Are gaps preferred to gluts?
A closer look at borderline contradictions*

Paul Égré
(IJN / ENS / SCAS)

Jérémy Zehr
(UPenn)

Abstract

This paper examines the acceptance of so-called borderline contradictions involving vague adjectives. A close look at the available data from previous studies indicates a preference for "gappy" descriptions of the form "x is neither P nor not P" over "glutty" descriptions of the form "x is P and not P". We present the results of an experiment in which we tested for that difference systematically, using relative gradable adjectives. Our findings confirm that both kinds of descriptions are accepted, but indeed that "neither"-descriptions are to a large extent preferred to "and"-descriptions. We examine several possible explanations for that preference. Our account relies on the distinction proposed by Cobreros et al. between *strict* and *tolerant* meaning for vague adjectives, as well as on a specific implementation of the strongest meaning hypothesis endorsed by Cobreros et al. as well as Alxatib and Pelletier. Our approach, however, argues in favor of local pragmatic strengthening instead of global strengthening in order to derive that preference.

1 Introduction

Several experiments in the last decade indicate that sentences that correspond to literal contradictions from the standpoint of classical logic are accepted to a significant extent by naive subjects to describe *borderline cases* of a vague predicate, in a way that they are not for so-called *clear cases* ([Ripley, 2011, Alxatib and Pelletier, 2011a, Serchuk et al., 2011, Egré et al., 2013]). For example, [Ripley, 2011] found that sentences like (1-a) and (1-b) are accepted significantly more by subjects presented with a circle-square pair at middling distance from one another, compared to a circle and a square touching each other, or to a circle and a square appearing further away from each other. Similarly, [Alxatib and Pelletier, 2011a] found that sentences of the form (2-a) and (2-b) are checked "True" significantly more by participants when pertaining to a man appearing of height 5'11" than for a man appearing of height 6'6" or for a man of height 5'4".

(1) a. The circle is and isn't near the square.
b. The circle neither is nor isn't near the square. [Ripley, 2011]

(2) a. Man x is tall and not tall.
b. Man x is neither tall nor not tall. [Alxatib and Pelletier, 2011a]

Similar sentence forms are intuitively unacceptable when involving *crisp* predicates (see (3)), setting aside cases of presupposition failure as in (4). Possibly, (4-b) might be acceptable, but only to

*[Version of 16/02/2016]. Acknowledgments.

convey that $\sqrt{2}$ is outside the domain of application of the predicate "prime number", assuming the latter applies to only integers (see [Zehr, 2014] for more on the difference between vagueness and presupposition failure):[1]

(3) a. # 9 is and isn't a prime number.
b. # 9 neither is nor isn't a prime number.

(4) a. # $\sqrt{2}$ is and isn't a prime number.
b. (?) $\sqrt{2}$ neither is nor isn't a prime number.

Prima facie, the acceptability of contradictory sentences in borderline cases (so-called "borderline contradictions", following [Ripley, 2011]'s terminology) may not appear so surprising, since a borderline case of a vague predicate is often characterized as a case for which one feels equally attracted toward applying and toward denying the predicate (e.g. [Peirce, 1902]). From a behavioral point of view, however, that characterization is compatible with subjects systematically rejecting descriptions of the form "x is P and not P", or "x is neither P nor not P", but judging "x is P" half the time, and denying "x is P" the other half. The results of [Ripley, 2011, Alxatib and Pelletier, 2011a] and [Egré et al., 2013] suggest that that picture is inadequate, however. From a semantic point of view, those findings are not easily accommodated by either supervaluationist or subvaluationist theories of vagueness, which predict sentences of the form "x is P and not P" and "x is neither P nor not P" to remain contradictions in borderline cases (see [Ripley, 2013, Alxatib and Pelletier, 2011a]). They are more easily accommodated in paraconsistent-friendly frameworks or in fuzzy logic, however, that is in theories in which classical contradictions do not automatically receive the value False. One framework of particular interest for us is the so-called strict-tolerant framework ([Cobreros et al., 2012]), in which both kinds of sentences can be true tolerantly in borderline cases.

