Vagueness and the derivation of structural implicatures*

Timothy Leffel,[1] Alexandre Cremers,[2] Nicole Gotzner,[3] Jacopo Romoli[4]

[1]University of Chicago   [2]Ecole Normale Supérieure

[3]Zentrum für Allgemeine   [4]Ulster University
  Sprachwissenschaft (ZAS)

Abstract

While vagueness and implicature are two of the most studied topics in semantics and pragmatics, their interaction has received little attention in the literature. Part of what makes this an interesting domain of investigation is that both implicature and vagueness involve an inherent amount of interpretive uncertainty: with vagueness, there is uncertainty about whether a predicate applies to an object; with implicature, there is uncertainty about whether a candidate inference was intended by the speaker. Gradable adjectives provide an ideal domain for investigating how multiple sources of uncertainty interact in semantic/pragmatic interpretation. In this paper, we argue that interaction between vagueness and implicature is the source of a puzzling and, to our knowledge heretofore unobserved, asymmetry between the interpretations of different classes of intensified gradable adjectives under negation. Specifically, the interpretation of sentences containing the collocation *not very* ADJ varies depending upon whether ADJ is a relative standard gradable adjective (e.g. *tall*) or a (minimum standard) absolute gradable adjective (e.g. *late*), in the sense of Kennedy & McNally (2005); Kennedy (2007); a.o. We first report an experiment investigating the interpretation of *not very* ADJ and related expressions across these two adjective types. Our results provide strong evidence that *not very* ADJ gives rise to an ADJ-implicature for minimum standard adjectives like *late* but not for relative standard adjectives like *tall*. We propose that this asymmetry follows from an interaction between scale structure and a novel constraint on implicature calculation according to which implicatures are not derived if they lead to "borderline contradictions" (Ripley 2011; Alxatib & Pelletier 2011; a.o.). We then provide evidence for this hypothesis based on reconstructed interpretations of logically complex predicates from the experimental data. In sum, the interpretive asymmetry is reducible to differences in scale structure and vagueness: relative adjectives are vague in a way that absolute adjectives are not, which we argue affects the relative likelihood that an implicature will be generated.

1 Introduction

1.1 The puzzle

Gradable adjectives are adjectives that can hold of objects to varying degrees. For example *tall* is gradable because two people can both be tall without having to be the same height. Adjectives like *dead* or *prime* are not gradable, because they have "all-or-nothing" interpretations: an organism is either dead or it isn't; a natural number is either prime or it isn't. A defining property of gradable adjectives is that they can participate in degree comparison constructions (e.g. *tall-er/-est*, *as tall as*) and can be modified by intensifiers like *very*.

The phenomenon we investigate in this paper is the interpretive contrast between different intensified gradable adjectives in sentences like (1) and (2). The most prominent interpretation of (1) appears to imply that John was in fact late, just not very much so. But (2) does not invite a corresponding inference to the positive form. In fact, the most natural interpretation of (2) might even be *in*compatible with John being tall.[1]

(1)	John was not very late.		(2)	John is not very tall.
	⤳ John was late.			⤫ John is tall.

The contrast between *tall* and *late* in (1)-(2) exemplifies a broader pattern: in the frame '*not very* ADJ', certain adjectives pattern like (1), as shown in (3), while others pattern like (2), as shown in (4).[2] To anticipate the next section: the difference in interpretation between the examples in (3) and those in (4) aligns with the relative/absolute distinction in gradable adjectives, which has to do with the scale structure an adjective makes reference to and the amount of vagueness involved in its meaning (Unger 1975; Kennedy & McNally 2005; Kennedy 2007).

(3)	a.	The antenna is not very bent.	(4)	a.	John isn't very smart.
		⤳ The antenna is bent.			⤫ John is smart.
	b.	The table isn't very dirty.		b.	The supermarket is not very far.
		⤳ The table is dirty.			⤫ The supermarket is far.
	c.	Mary isn't very sick.		c.	The line is not very long.
		⤳ Mary is sick.			⤫ The line is long.

The fact that the interpretation of intensified adjectives under negation is sensitive to the relative/absolute distinction is, to our knowledge, a new observation. Bolinger (1972) and Horn (1989) pointed to a *not*-Adj interpretation of examples like (2), and proposed that it involves a form of euphemism or is related to politeness. Horn in particular refers to *not very* ADJ as a kind of "negative understatement," suggesting that its meaning is derived from a "conventionalized strengthening rule" that interprets '*not* intensifier ADJ' as '*rather un*-ADJ' " (Horn 1989:353-4). Horn cites pairs like *happy/sad* and *smart/stupid* to argue that the 'not ADJ' interpretation is euphemistic: in the frame '*not* intensifier ADJ', the "positive" (in the evaluative sense) predicates of these pairs have the 'not ADJ' reading, while the "negative" predicates seem to be more neutral.

However, while an explanation in terms of euphemism feels plausible for evaluative adjectives (*smart/stupid*) or adjectives that encode a desirable property (*tall*), it is unclear how it could extend

---

[1] The inference to the positive form from (2) is more attractive if *very* is stressed. We return to this point below.

[2] Contraction of negation (*isn't* versus *is not*) does not seem to make a difference, at least in our judgments.

to cases of purely dimensional relative adjectives like *far* in (4b). In fact, looking at the pair of absolute antonyms *early/late* and the pair of relative antonyms *close/far* suggests that there is a double dissociation between "positivity" and the availability of the 'not ADJ' interpretation. Indeed, both absolute adjectives *early* and *late* give rise to an 'ADJ' inference, while both relatives adjectives *close* and *far* are compatible with the 'not ADJ' interpretation. Because they are antonyms, an explanation based on euphemism would predict differences *within* pairs, but the observed contrast is *between* pairs instead, aligning with the relative/absolute distinction.

The observation that underlying lexical semantics systematically affects whether or not an adjective in the '*not very* ADJ'-frame is associated with the positive 'ADJ' inference shows that Horn's (1989) passing characterization of the construction is at the very least incomplete. More broadly, this variation between adjective subclasses is problematic for *any* general explanation that does not make reference to the lexical semantics of particular adjectives.

We propose instead that the contrast illustrated in (1)-(4) is a byproduct of interactions between principles regulating conversational implicature and vagueness. More specifically, we will advance a theory according to which the inference to the positive form from (1) is a structural manner implicature that is blocked in the case of (2) by a constraint on vague extension boundaries. In the conclusion, we will see how marginal effects of evaluativity may be present on top of this.

The rest of the paper is organized as follows: in the rest of §1 we introduce some generalizations about the interpretation of *not very* ADJ and related constructions, illustrating why they jointly constitute a genuine semantic puzzle. In §2 we present an experiment designed to provide detailed information about the semantic relationship between *not very* ADJ and related expressions. In §3 we formulate a theory of the interaction between vagueness and implicature calculation based on the idea that implicatures are blocked if they lead to "borderline contradictions" when conjoined with the literal meaning. We then provide evidence for the proposal based on reconstructed interpretations of logically complex predicates from the experimental data in §2. §4 concludes.

1.2 Vagueness and scale structure

Vagueness can be characterized in terms of systematic uncertainty about whether an expression applies to an object. For example does five hours from now count as 'soon'? Does a $6'$-tall person count as 'tall'? Put another way, vagueness involves the existence of so-called "borderline" cases: if we characterize an object *A* as heavy and an object *B* as not heavy, we can always find an object *C* with weight in between *A* and *B*, for which we are not sure whether we should call it heavy or not (see Barker 2006 for a broad overview on vagueness in language).

