.

Stojanovic, Isidora (2007). Talking about taste: Disagreement, implicit arguments, and relative truth. *Linguistics and Philosophy* 30.6, pp. 691–706.

– (2019). Disagreements about Taste vs. Disagreements about Moral Issues. *American Philosophical Quarterly*, 56(1), 29–42.

Väyrynen, Pekka (2013). *The Lewd, the Rude, and the Nasty*. Oxford University Press.

– (2017). “Thick Ethical Concepts”. In: *The Stanford Encyclopaedia of Philosophy*. Ed. by Edward N Zalta. Fall 2017. Metaphysics Research Lab, Stanford University.

Willer, M., & Kennedy, C. (2019). Assertion, Expression, Experience. Retrieved from <https://semanticsarchive.net/Archive/jE5ZTAyY/>

Appendix

This appendix provides the list of critical and control items used in the experiment.

- Critical items
 - (4) a. What Rose did is morally better than what her boss did.
 - b. No, I disagree, what her boss did was better.
 - (5) a. The previous regime was more moral than the government that was established after the coup.
 - b. I disagree, the new government was more moral.
 - (6) a. Bill leads a more ethical lifestyle than Amy.
 - b. I disagree, Amy’s lifestyle is more ethical.
 - (7) a. The current CEO of that company behaved more virtuously than the previous one.
 - b. I disagree, the previous CEO was more virtuous.
 - (8) a. General McAdam was more coward than General Smith in that particular

battle.

b. No, General Smith was more coward.

(9) a. The members of the committee were more generous than the invited participants.

b. No, the invited participants were more generous!

(10) a. Bill's friends are more loyal than Mary's.

b. No, Mary's friends are more loyal.

(11) a. The treasurer turned out to be more honest than the president.

b. I don't agree with you, the president was more honest.

(12) a. It's more important to follow Mary's than Bill's advice.

b. I don't agree with you, Bill's advice is more important.

(13) a. What she did is more justified than what he did to her.

b. I don't agree, what he did is more justified.

(14) a. Jane acted more rationally than Joe in the aftermath of the tragedy.

b. I disagree, Joe was more rational than Jane.

(15) a. Everyone knows that Ann's work is more valuable than Jim's.

b. I don't agree with you, Jim's work is of far more value.

- Control items (from Solt 2018)

(1) a. Frank is taller than his friend Jimmy.

b. No, Jimmy is the taller one.

(2) a. The old Ipad is more expensive than the new one.

b. No, the new one is the more expensive device.

(3) a. The movie theater is emptier today than it was yesterday.

b. No it isn't. It was emptier yesterday.

(4) a. The wine bottle is fuller than the champagne bottle.

b. No, the champagne bottle is fuller.

- (5) a. The vegetable soup is saltier than the chicken soup.
b. No, the chicken soup is saltier.
- (6) a. The green towel is wetter than the red one.
b. No, the red towel is wetter.
- (7) a. Caryl and Tina both have blond hair, but Caryl's is lighter than Tina's.
b. No, Tina's hair is definitely lighter than Caryl's.
- (8) a. I would rather use the yellow pillow—it's harder than the white one.
b. No, the white pillow is the harder one of the two.
- (9) a. The cake is tastier than the chocolate cookies.
b. No, the chocolate cookies are tastier.
- (10) a. Those sneakers are uglier than the sneakers you tried on earlier.
b. No, the earlier sneakers were uglier.
- (11) a. This cat is happier than that dog.
b. No, the dog is the happier one of them.
- (12) a. The math professor is more intelligent than the physics professor.
b. No, I disagree. The physics professor is the more intelligent one.