In this paper, we propose to dig further into the explanation of borderline contradictions. In what follows we shall refer to sentences of the form "x is P and not P" as *conjunctive* descriptions of borderline cases, or *conjunctions* for short, and to sentences of the form "x is neither P nor not P" as *negated disjunctive* descriptions, or *negated disjunctions* for short. More often, we will refer to them simply as "and"-descriptions and "neither"-descriptions. Although we shall sometimes use the expressions "glutty" descriptions and "gappy" descriptions, we will favor the "and" vs. "neither" denominations, which in a sense are the most theory-neutral. The phenomenon we are interested in is whether the two kinds of descriptions are equally accepted in borderline cases, or whether there is a preference for one kind of description over the other. Our intuition tells us that negated disjunctions might be preferred to conjunctions, that is, that it might be easer to describe a borderline case as "neither tall nor not tall" than as "tall and not tall".

A closer look at extant results suggest that this is likely to be the case (see the review section 2). The difference, however, has not been an object of attention in previous studies. We propose to test for that preference. One motivation to do so is that the difference is susceptible to cast further light on the selection between two kinds of meaning for vague predicates. [Alxatib and Pelletier, 2011a, Alxatib and Pelletier, 2011b] and [Cobreros et al., 2012] both treat vague predicates as pragmatically ambiguous between a strong and a weak meaning.

[1] Sentence (4)-b may be judged outright false of course, since $\sqrt{2}$ *is not* a prime number. It seems to us acceptable in a context in which a teacher, let us say, would want to cut short a dispute between two pupils, one of them arguing that $\sqrt{2}$ is a prime integer, and the other arguing that $\sqrt{2}$ is not a prime integer, but mistakenly thinking it is an integer. Both pupils would wrongly presuppose that $\sqrt{2}$ is an integer, and the point of (4)-b would be to reject that presupposition.

[Alxatib and Pelletier, 2011a, Alxatib and Pelletier, 2011b] and [Cobreros et al., 2012] both formulate the hypothesis that the strong meaning ought to be selected first, but do not look at whether the principle should entail a preference for negated disjunctions over conjunctions. In section 3, we present the results of an experiment confirming our main intuition. In section 4, we present a specific algorithm intended to account for that result. This algorithm too implements the idea that the strict meaning is selected before the tolerant meaning, but importantly it rests on the idea that pragmatic strengthening is done locally, rather than globally for whole sentences.

2 Gaps and Gluts: a Brief Review

Borderline cases of a vague predicate are commonly described either as cases leaving a gap between the positive extension of the predicate and its negative extension, or in a dual manner, as cases where the positive extension of the predicate and its negative extension overlap, thereby creating a glut (see [Fine, 1975, Egré et al., 2013]). Intuitively, a negated description of the form "x is neither P nor not P" matches the idea of a gap between the positive and the negative extension, and a conjunctive description of the form "x is both P and not P" the idea of a glut. The first question we seek to clarify is whether "glutty" and "gappy" descriptions of borderline cases are used to the same extent.

[Ripley, 2011] presented participants with two kinds of conjunctive and disjunctive descriptions, which he calls elided vs. non-elided. In an elided conjunction, the conjunction is internal to the VP ("The circle both is and isn't near the square"), whereas in a non-elided conjunction, it is sentential ("The circle is near the square and it isn't near the square"). Whether in what Ripley calls non-elided disjunctions ("The circle neither is near the square nor isn't near the square"), or elided disjunctions ("The circle neither is nor isn't near the square"), the disjunction is VP-internal, but part of the VP is elided or not. Ripley presented participants with 7 pairs of a circle and square at varying distances from one another, and asked them to rate each description for each stimulus on a 1-7 scale, with 1 labeled 'Disagree' and 7 labeled 'Agree'. In Figure 2, we give a summary of his data, where we aggregated scores for so-called elided vs. non-elided description types. Prima facie, we see no preference for one description type over the other, neither globally, nor in the specific case of stimulus C which receives the highest assent to both description types.