Vagueness arises with diverse categories of linguistic expressions and in particular with some gradable adjectives like *tall* and *clever*. But not all gradable adjectives are alike with respect to vagueness. For instance, it is doubtful that one could ever say *exactly* what it takes for someone to be tall: there will always be borderline cases and some amount of uncertainty about who counts as tall. But adjectives like *full*, at least in principle, have *precise* meanings.[3] If a class allows a maximum of 30 students and only 27 have registered, the class is not full and there is no uncertainty about the matter; if 30 have registered, then the class is full (and again, no uncertainty).

Gradable adjectives have been studied extensively and their vagueness (or absence thereof) can be explicitly modeled. Semantics for gradable adjectives are typically stated in terms of scales – ordered sets of reified "degrees" – and threshold values on those scales ("degree semantics," Bartsch

---

[3] Although some have suggested that all gradable adjectives are vague (e.g. Lassiter & Goodman 2014), they at least agree that there are differences in the amount of uncertainty introduced by different adjectives.

& Vennemann 1972; Cresswell 1977; von Stechow 1984; Bierwisch 1989; Heim 1985,2000; Kennedy 1999,2007). In this framework an adjective ADJ applies to an object x iff x's position on the ADJ-scale exceeds a contextually or compositionally designated value called the "threshold of application" (or "standard of comparison"), written 'θ_{ADJ}'. For example John counts as tall just in case John's position on the tallness scale (i.e. his height) exceeds whatever the context determines is the minimum height required to count as 'tall' – i.e. θ_{tall}. Thus the meaning of *John is tall* can be represented in terms of degree semantics as $\textbf{height}(\textbf{john}) > \theta_{tall}$. This affords a natural analysis of degree modifiers as items that shift threshold values (Kennedy & McNally 1999; Barker 2002; a.o.). If the threshold for *tall* is (say) around $6'$ in some context, then the threshold for *very tall* might be around $6'2''$. Note that the shift itself ($2''$ here) should itself be vague and context dependent.

The distinction between *tall* and *full* is commonly modeled as a difference in underlying scale structure: "relative standard" gradable adjectives (*tall*) are associated with open scales (scales without endpoints) and have contextually variable thresholds; "absolute" gradable adjectives (*full*) are associated with scales that have either a maximum or minimum boundary (or both), and by default have thresholds located at a scalar endpoint (see Kennedy 2007; Kennedy & McNally 2005; Solt 2015 and others, details differ by implementation). Two important subclasses of absolute gradable adjectives can be identified by the distribution of modifiers: maximum standard adjectives and minimum standard adjectives, which have default thresholds at scalar maxima and minima, respectively.[4] Maximizers like *completely* and *totally* are grammatical only with maximum standard adjectives like *full*, *empty*, and *straight* (cf. **completely tall/late*). Minimizers like *slightly* and *a little* are grammatical only with minimum standard adjectives like *bent*, *curved*, *early*, and *late* (cf. **slightly full/tall*).

In sum, gradable adjectives are commonly divided between *absolute* gradable adjectives, so-called because there is a sense in which something can be, for instance, full or bent or late without reference to other objects, and *relative* standard adjectives, which are inherently context-dependent: something can only be tall or big relative to other objects (a "comparison class" in standard parlance, see Klein 1980; Solt & Gotzner 2012). The two adjective subtypes differ with respect to whether vagueness is involved in their interpretation: because the threshold for absolute adjectives is a known value, it is fixed and well defined, while the threshold of a relative adjective is context-dependent and vague. We argue below that this difference is one of the two major sources of the contrast we identified in the interpretation of *not very* ADJ. The other source is a particular kind of interaction between restrictive modification and the Gricean Maxim of Manner, to which we now turn.

1.3 Structural "manner" implicatures from modifiers

Grice's Maxim of Manner states (roughly) that a speaker will not use more words than are necessary to get her point across (Grice 1975 and much subsequent work). In combination with the Maxim of Quality, this leads to implicatures when a speaker uses an utterance which is both more complex than a readily available alternative and less informative than the simpler alternative. Inferences generated in this fashion have received considerable attention in the literature and have been treated as "manner implicatures" or (quasi)-presuppositions (Simons 2001/2013; Schlenker 2008), or simply as

---

[4] Note that absolute adjectives are also context-dependent, but arguably in a different kind of way: clearly what it means to be *late* or *full* varies by situation. Five minutes late might count as 'late' for work, but not for a casual dinner, for example. Most theories of gradability model such non-endpoint oriented readings as a form of imprecision (e.g. as in approximative readings of numerals), which introduces some flexibility into the interpretation. See Kennedy 2007 or Lasersohn 1999 for a pragmatic analysis of imprecise interpretations of absolute adjectives.

Quantity implicatures given a theory of alternatives which includes all sentences that are structurally simpler than the asserted one (Katzir 2007,2014; Fox & Katzir 2011; see also Matsumoto 1995; Sauerland 2004; Chemla 2009). A hallmark case of these implicatures involves restrictive modifiers in negative and downward-entailing environments. For example Simons (2001/2013) notes that sentences like (5a) strongly imply that the unmodified alternative sentence (5b) is false, i.e. that John did indeed vote (for someone other than Nader).

(5) a. John didn't vote for Nader.
     b. John didn't vote.
     c. John voted.

The inference of (5c) from (5a) can be shown to follow from standard (neo)-Gricean reasoning. The following informal sketch is adapted from Simons 2001/2013 and Katzir 2007.

  (*i*) You, the speaker, assert (5a).
 (*ii*) I, the addressee, observe that (5b) is a "better" alternative to (5a), since (5b) asymmetrically entails (5a) (Quantity) and contains a proper subset of words (Manner).[5]
(*iii*) I observe that you chose not to assert (5b) despite this fact.
 (*iv*) By Quality, I therefore conclude that you must not believe (5b), because if you did, you would have uttered it instead.
  (*v*) I assume that if you know John didn't vote for Nader, then you probably know whether or not he voted (you are an "opinionated authority" in Sauerland (2004) and others' terms).
 (*vi*) Given (*iv*) and (*v*), it follows that you probably believe (5c).

Implicatures like the one in (5) arise in syntactic environments other than matrix negation, as well. This is because, as Katzir (2007) and others discuss, a modified sentence generally asymmetrically entails its unmodified version in upward-entailing contexts (e.g. *John passed the exam with an A* entails *John passed the exam*). In downward-entailing contexts entailment relations are reversed, so in such environments we expect an implicature to the negation of the unmodified version from the assertion of the modified one. For instance (6a) naturally gives rise to the negation of (6b) – *I think John passed the exam* – as an implicature. Similarly (7a) naturally implicates the negation of (7b), namely that some of the students passed.

(6) a. I don't think John passed the exam with an A.
     b. I don't think John passed the exam.

(7) a. None of the students passed the exam with an A.
     b. None of the students passed the exam.

1.4 Back to the puzzle

We propose that the puzzle about *not very* in (1)-(4) is partially explained by manner-based pragmatic reasoning parallel to cases like (5). The account runs as follows: a modified sentence like (8a) entails its unmodified counterpart (8b). In other words, *very* is a restrictive modifier on gradable adjectives, so anything that's 'very ADJ' necessarily exceeds the threshold for ADJ.

---

[5] However one wants to measure utterance length or structural complexity, it should be clear that (5b) is "briefer" and "simpler" than (5a); see Katzir 2007; Fox & Katzir 2011 for some relevant formalizations of these notions.

(8) a.   John was very late.
    b.   John was late.

In a downward entailing environment like matrix negation, the entailment relation is reversed, so (9a) is entailed by (9b). And crucially (9b) is structurally less complex than (9a) (uncontroversially so, we believe), since the former is a substructure of the latter. Therefore, given the reasoning in (*i-vi*) above, (9a) should be associated with an implicature to the negation of (9b) – *John was late*.

(9) a.   John was not very late.
    b.   John was not late.

And as we pointed out above, (9a) does seem to suggest that John was late. Similarly, cases like those in (10) conform to this expected pattern.

(10) a.   The kitchen is not/isn't very dirty.   ⤳ The kitchen is dirty.
     b.   The chair didn't get/isn't very warped.   ⤳ The chair is warped.
     c.   The antenna isn't very bent.   ⤳ The antenna is bent.

A key premise of the analysis is that *very* triggers manner-based reasoning due to its function as a restrictive modifier. It is therefore expected that other restrictive modifiers in this syntactic frame should give rise to corresponding inferences. And indeed the pattern does appear to extend to modifiers beyond *very*, including *super*, *really*, *extremely*, etc. (see Bolinger 1972 and Horn 1989 for discussion). For instance our intuition is that *not super late* tends to imply *late*. Additionally, (11) shows that the same inference pattern applies in downward-entailing contexts beyond just matrix negation, further strengthening the parallel between intensified adjectives and implicatures from phrasal modifiers.[6] This generality across different adverbs and types of downward-entailing environment casts further doubt on the idea that *not very* ADJ is a frozen expression with idiosyncratic meaning.

(11) a.   I don't think John was very late.   ⤳ I think he was late.
     b.   None of the students were very late. ⤳ Some of the students were late.

The simple manner-based account as stated above does not, however, solve the entire puzzle. Because nothing in the derivation proposed for (8)-(9) above depends on the scale structure of *late*, a natural expectation is that *all* sentences of the form *X is not very* ADJ should be associated with implicatures of the form *X is* ADJ. But while we have seen that this holds for minimum-standard adjectives, it does not extend to relative gradable adjectives: without focus intonation on *very*, the sentences in (12) all seem to lack the '*X be* ADJ' inferences observed in (10). Similarly, the implicature is absent in other downward-entailing environments, as illustrated in (13), and also absent with other intensifiers beyond *very* (e.g. *not super tall*).[7]

(12) a.   John is not very tall.   ⤳̸ John is tall.

---

[6] Intuitions are also less robust in the restrictor of *every*, e.g. it is not obvious whether there is a difference in inferences to the positive form between sentences like *Everyone who was very late was sent to the principal's office* and *Everyone who is very tall will be considered for the basketball team*. We leave this as an issue for future research.

[7] *Extremely* may be an exception, but since the implicature resurfaces in all examples in (12) when *very* receives focus, a natural explanation might be that *extremely* attracts focus. We will propose an alternative explanation in fn14.

b.	The bed is not very comfortable.	↛ The bed is comfortable.
c.	Bill is not very happy.	↛ Bill is happy.
d.	John's lifestyle is not very responsible.	↛ John's lifestyle is responsible.

(13) a. I don't think John is very tall.   ↛ I think he is tall.
     b. None of the students are very tall.   ↛ Some of the students are tall.

The fact that the generalization targets a specific subclass of gradable adjectives strongly suggests that the underlying cause of the asymmetry is somehow related to scale structure. Even stronger support comes from parallel contrasts in other syntactic environments or with other syntactic categories: for attributive adjectives, compare *John is not a very tall guy* with *This is not a very bent stick*; for the adverbial domain compare *I can't jump very high* with *I can't arrive very early*. In both cases minimum standard predicates (*bent*, *early*) seem to license inferences to the positive form while relative predicates (*tall*, *high*) fail to.

The status of maximum standard absolute adjectives in the '*not very*'-frame is less clear, in part because their thresholds are located (by default) at scalar maxima: how could *very* shift a threshold to a higher value if it is already at the scale maximum? To the extent that *very* can occur felicitously with maximum standard adjectives (as in *very full*), there seems to be a kind of pragmatic weakening involved whereby *full*'s threshold is relaxed from the strict endpoint (Lasersohn 1999), in effect coercing it into a relative adjective (Kennedy & McNally 2005).[8] Because of these complications, we do not systematically investigate maximum standard adjectives in this paper. The comparison between relative and minimum standard absolute adjectives is sufficient for our present purposes since they differ semantically from one another with respect to vagueness.

So far, we have presented judgments which suggest that the interpretation of *not very* ADJ varies depending on whether ADJ is an absolute or a relative gradable adjective. Our own intuitions support the following generalization:

(14)   *X is not very* ADJ is most naturally interpreted as

    a. *X is* ADJ*, but not: very* ADJ   when ADJ is absolute (min. std.)
    b. *X is not: very* ADJ (and potentially *not* ADJ)   when ADJ is relative std.

We have shown that (14a) can be naturally explained as a garden-variety manner implicature. But the interpretation of relative adjectives in the '*not very* ADJ'-frame is a murkier phenomenon. The problem can be broken down into two independent questions. On the one hand there is the question of why relative adjectives do not trigger implicatures to the positive form:

(*i*) **lack of expected inference to the positive form**
Even though *tall* is restrictively modified by *very* in a downward-entailing environment, *John is not very tall* does not have the negation of the unmodified alternative as an implicature (i.e. *John is tall* is not implicated).

However, as indicated by the phrasing of (14b), it is less than clear whether *not very* ADJ with a relative adjective entails *not* ADJ (as suggested by Horn 1989), or merely fails to imply ADJ:

---

[8] Kennedy & McNally (2005:370-1) have in fact suggested that *very* is actually ungrammatical with absolute adjectives altogether (except possibly when used in a "relative like, imprecise" way). However, the use of *very* with absolute adjectives is widespread in everyday speech and in corpora, and examples like *not very late* seem to be perfectly acceptable and interpretable (as confirmed in §2).

(*ii*) **unexpected strengthening(?)**

Without the implicature to the positive form, one should expect *John is not very tall* (with the parse [*not* [*very tall*]]) to mean "John is less tall than is required to count as very tall," which is compatible with John being either tall or not tall. This contradicts the intuition that *John is not very tall* might even convey that John is not tall.

1.5 Motivating an experimental approach

The generalizations we have advanced so far have been based solely upon our intuitive judgments. But there are reasons to believe that intuition alone is inadequate for a complete understanding of the present phenomenon. For one, intuitions about the exact meaning of *not very* ADJ are not always crystal clear, especially as they relate to expressions like *not* ADJ or the antonym of ADJ. A particularly salient example: does *not very tall* imply *short*, *not tall*, or does it merely fail to imply *tall*? The answers to such questions constrain candidate theories of the puzzle, but cannot be obtained through introspection alone. The phenomenon also seems to be *gradient* in a number of ways, which makes theorizing on the basis of binary judgments more difficult. For example inference to the positive form is very attractive with *not very late*, not at all attractive with *not very tall*, and perhaps somewhere in between with (say) *not very fat*. Gradience in the interpretation of gradable adjectives is of course not surprising, but it can sometimes call for methodologies that yield finer-grained data than categorical introspective judgments can; this is especially true in the present case, where the relevant judgments concern availability of various inferences. The recent body of experimental research on the semantics, pragmatics, and processing of gradable predicates reflects a growing consensus that quantitative analysis can produce valuable insights about gradability and scalar meaning (Syrett et al. 2010; Foppolo & Panzeri 2011; Solt & Gotzner 2012; Lassiter & Goodman 2014; Qing & Franke 2014; a.o.).