On the other hand, we do observe an overall preference for disjunctions over conjunctions in the experiment run by [Alxatib and Pelletier, 2011a]. Alxatib and Pelletier's methodology was different from Ripley's, since participants had to check True, False, or Can't Tell to four kinds of description including "Tall", "not Tall", "Tall and not Tall", and "Neither tall nor not tall". In Figure 2, we reproduce the proportions of 'True' checks to the latter two kinds, namely conjunctions and disjunctions. What we can observe is a higher level of 'True' checks to disjunctions, in each stimulus. For the stimulus of middling height in their stimulus set, [Alxatib and Pelletier, 2011b] report the following data, reproduced in Table 1, where Tn, Fn, and Cn stand for the numbers of participants responding "True", "False" and "Can't tell" to "#2 is neither tall nor not tall", and where Tb, Fb, and Cb give the corresponding numbers for "#2 is both tall and not tall". As they point out, a McNemar-Bowker test for symmetry gives a value of 8.04, with $p < 0.05$, suggestive of a difference between the two descriptions.

Serchuk and colleagues also investigated borderline contradictions, this time comparing more explicitly than Ripley the effect of having two kinds of negation, VP-internal or sentential. Unlike Ripley or Alxatib and Pelletier, they did not use perceptual stimuli, but asked participants to

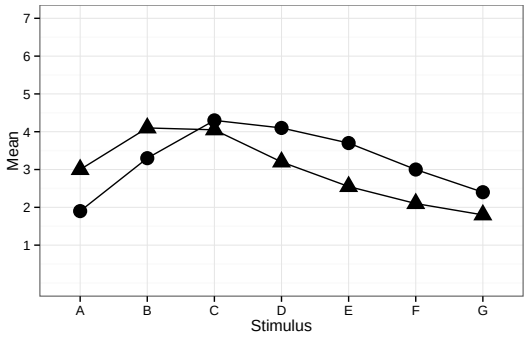

Figure 1: Ripley's data: triangles represent aggregate scores for disjunctions, and squares aggregate scores for conjunctions

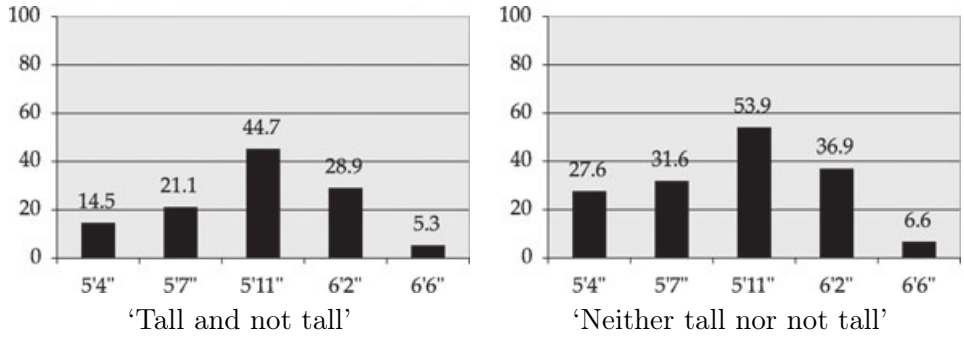

Figure 2: Alxatib and Pelletier's data: proportion of 'True' checks in conjunctions and disjunctions