Finally, quantitative data provide a more direct way of evaluating the manner-based theory stated above. In particular, the idea that *not very* ADJ has ADJ as a structural implicature makes a few concrete predictions about the meaning of *not very* ADJ as it relates to *not* ADJ and ADJ. First, the implicature-based account predicts that *not* ADJ is a (structural) *alternative* to *not very* ADJ; if correct, then we expect to find evidence of *competition*: two competing forms should never be fully acceptable at the same time. Second, if the positive form is in fact derived from *not very* ADJ as an implicature, then we expect to see more uncertainty (or greater variance) in judgments of truth in situations in which the literal reading is true but the strengthened reading is false, since implicatures are defeasible and not always computed. The degree to which these predictions are borne out for absolute versus relative gradable adjectives constrains possible theories of the unexplained facts about relative adjectives discussed above: if there is strong evidence of competition between *not very* ADJ and *not* ADJ, for example, then this would rule out theories that do not invoke alternatives-based reasoning in the interpretation of the relevant expressions. And assuming that entailments are generally associated with less uncertainty than are implicatures, we can gain clues about the status of the putative 'not tall' inference from *not very tall* viz. entailment versus implicature versus neither (which, in turn, further constrains possible theories of the phenomenon).

In the next section we report an experiment designed to yield fine-grained information about how speakers interpret sets of related gradable expressions. The experimental data come in the form of gradient truth-value judgments from a fairly large sample of speakers, collected in a large range of precisely manipulated situations/contexts.

2 Experiment

We estimated speakers' interpretations of gradable expressions by creating continua of ordered degrees on scales and then asking people how strongly they agreed or disagreed with statements about objects located at various positions on the scales (e.g. is a $6'$-tall man "tall"?). We investigated *tall*, which for current purposes we take to be representative of the class of relative standard gradable adjectives, and *late* which we take to be representative of the class of (minimum standard) absolute gradable adjectives. Because the two adjectives required different background stories, adjective type was a between-subjects factor.

2.1 Methods

2.1.1 Participants

Experimental participants were recruited via Amazon Mechanical Turk in two survey versions. 36 participants took part in the *tall*-version, and 35 in the *late*-version (age range: 19–60). One participant was excluded from the analyses because their self-reported native language was not English.

2.1.2 Materials and procedure

In each trial, a participant was presented with a fact concerning the height or arrival time of a different person (in the *late*-version, initial instructions provided global context by specifying that newly hired people were expected to arrive at 9am for their first day at work). The fact was paired with a statement, uttered by a character named Mary, involving one of many different adjective constructions. The participant's task was to indicate whether they agreed or disagreed with Mary by adjusting a slider with values ranging from 0 to 100; see Figure 1 for sample displays.

Fact: Donna showed up to work at 8:48am.

Mary said: "Donna was not very late."

disagree _____ agree

Continue

Fact: Ronald is 5 ft 10 in.

Mary said: "Ronald is not very tall."

disagree _____ agree

Continue
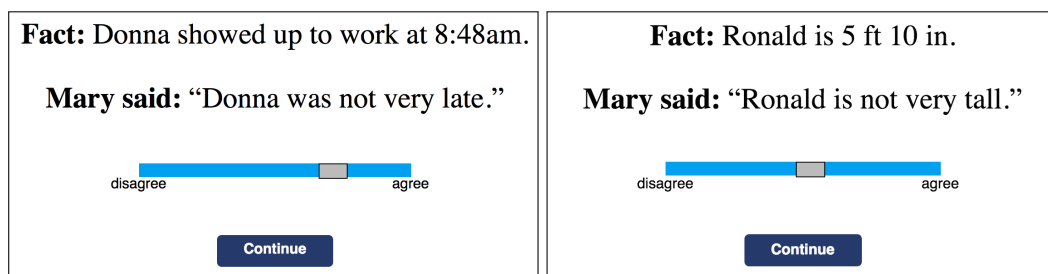
Figure 1:   Sample displays for *late* (left panel) and *tall* (right panel)

In the *tall*-version, participants evaluated Mary's statements about men of different heights.[9] We tested the following seven constructions: "*X* is tall," "*X* is not tall," "*X* is very tall," "*X* is not very tall," "*X* is short," "*X* is not short," and "*X* is neither tall nor short," paired with 13 different heights ranging from 5ft 3in (160cm) to 6ft 10in (208cm).[10] In the *late*-version, Mary made statements about newly hired employees, who were all expected at 9am for their first day of work. Participants evaluated the following seven constructions: "*X* was late," "*X* was not late," "*X* was very late," "*X* was not very late," "*X* was early," "*X* was not early," and "*X* was on time," at 13

---

[9] We used only unambiguous male names to keep the comparison class for tallness as homogeneous as possible.

[10] The heights we used were most dense close to 5ft 10 in (the average American adult male height), and sparser at lower and higher extremes. The heights (in inches) were: 63, 66, 68, 69, 70, 71, 72, 73, 74, 75, 77, 79, and 82.

Figure 2: Mean % agree by construction across degrees. Left panel: *late*. Right panel: *tall*. Ribbon width represents standard error.

arrival times ranging from 8:39am to 9:48am.[11] Each participant saw all possible combinations of scale points and constructions twice (with a different name in each repetition), totaling in 182 trials per experiment version. Trials were presented in a different random order for each participant.
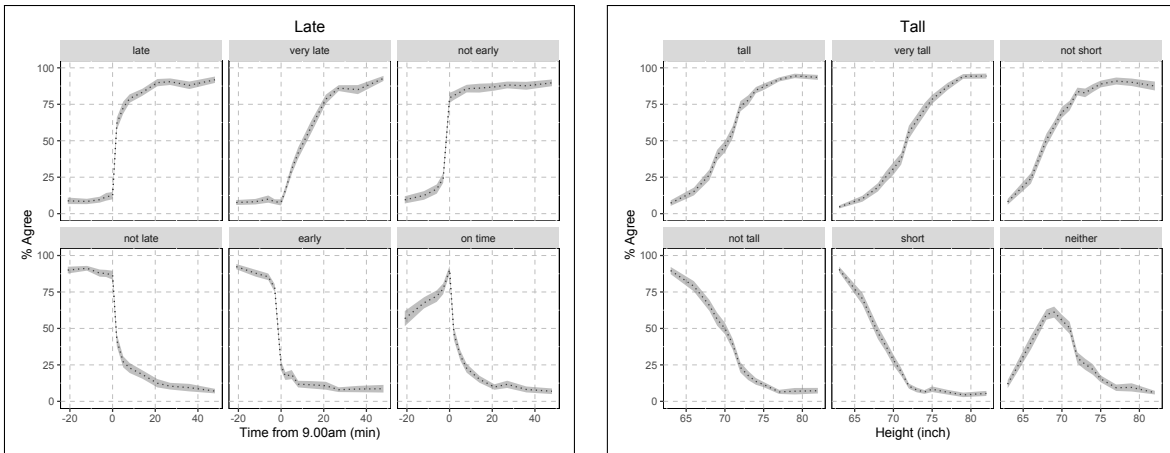
2.2 Results

Figure 2 shows judgments averaged over participants and items for all constructions except *not very* ADJ, for the *late*-experiment (left panel) and the *tall*-experiment (right panel). The curves associated with *tall* and *late* can be seen as visualizations of scale structure: acceptance of the predicate *tall* follows an approximately sigmoidal shape, increasing as height increases for nearly the entire interval of heights. This reflects uncertainty about the exact value of the tallness threshold. For *late*, acceptance remains relatively constant and close to zero prior to 9am, and then shows a rapid increase as time moves beyond 9am, eventually leveling off near 100% acceptance. This reflects the virtual certainty that the threshold for lateness is located at 9am.[12]

Some entailment relations (or relative strengths) of expressions can be visualized by relative height on the coordinate system: if two curves in Fig. 2 have similar shapes, then the one that has consistently lower agree-% values entails the one with consistently higher values. For example *very tall* has a quite similar shape to *tall*, but *very tall* has consistently lower values than *tall* across all heights. This reflects the fact that the former entails the latter. Similar remarks apply to *late* versus *very late*. In general, the results depicted in Fig. 2 are consistent with pre-theoretical expectations about the respective meanings of the constructions involved.

The comparison of interest for determining the meaning of *not very* ADJ is between the mean agree-% of *not very* ADJ and that of *not* ADJ. Figure 3 plots participants' mean judgments for the

---

[11] The arrival times we used were distributed around 9am in the same fashion that heights were distributed around 70in. They were: 8:39, 8:48, 8:54, 8:57, 9:00, 9:02, 9:05, 9:08, 9:14, 9:21, 9:27, 9:36, and 9:48.

[12] In fn4 we mentioned that although scalar endpoints are the default threshold for absolute adjectives, in reality the threshold may vary marginally across contexts. And indeed we observe that participants consider that 2min late does not count as perfectly 'late'. However, note that this uncertainty only extends to times *after* 9am, but crucially not before (and the pattern is reversed for *early*). This is an indication that the underlying mapping from arrival times to degrees treats all times earlier than 9am as 0, so that the scale for *late* is indeed lower-bound, and 9am gets mapped to a scalar endpoint even though it is not an endpoint on the physical timescale (similarly for *early*, all times past 9am get mapped to 0).
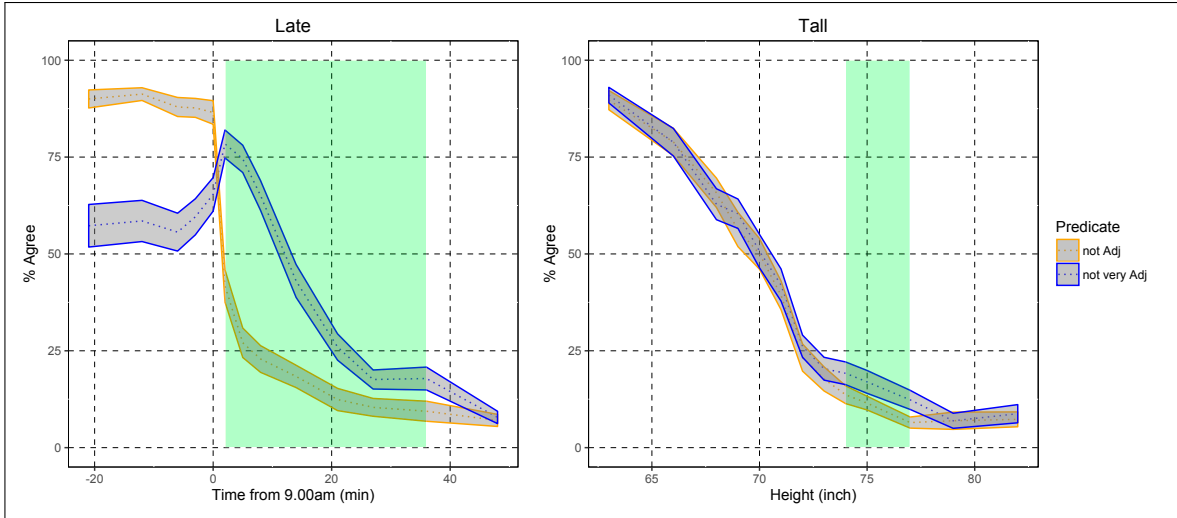
Figure 3: Mean % agree by construction across degrees. Left panel: *not late* (orange) and *not very late* (blue). Right panel: *not tall* (orange) and *not very tall* (blue). Significant clusters are indicated in green when *not very* ADJ is above *not* ADJ and in red otherwise. Ribbon width represents standard error.

critical constructions *not very* ADJ and *not* ADJ, across varying degrees of height or arrival time (again with *late* on the left and *tall* on the right). As can be seen from the plots, *not late* shows a sharp decline in acceptance in the region just beyond the threshold (9:00am), the very same region in which *not very late* shows the highest acceptance rate. On the other hand, *not very tall* displays virtually the same response profile as *not tall*. While acceptance of the predicate *tall* increases with height (Fig. 2, right), a roughly mirror-image pattern is present for *not very tall* (Fig. 3, right). This approximates the relationship between an expression and its negation (cf. *tall* versus *not tall* or *late* versus *not late* in Fig. 2).

To analyze the responses for *not very* ADJ and *not* ADJ statistically, we adapted the cluster analysis proposed by Maris & Oostenveld (2007) for EEG signals based on Monte Carlo sampling, treating the degree scales as our continuous dimension (instead of time) so that all responses from one participant to one sentence across the scale give us two curves (since each item was repeated twice). In short, the idea is to reassign each curve randomly to one of the sentence *not very* ADJ and *not* ADJ and see how likely it would be to observe differences at least as big as in the actual data.

After running 5000 such random permutations, we found two significant clusters for "late:" from 21min early to exactly 9:00am, *not very late* is significantly degraded compared to *not late* (59.3% (33) versus 88.7% (16) mean rating (*sd*), respectively); and from 2min late to 36min late, we observed the opposite (46% (34) versus 20% (24); see Fig. 3, left panel). Both clusters were larger than anything obtained from the 5000 random permutations, hence $p < .0002$.

For "tall," acceptability of the constructions *not very tall* and *not tall* coincided in most intervals. But the analysis revealed a small significant cluster from 6ft 2in (188cm) to 6ft 5in (196cm), during which *not very tall* was more acceptable than *not tall* (16.2% (20) versus 10.5% (13); see Fig. 3, right panel), with $p = .005$.

2.3 Discussion of results

Our results show that speakers clearly distinguish between the meanings of *not late* and *not very late*, and provide strong support the hypothesis that the latter is interpreted with an implicature to

the negation of the former. In particular, the signature of this implicature is visible in the results in three ways.

First, in the range of (strictly) late arrivals, participants treat *not very late* as more acceptable than *not late*, with the largest difference in acceptability – as well as the highest rate of agree-% for *not very late* – occurring in the region just after 9am. This shows that, in this range, *not very late* is strictly less informative than *not late*. This is a pre-condition for the derivation of an implicature from *not very late* to *late* (see step (*ii*) of the derivation proposed in §1.3)

Second, for early arrivals, *not very late* is instead degraded relative to *not late* – crucially in spite of the fact that early arrival is perfectly compatible with the literal meaning of *not very late*. This is strong evidence for the inference in (1): the fact that *not very late* is degraded precisely when *not late* is most acceptable suggests that the two expressions are in competition, as expected if *late* is an implicature of *not very late*.

Third, note that mean agreement rating for *not very late* is only at 57.8% prior to 9am – the region in which the literal reading is true but the strengthened reading is false. Variance is also higher for *not very late* than for all other constructions in this region (*sd* = 33, compared to the average *sd* of 18 for all others; difference visible from relative width of error ribbons). This pattern demonstrates uncertainty about whether being early counts as 'not very late', which is expected since the literal meaning is compatible with being early but the strengthened meaning is not. Again, this provides evidence for the *late* implicature of *not very late*.

A different picture emerges from the results of the *tall* conditions. The constructions *not tall* and *not very tall* are treated almost completely alike by participants, with no indication of the implicature *tall* from *not very tall*. The close similarity of the curves for *not tall* and *not very tall* confirms our intuition that the implicature to the positive form is not drawn from *not very tall* in the way it is drawn from *not very late*. Unlike the results for *late*, we found no evidence for competition between these two expressions – there is no interval during which acceptance of one spikes while acceptance of the other drops – and thus no evidence that an implicature would be derived from an assertion of one expression via reasoning about the non-asserted one. Moreover, the fact that the meaning of *not very tall* is almost identical to the meaning of *not tall* provides evidence not only for the lack of an implicature, but also for a meaning stronger than the mere negation of *very tall*.

Going back to the two puzzles outlined in §1, our results provide strong evidence that (i) unlike *not very late*, *not very tall* does not give rise to the positive inference 'tall'; and (ii) *not very tall* seems to be further strengthened to have a meaning very close to *not tall*.