imagine borderline cases, describing those in terms of the operator "clearly". For instance, in one of their scenarios, a woman named Susan is described as being "somewhere between women who are clearly rich and women who are clearly non-rich". For disjunctions, they used unnegated disjunctions of two kinds: "Either Susan is rich or Susan is not rich", and "Susan is rich or it is not the case that Susan is rich". Each time, however, the disjunction is sentential and not VP-internal: they did not present participants with "Susan is rich or not rich". They used a similar pair of probes for conjunctions, namely "Susan is rich and Susan is not rich", and "Susan is rich and it is not the case that Susan is rich" (here too, they did not use "Susan is rich and not rich"). Participants had to check exactly one answer among six possible answers for each sentence, within the set consisting of {true, not true but also not false, partially true and partially false, false, both true and false, true or false but I don't know which}. In Table 2, we present Serchuk et al.'s data in a table in which we collapsed the responses other than True and False under a third category "Other". In this case, we cannot directly compare "True" responses to "neither" vs. "both". However, we can see that there is a higher proportion of False answers to the disjunction "either x is P or x is not P" than of True answers to "x is P and not P", and similarly when we compare False answers to "x is P or it is not the case that x is P" to true answers to "x is P and it is not the case that x is

	Tn	Fn	Cn	Total
Tb	22	12	0	34
Fb	13	18	0	31
Cb	6	2	3	11
Total	41	32	3	76

Table 1: Alxatib and Pelletier's data for the central stimulus of height 5'11"

P". A Fisher test based on that comparison yields a very significant difference for the former two sentences ($p < 10^{-10}$) as well as for the latter two ($p < 10^{-16}$).

Description	'True'	'False'	Other
Either x is P or x is not P	113	137	100
x is P or it is not the case that x is P	141	88	121
x is P and x is not P	66	195	89
x is P and it is not the case that x is P	25	247	78

Table 2: Serchuk et al.'s data on disjunctive and conjunctive descriptions of borderline cases

Finally, we mention the results of an unpublished study, by [Solt and Gotzner, 2010]. They showed participants picture series depicting either suitcases of various sizes, bluejeans of varying price, or cities of different distances from Berlin. The goal of the experiment was to see which pictures in each series would be classified by participants as satisfying the respective adjectives "big" (*gross*), "expensive" (*teuer*), "far" (*weit*) in comparison to their negation or the polar antonym. In one condition participants had to decide which pictures satisfied the adjective, and which satisfied the syntactic negation. In another condition a different group had to decide which pictures satisfied the adjective, and which satisfied the polar antonym. Their results indicate that only a small minority of participants left no gap between the adjective and its opposite (whether polar or syntactic), or ascribed the same item both descriptions. Although that study does not investigate the acceptance of complex sentences like the ones we are interested in, it supports the idea of a preference for gaps over gluts when people are asked to judge of a sentence and its negation separately.

What this short review indicates, therefore, is a preference for the description of borderline cases in terms of disjunctive descriptions ('neither P nor not P") over conjunctive descriptions ('P and not P'), and plausibly, we should see the contrast enhanced with polar antonyms used instead of the syntactic negations. That is, participants should accepts gluts even less with polar antonyms than with syntactic negations, and conversely they should accept gaps more with polar antonyms than with syntactic negations. Because the data we reviewed are partial and unsystematic, we proceed to test that hypothesis more systematically in what follows.

3 Experiment

In order to test the preference for "gappy" descriptions over "glutty" descriptions of borderline cases, we designed an experiment intended to compare the acceptance of both kinds of descriptions. The

experiment we report here is the third and main one in a series of three, and it replicates the results of the two previous, pilot versions. We presented participants with different scenarios involving different adjectives, each time involving the verbal description of a borderline case, drawing inspiration from [Serchuk et al., 2011], who basically asked participants to imagine borderline cases based on similar verbal descriptions. We then asked participants to judge the adequacy of contradictory descriptions of the forms in (5), leaving them the possibility to either accept or reject the description.

(5) a. *Neither*-descriptions: "BORDERLINE is neither ADJECTIVE nor NOT ADJECTIVE"
b. *And*-descriptions: "BORDERLINE is ADJECTIVE and NOT ADJECTIVE"

Our prediction was that we should see a higher rate of acceptance of "neither"-descriptions over "and"-descriptions, but also that either should be significantly more accepted than outright false sentences. In order to get a homogenous set of results, however, our first step was to define general principles for the selection of adjectives and for the construction of our scenarios.