Two final points relating to this aspect of the results: first, Horn's (1989) proposed interpretation for *not very tall* ('rather short') was not precise enough for a predictive theory. Our data, however, precisely characterizes the relationship between the interpretations of *not very tall*, *not tall*, and *short*: as can be seen from visual inspection of Figs. 3 and 2, we found that *short* is consistently interpreted as stronger than *not very tall*, with a mean agreement rate of 26% for the former versus 38% for the latter (collapsed over heights).

Second, we found that the response profiles for *not tall* and *not very tall* coincide for nearly the entire range of heights evaluated in the present study. Nonetheless, there is one exception: in the interval between 6ft 2in and 6ft 5in, *not very tall* is slightly more acceptable than *not tall* (16.2% (20) versus 10.5% (13) mean rating (*sd*), respectively). In some sense this is unsurprising, as these heights intuitively correspond to someone who would be tall, but less than very tall. Albeit small in magnitude and range, this difference could be an indication that *not very tall* is still marginally weaker than *not tall*, and hence that the "strengthening" of *not very tall* to *not tall* is more of a

gradient phenomenon than has been assumed in the past.

3 Vagueness and implicature as the source of the contrast

In our results, we found that a sentence like (15a) is interpreted with an inference to the negation of (15b). By contrast, a sentence like (16a) is interpreted as approximately synonymous with (16b) – with the caveat discussed in the previous section.

(15) a. John was not very late.   (16) a. John is not very tall.
      b. John was not late.   b. John is not tall.

In §1.4 we argued that the first case can be explained as a standard structural (manner) implicature triggered by competition between (15a) and (15b), and our experimental results confirmed this view. We have also seen that euphemism may partially explain how it is that (16a) can be used to communicate something that approximates (16b), instead of simply "less than very tall." But the explanation is inherently incomplete: euphemism cannot not explain why (minimum standard) absolute adjectives do not follow the same pattern, nor does it explain why reasoning about euphemism dictates meaning *instead of* the kind of manner-based Gricean reasoning that explains the interpretation of (15a). The remaining puzzle, then, essentially boils down to the question of why (16a) does not give rise to a *tall* implicature parallel to the *late* implicature of (15a).

Since the difference between adjectives that implicate the positive form in *not very* ADJ and those that don't systematically aligns with the distinction between minimum standard and relative standard gradable adjectives, we believe that the explanation must be sought in the interaction between vagueness – the semantic dimension along which the two classes vary – and the derivation of structural implicatures – the pragmatic mechanism that accounts for the positive form inference from *not very* ADJ (when present).

3.1 Deriving the relative/absolute asymmetry

We propose that the asymmetry in implicature to the positive form between *not very tall* (or *nvt*) and *not very late* (or *nvl*) follows directly from basic facts about relative versus absolute scale structure, combined with a novel constraint on the derivation of implicatures. Our constraint builds on the idea that implicatures should not contradict assertions, extending it to Ripley's (2011) notion of a *borderline contradiction*. Ripley defines a borderline contradiction as a conjunction of the form $(Fa \land \neg Fa)$, where F is a vague predicate and a is a borderline case for F (meaning it is not clear whether a is F or non-F). Such constructions have been shown to be more acceptable than ordinary contradictions, but still clearly degraded (Alxatib & Pelletier 2011; Serchuk et al. 2011; Egré et al. 2013; Zehr 2014). We propose to generalize this notion to the case of two (possibly) distinct but related predicates, using the definition in (17).

(17)   If F and G are gradable predicates on the same scale and with the same polarity: $(Fa \land \neg Ga)$ is a borderline contradiction iff a is a borderline case for both F and G.

Using the notion of a borderline contradiction as defined in (17), we propose the following novel constraint on implicature calculation:

(18)   **Constraint on vague implicatures**
      If a sentence S has an alternative S', the potential implicature $\neg S'$ is not drawn
      if $S \land \neg S'$ would necessarily be a *borderline contradiction*.

The intuition here is that borderline contradictions are generally infelicitous – or at least unlikely to be true – and therefore if there is an option to not interpret a sentence with a borderline-contradiction implicature, then such a strategy should be preferred. Similar constraints have been proposed in more restricted form: for example Fox & Hackl (2006) argue that implicatures are not derived if they lead to contradiction when combined with the literal meaning. (18) simply extends this idea to the generalized notion of borderline contradiction in (17): avoid deriving an inference if there is uncertainty as to whether the resulting meaning would *ever* be satisfiable.[13]

Our explanation of §2's results in terms of (18) runs as follows. Consider first the case of *not very late*. The meaning conveyed by the assertion with its implicature can be paraphrased as "late but not very late." Treating *late* as *F* and *very late* as *G* in the definition (17), we predict that the implicature is derived only if there are arrival times which do not lead to borderline contradictions. Because *late* is minimum-standard, any time shortly after 9am counts as clearly late, and by getting sufficiently close to 9am, there will be some times which also count as clearly less than very late. Since there are times which clearly count as late but not as very late, there is no borderline contradiction and hence (18) does not prevent an inference from being drawn; manner-based reasoning can then apply and the implicature may be drawn.

By contrast, the constraint (18) could block the derivation of the implicature for *nvt* in the following way. For the implicature to be derived, we would need to find individuals which are simultaneously clearly tall and clearly less than very tall. However, in this case both predicates are vague. We argue that in general there are simply no heights which clearly satisfy both predicates at the same time, because *very* does not increase the threshold enough to make *nvt* clearly compatible with *tall*.[14] Thus a borderline contradiction arises and the implicature is blocked by (18).

3.2 Evidence from further data analysis

Our theory of *not very* ADJ relies upon the crucial assumption that the conjunction of "tall" with "not very tall" is a borderline contradiction in a way that the conjunction of "late" with "not very late" is not. However, we have not justified this claim beyond intuition thus far. We will now do exactly this. Establishing the claim essentially reduces to showing that the phrase *tall but not very tall* is significantly less acceptable ("more contradictory") than *late but not very late*, which has a determinate lower-bound and hence should not give rise to borderline contradictions.

The most direct way to evaluate this assumption would be to elicit judgments about the acceptability of ADJ *and not very* ADJ. This strategy, however, is clearly problematic: the phrase *late and not very late*, for instance, seems pragmatically deviant, presumably because of lexical competition with *but*.[15] One could then consider using *but* to conjoin ADJ with *not very* ADJ, which has the same truth-functional meaning as *and* and is a more natural connective in this context. However, doing so would be equally problematic for the following reason: using *but* to conjoin ADJ *and not very* ADJ introduces an implication that the two conjuncts contrast with one another, and thus (since the main predicates are identical) attracts focus to the modifier *very*. As we noted earlier, focal

---

[13] A similar proposal is advanced by Chemla & Romoli (2015), who argue that implicatures are not derived if the probability of the implicature given the assertion is substantially lower than the probability of the implicature alone.

[14] This suggests that if we replace *very* with *extremely*, the implicature should resurface, given that it is easier to find heights which are clearly tall without being anything close to *extremely* tall. This feels intuitively correct: *John is not extremely tall* seems to suggest that John is tall to a greater degree than does *John is not very tall*.

[15] Corresponding effects are visible in other domains, e.g. conjoining *some* with *not all* seems to require *but* as opposed to *and*.

stress on *very* has the effect of forcing the derivation of the positive form implicature, thus avoiding contradiction (perhaps because stress has the effect of pushing the threshold further to the right, by increasing the strength of the intensification). Therefore, using *and* would make the conjunction of ADJ with *not very* ADJ infelicitous for independent pragmatic reasons, while using *but* would affect focus structure and hence provide judgments about a meaning that is probably distinct from the logical conjunction of the relevant predicates.

Instead of collecting explicit judgments, we employed an indirect strategy that allowed us to estimate the interpretation of the abstract construction "ADJ $\wedge \neg$ *very* ADJ," while avoiding the focus-related confound, as well as the confound introduced by pragmatic competition between *and* and *but*. In the following, we describe a post-hoc analysis on the data from §2 establishing that *tall* $\wedge \neg$ *very tall* does indeed have a degraded interpretation when compared to *late* $\wedge \neg$ *very late*, as hypothesized in §3.1 above.

By combining data about ADJ and data about *very* ADJ, we were able to "construct" estimated agree-%'s for complex expressions of the form ADJ $\wedge \neg$ *very* ADJ on the same 0-100 scale, and test whether they behaved like borderline contradictions (we take acceptability judgments in this task to reflect the perceived probability that the sentence is true). We used the definitions of negation and conjunction from fuzzy logic (Zadeh 1965), presented in (19), which is the standard for vagueness: where A is a proposition containing a vague item and $v(A)$ is its value on our $[0,100]$ scale, we define the negation of A to just be the proposition $\neg A$ such that $v(\neg A) = 100 - v(A)$. In other words, if a participant's mean rating for *Bill is tall* when Bill is 6ft is 70, then we can infer their mean rating of $\neg(Bill\ is\ tall)$ to be around $100 - 70 = 30$. And indeed, this "artificial" meaning almost completely coincides with participants' actual judgments about the sentence *Bill is not tall*. The conjunction $A \wedge B$ is defined as the proposition with value equal to the minimum of $v(A)$ and $v(B)$.[16]

(19)   Definitions of logical operators for reconstructing gradable pseudo-predicates:

    a. $v(\neg A)$ $=_{\mathrm{df}}$ $100 - v(A)$ (negation of proposition A)
    b. $v(A \wedge B)$ $=_{\mathrm{df}} \min(v(A), v(B))$ (conjunction of propositions A, B)

Using these definitions, we can estimate interpretations of propositions of the form ADJ $\wedge \neg$ *very* ADJ for each participant and scale point (e.g. someone's judgment about the degree to which a 6ft-tall man counts as 'tall') as follows: we first take the actual judgments for ADJ and for *very* ADJ. We then take, for each participant and each height/time, whichever of the following two quantities is smaller: the mean value of their judgments about ADJ, and 100 minus the mean value of their judgments about *very* ADJ.[17] This allows us to compare estimated meanings of logically complex predicates in order to evaluate our assumption about the role of borderline contradictions in the present phenomenon. We also calculated estimates for ADJ $\wedge$ *not* ADJ (using the judgments for *not* ADJ) as a reference point, since they are known cases of borderline contradictions.

Estimated mean agree-%'s for each of the four reconstructed predicates are plotted in Fig. 4. Visual inspection suggests that *tall* $\wedge \neg$*very tall* is clearly degraded compared to *late* $\wedge \neg$*very*

---

[16] Note that in general, the probability of a conjunction $p(A \wedge B)$ is not equal to $\min(p(A), p(B))$. However, if the events we consider are $A =$ 'attributing property F to x' and $B =$ 'attributing property G to x' with F and G two gradable expressions on the same scale, it is likely that one of A or B will entail the other. The choice of the min function translates this property and it is validated by an inspection of the curves for *not tall*, *not short*, and *neither tall nor short* in Fig. 3.

[17] Crucially, we used the judgments for $100-$*very* ADJ and not those for *not very* ADJ, which we aim to explain, as the latter strategy would have been circular.
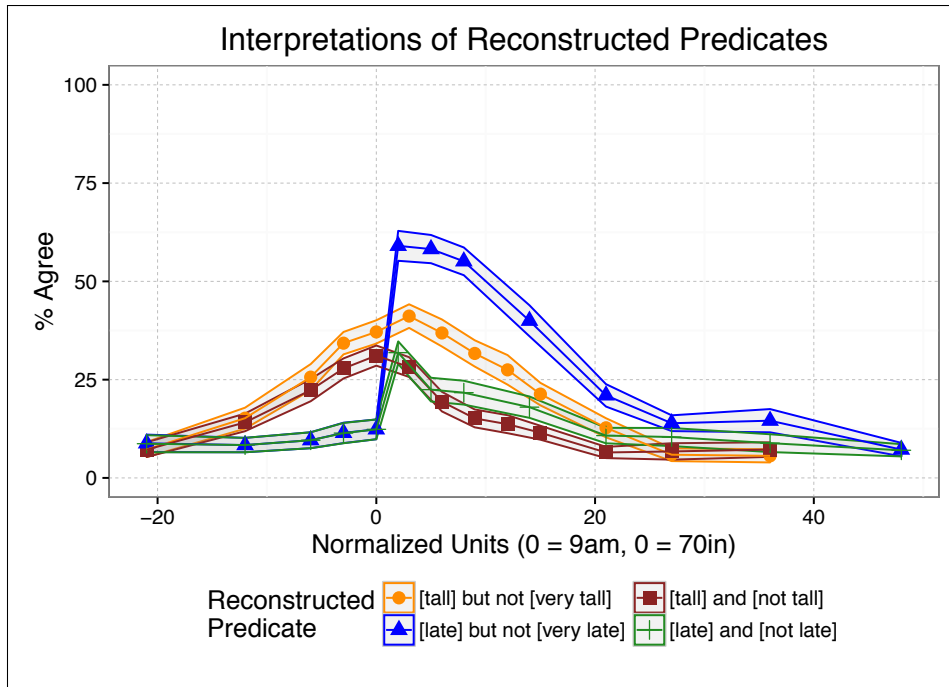
15

Figure 4: Estimated interpretations of complex expressions which would be potentially problematic to explicitly ask for judgments about. The units are normalized so that 0 is 9am for *late*-predicates, and $70''$ for *tall*-predicates (the average adult American male height). Ribbon width represents standard errors. 1 unit = 1min = 1/3in.

late, and is qualitatively closer to known borderline contradictions. To assess this quantitatively, we looked at the highest values a reconstructed predicates would take across the whole scale (a borderline contradiction should maintain low acceptability across the scale, whereas a non-contradictory proposition should have at least some range during which it has a reasonably high acceptability rate). A *t*-test on these reconstructed peak values across the scale for each participant in each construction confirmed the observation: mean estimated peak rating was significantly lower for *tall* $\wedge$ ¬*very tall* than it was for *late* $\wedge$ ¬*very late* ($t(68) = 3.7$, $p < .001$). Furthermore, while the calculated position of the peak for *late* $\wedge$ ¬*very late* was 9:02am for more than half of the participants (with very limited variation: $sd = 2.1 \approx 2$min), the most frequent peak for *tall* $\wedge$ ¬*very tall*, 71in (180cm), corresponded to only 8 participants and there was considerably more variation around this value ($sd = 6.0 \approx 2$in/5cm).

In sum, the interpretive difference between *not very tall* and *not very late* follows from the fact that deriving the inference from *nvt* would result in too much uncertainty, whereas the absolute nature of *late* ensures agreement on at least a part of the resulting strengthened interpretation: while few (if any) heights seem to count as 'tall but not very tall', and there is variation in what exactly those heights are, it is clear across the board that 2min late qualifies as 'late but not very late'.

Notice that the proposal above explains why the implicature to the positive form is not derived for relative adjectives. However, we have seen that in our data the interpretation of *not very tall* is actually stronger than the mere negation of *very tall*, what we called the "unexpected strengthening" puzzle above. Therefore, something else needs to be said if we are to derive this stronger meaning. Evaluativity and understatement may play a role (Krifka 2007; Horn 1989), although the exact mechanism remains to be determined and as discussed above, it is unlikely that it could explain

all of the relevant cases. A solution to this final aspect of the puzzle is left as a topic for future investigation.