3.1 Design

The experiment was a block design: each participant faced two series of descriptions. In one block, the descriptions were as exemplified in (5). In the other block, NOT ADJECTIVE was replaced with a corresponding antonym (e.g. *short* in place of *not tall*). In this paper, we do not report the results for these antonymic descriptions: we treat them as fillers, even though they served as critical conditions for the purposes of another study investigating the relation between the two types of negations (syntactic vs. lexical). We report the results of the critical (i.e. syntactic) conditions after collapsing the two blocks.[2]

The design took the form of an acceptability task: participants were presented with a fictive scenario verbally depicting a borderline case on a given gradable dimension and had to tell whether they would accept four descriptions of this borderline case, as exemplified in Figure 3. On each trial, they saw the scenario first, followed by the instruction *Click to see the first description.* The instruction would disappear after clicking and simultaneously reveal the first of the four descriptions to be assessed. Each time, clicking on one of *Yes* and *No* would dynamically let the next description appear on screen. For the last description, checking *Yes* nor *No* would reveal a link needed to validate all four answers of the current trial, reading *Click here to continue.* As long as the descriptions were visible on the screen, the participants were able to modify their judgments. Once they clicked on *Click here to continue*, they would see the next trial, with a new scenario targeting a distinct adjective.

3.2 Materials

3.2.1 Scenarios

[Serchuk et al., 2011] used rather explicit descriptions of borderline cases in their experiments, as exemplified in (6), and then asked their participants to judge a list of sentences about these borderline cases. What follows is an example of the sentences they used:

(6) "Imagine that on the spectrum of rich women, Susan is somewhere between women who are clearly rich and women who are clearly non-rich".

[2]The methods we used to analyze the results are discussed in the *Results* section.

A survey on heights has been conducted in your country. In the population there are people of a very high height, and people of a very low height. Then there are people who lie in the middle between these two areas.

Imagine that Betty is one of the people in the middle range. Comparing Betty to other people in the population, is it true to say the following?

[Click to see the first description]

[Betty is neither tall nor not tall	❍ Yes	❍ No]
[Betty is tall and not tall	❍ Yes	❍ No]
[Betty is taller than at least one person	❍ Yes	❍ No]
[Betty is taller than everybody else	❍ Yes	❍ No]

[→ Click here to continue]

Figure 3: Example of a trial. [Bracketed texts] appeared dynamically: each description appeared after answering the preceding one or clicking Click to see the first description. The set of answers was validated by clicking on Click here to continue (appearing after the last click).

To the extent that their results revealed acceptance of contradictory descriptions, they show that participants readily represent borderline cases when explicitly asked to do so. This motivated us in describing borderline cases in similar verbal ways. However, in contrast with Serchuk et al.'s descriptions, we did not let the target adjectives (e.g. *rich* in (6)) appear in the scenarios, to avoid any priming effect. Indeed, the description in (6) might prime a *clearly* reading of *rich* that would lead to systematically exclude the borderline cases form the positive extension of *rich*. Alternatively, it might prime a contrastive, looser use of *rich* leading one to categorize the borderline cases in the positive extension. If any of these effects were real, they could bias our results in a way that does not appear to be the case in [Alxatib and Pelletier, 2011a]'s pictorial context. Because of that, we formulated our scenarios by referring to the *scale* associated with the adjectives, using non-morphologically-related nouns whenever possible (cf. Fig. 3 where we used the word *heights* to refer to the scale associated with *tall*). We hypothesised that participants would naturally represent the individuals in the middle of these scales as borderline cases for the target adjectives. We take our results as evidence that participants indeed pictured borderline cases, as expected. Note that contrary to pictorial representations, verbal descriptions have the advantage to let subjects build their own, ideal representations of what a borderline case might be on the given scale. The particular scenarios we used are reported in Appendix A.

3.2.2 Selection of adjectives: four principles

Not all adjectives come with borderline cases, and not all adjectives seem to define their borderline area in the same way. In order to have a set of adjectives as homogeneous as possible, we selected them along four criteria that we present here. Our selection was partly guided by principles investigated in the work of [Roche, 2012] and [Ruytenbeek, 2013], done under the supervision of Benjamin Spector, concerning the negation of adjectives.

Gradability. As exemplified in the scenario in Fig. 3 all the borderline cases we described were presented as lying in the central region of *a given scalar dimension*.[3] All the adjectives we used are gradable and therefore define a scale along which their arguments are non-trivially ordered. We say that an adjective is gradable if and only if one the following constructions is perceived as natural.

(7) a. X is ADJECTIVE-er than Y.
b. X is more ADJECTIVE than Y.

For example *tall* is gradable (8-a), whereas *underage* is not (8-b).

(8) a. Bill is taller than Sue.
b. #Bill is more underage than Sue.

Criterion 1: We excluded any non-gradable adjective from our list.

Relativity. Two kinds of gradable adjectives have been distinguished in the literature: relative (gradable) adjectives like *tall* and absolute (gradable) adjectives like *full* [Unger, 1975, Kennedy, 2007]. Following one of Kennedy's tests, we say that a gradable adjective is relative if and only if the following inference is not systematic.

(9) X is the ADJECTIVE one. $\not\rightsquigarrow$ X is ADJECTIVE (generally speaking)
a. Relative: My glass is the tall one. $\not\rightsquigarrow$ my glass is tall (generally speaking)
b. Absolute: My glass is the full one. $\rightsquigarrow$ my glass is full (generally speaking)

For example *tall* is relative because the inference is *not* systematic (9-a), but *full* is absolute because the inference *is* systematic (9-b). For the purpose of our study we chose to include only relative gradable adjectives. Our reason to do so was the following: intuitively, absolute adjectives denote an endpoint on a scale. For example, *full* denotes the maximum extent to which a recipient can be filled with substance. A relative adjective like *tall*, on the other hand, does not select any context-invariant standard on a scale. Instead, it refers to context-sensitive standard. In the way we constructed our stimuli, we always asked participants to imagine people or objects standing in the "middle range" of two extreme regions denoted by "very high" and "very low" along the corresponding dimension. For absolute gradable adjectives, we would have to adapt the descriptions. For example: if we were to ask participants to consider how they would describe glasses in the middle range between those that are "very filled" and those that are "filled very little", we would very likely fail to target the borderline region for what count as "full". Intuitively, the borderline region for "full" is a region that is close enough to the maximum degree to which a recipient can be filled. Another reason we had not to include absolute adjectives in this experiment is because we also looked at antonyms, and here again, we can expect antonyms of absolute adjectives to not behave exactly like antonyms of relative adjectives (see [Burnett, 2012] for discussion). In summary, we did not include absolute adjectives in our experiment mostly to have a homogenous set of descriptions and adjectives to test.[4]

[3] In this paper, we take this *non-extreme* property to be definitory of borderline cases. By contrast, one might consider that an *extremely* insane person in a *perfect* physical shape would be a borderline case for *healthy* to the extent that our judgment for *this person is healthy* would be unclear. As outlined in the third point of this list, such a case does not fall under the concept of *borderline cases* that we use.

[4] Some authors consider that the status of what we call borderline cases for absolute adjectives is due to a different phenomenon from the one at play with relative adjectives. For instance [Kennedy, 2007] claims that calling an almost-full glass "full" is a manifestation of *imprecision*, but that there is in fact a sense in which such a glass is

We return to the question of whether our results would extend to the class of absolute adjectives in section 4 below.

Criterion 2: We excluded absolute gradable adjectives from our list.

Note that as [Kennedy, 2013] notes, all relative adjectives seem to be subjective. That is to say, they seem to systematically allow for faultless disagreement: substituting *tall* for ADJECTIVE in (10), as in (10-a), does not necessarily imply that either Mary or Sue is wrong, but doing the same with *full* as in (10-b) *does* imply that either Mary or Sue is wrong. While Mary and Sue can truly diverge on what heights they consider to be tall, there is an objective threshold for fullness (complete filling) and therefore one of them has to be wrong.