4 Conclusion

Sentences like *John is not very tall* are often interpreted as roughly synonymous with *John is not tall*, a puzzle noted in previous literature and treated as some form of euphemism. In this paper, however, we observed that in a systematic set of cases, there is instead an inference from *not very* ADJ to the positive form ADJ. We showed that (*a*) the inference of 'ADJ' from *not very* ADJ is sensitive to the scale structure associated with ADJ – more specifically to the absolute/relative distinction; furthermore, we showed that (*b*) when present, the inference is derived as an implicature via manner-based Gricean reasoning.

We gathered experimental data which confirm the original puzzle about *tall* versus *late*, as well as our hypothesis that when present, inference to the positive form is a manner-based structural implicature. Because of the systematic nature of the difference between vague relative adjectives and precise absolute adjectives, we argued that the phenomenon must be accounted for as an interaction between the mechanisms at the source of implicature and of vagueness, and showed how a well motivated constraint on implicature derivation interacts with scale structure to derive the facts directly.

This analysis highlights the fact that uncertainty in one interpretive domain – the application of gradable predicates to objects – systematically affects uncertainty in another domain – whether a meaning should be strengthened via an implicature. Bayesian approaches to semantics and pragmatics (see e.g. Lassiter & Goodman 2014; Qing & Franke 2014 for applications to gradability) would seem ideally suited for constructing more technically sophisticated models for this kind of interaction, but we have shown how rather simple hypotheses may explain a complex pattern of interaction between semantic and pragmatic modules.

References

Alxatib, Sam & Jeff Pelletier. 2011. On the psychology of truth-gaps. In *Vagueness in communication*, 13–36. Springer.

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

Barker, Chris. 2006. Vagueness. *Encyclopedia of Language and Linguistics, 2nd ed* .

Bartsch, Renate & Theo Vennemann. 1972. The grammar of relative adjectives and comparison. *Linguistische Berichte* 20. 19–32.

Bierwisch, Manfred. 1989. The semantics of gradation. In *Dimensional adjectives: Grammatical structure and conceptual interpretation*, 71–261. Springer-Verlag.

Bolinger, Dwight. 1972. *Degree words*, vol. 53. Walter de Gruyter.

Chemla, Emmanuel. 2009. Similarity: Towards a unified account of scalar implicatures, free choice permission, and presupposition projection. Under Revision for *Semantics and Pragmatics*.

Chemla, Emmanuel & Jacopo Romoli. 2015. The role of probability in the accessibility of scalar inferences. *manuscript* .

Cresswell, M. J. 1977. The semantics of degree. In Barbara Partee (ed.), *Montague grammar*, 261–292. New York: Academic Press.

Egré, Paul, Vincent de Gardelle & David Ripley. 2013. Vagueness and order effects in color categorization. *Journal of Logic, Language and Information* 22(4). 391–420. doi:10.1007/s10849-013-9183-7. http://dx.doi.org/10.1007/s10849-013-9183-7.

Foppolo, Francesca & Francesca Panzeri. 2011. Do children know when their room counts as clean. *Proceedings of NELS42. Amherst: GLSA Publications* .

Fox, Danny & Martin Hackl. 2006. The universal density of measurement. *Linguistics and Philosophy* 29. 537–586.

Fox, Danny & Roni Katzir. 2011. On the characterization of alternatives. *Natural Language Semantics* 19(1). 87–107.

Grice, Paul. 1975. Logic and conversation. In *The Logic of Grammar*, 64–75. Dickenson.

Heim, Irene. 1985. Notes on comparatives and related matters. Ms., University of Texas.

Heim, Irene. 2000. Degree operators and scope. In *Proceedings of SALT 10*, 40–64.

Horn, Laurence R. 1989. *A natural history of negation*, vol. 960. Chicago: University of Chicago Press.

Katzir, R. 2007. Structurally-defined alternatives. *Linguistics and Philosophy* 30(6). 669–690.

Katzir, Roni. 2014. On the roles of markedness and contradiction in the use of alternatives. In *Pragmatics, semantics and the case of scalar implicatures*, 40–71. Springer.

Kennedy, Christopher. 1999. *Projecting the Adjective: The Syntax and Semantics of Gradability and Comparison*. Garland.

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

Kennedy, Christopher & Louise McNally. 1999. From event structure to scale structure: Degree modification in deverbal adjectives. In Tanya Matthews & Devon Strolovitch (eds.), *Semantics and linguistic theory 9*, 163–180. Ithaca, NY: CLC Publications.

Kennedy, Christopher & Louise McNally. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language* 81(2). 345–381.

Klein, Ewan. 1980. A semantics for positive and comparative adjectives. *Linguistics and Philosophy* 4(1). 1–45.

Krifka, Manfred. 2007. Negated antonyms: Creating and filling the gap. In *Presupposition and implicature in compositional semantics*, 163–177. Springer.

Lasersohn, Peter. 1999. Pragmatic halos. *Language* 75(3). 522–551.

Lassiter, Daniel & Noah Goodman. 2014. Context, scale structure, and statistics in the interpretation of positive-form adjectives. In *Proceedings of semantics and linguistic theory*, vol. 24, 587–610.

Maris, Eric & Robert Oostenveld. 2007. Non-parametric statistical testing of eeg-and meg-data. *Journal of neuroscience methods* 164(1). 177–190.

Matsumoto, Yo. 1995. The conversational condition on horn scales. *Linguistics and philosophy* 18(1). 21–60.

Qing, Ciyang & Michael Franke. 2014. Gradable adjectives, vagueness, and optimal language use: A speaker-oriented model. In *Semantics and linguistic theory*, vol. 24, 23–41.

Ripley, David. 2011. Contradictions at the borders. In Rick Nouwen, Robert van Rooij, Uli Sauerland & Hans-Christian Schmitz (eds.), *Vagueness in communication: International workshop, vic 2009, held as part of esslli 2009, bordeaux, france, july 20-24, 2009. revised selected papers*, 169–188. Berlin, Heidelberg: Springer Berlin Heidelberg.

Sauerland, Uli. 2004. Scalar implicatures in complex sentences. *Linguistics and Philosophy* 27(3). 367–391.

Schlenker, Philippe. 2008. Be articulate: A pragmatic theory of presupposition projection. *Theoretical Linguistics* 34(3). 157–212.

Serchuk, Phil, Ian Hargreaves & Richard Zach. 2011. Vagueness, logic and use: Four experimental studies on vagueness. *Mind & Language* 26(5). 540–573.

Simons, Mandy. 2001/2013. On the conversational basis of some presuppositions. In *Perspectives on linguistic pragmatics*, 329–348. Springer.

Solt, Stephanie. 2015. Measurement scales in natural language. *Language and Linguistics Compass* 9(1). 14–32.

Solt, Stephanie & Nicole Gotzner. 2012. Experimenting with degree. In *Proceedings of salt*, vol. 22 166187, 353364.

von Stechow, Arnim. 1984. Comparing semantic theories of comparison. *Journal of Semantics* 3. 1–77. doi:10.1093/jos/3.1-2.1.

Syrett, Kristen, Christopher Kennedy & Jeffrey Lidz. 2010. Meaning and context in children's understanding of gradable adjectives. *Journal of semantics* 27(1). 1–35.

Unger, Peter. 1975. *Ignorance*. Clarendon Press.

Zadeh, Lotfi A. 1965. Fuzzy sets. *Information and control* 8(3). 338–353.

Zehr, Jérémy. 2014. *Vagueness, presupposition and truth-value judgments*: École Normale Supérieure de Paris dissertation.