(10) X thinks Z is ADJECTIVE but Y does not.

a. Subjective: Mary thinks Paul is tall but Sue does not $\not\rightsquigarrow$ either Mary or Sue is wrong.

b. Not subjective: Mary thinks the glass is full but Sue does not $\rightsquigarrow$ either Mary or Sue is wrong.

Subjectivity means that the evaluation of the predicate can be relativized to the interpreter's standards. In that sense, one might anticipate acceptance of borderline contradictions of the form *X is* ADJECTIVE *and* NOT ADJECTIVE to result from a *relativistic* interpretation along the lines of *X is* ADJECTIVE *for some interpreter(s) and* NOT ADJECTIVE *for some other interpreter(s)*. One could therefore suspect that we are not actually testing acceptance of contradictions due to *borderlineness*, but acceptance of contradictions simply due to *subjectivity*. For this point to be fair, one would need to show that contradictory descriptions are accepted/rejected to the same extent when formed with *non-vague subjective* predicates. The assumption behind this project is that *there are* subjective predicates that are not vague. However we are not aware of any such predicate.[5] In absence of evidence that vagueness is not reducible to subjectivity, or that non-vague subjective predicates (granted their existence) yield acceptable contradictory descriptions, we do not consider relativistic approaches to undermine the results of our study. As a matter of fact, therefore, all the adjectives we used are subjective adjectives.

Non-Evaluativity. We use the category *evaluative* in the sense of [Kennedy, 2013]. That is, we call an adjective *evaluative* if substituting it for ADJECTIVE in (11) results in a non-deviant sentence.[6] The reader may use the analytical comparison form when required.

(11) Z finds that X is ADJECTIVE-er than Y.

a. Evaluative: Mary finds that Paul is smarter than Joe.

b. Not evaluative: # Mary finds that Paul is taller than Joe.

uncontroversially *not* full. By contrast, for relative adjectives, there would be no non-arbitrary way to settle the question for borderline cases for they are a manifestation of *vagueness*.

[5] At first glance "indexical" words such as *local* might be such instances as their content is interpreter-dependent (local *for whom*?), but it seems to us that (i-a) is well less acceptable than (i-b).

(i) a. Chocolate from Ecuador is local and not local.

b. This man of average height is tall and not tall.

[6] See and [Sæbø, 2009] who first proposed such tests. This sense of "evaluative", although related, is more specific than [Rett, 2007]'s sense, who calls an expression "evaluative if it makes reference to a degree that exceeds a contextually specified standard".

As it appears and as pointed by [Kennedy, 2013], evaluativity is a kind of subjectivity, as it correlates with faultless disagreement. But subjectivity as defined in the previous point seems to result from a range of different possible thresholds on a given, easily identifiable scale and is tightly correlated with borderline cases. By contrast, evaluativity results from a range of different possible *criteria* to build a scale on which one later establishes a threshold. One could be tempted to entertain another relativistic approach where *X is* ADJECTIVE *and* NOT ADJECTIVE is interpreted as *X is* ADJECTIVE *according to some criteria and* NOT ADJECTIVE *according to some other criteria*. To exclude any such explanation of our results,

Criterion 3: we excluded any evaluative adjective from our list.

Individual- vs Stage-level. The last criterion we used is based on the distinction between *individual*-level predicates and *stage*-level predicates. After [Carlson, 1977], we will call an adjective *individual-level* if substituting it for ADJECTIVE in (12) results in a deviant sentence, and *stage-level* otherwise.

(12) There are two *X* ADJECTIVE.

a. Individual-level: ??There are two men tall.

b. Stage-level: There are two men happy.

The properties attributed by individual-level predicates seem to be more inherent to their argument than those attributed by stage-level predicates. Once again, a relativistic approach could try to explain the acceptance of contradictory descriptions under a reading like *X is* ADJECTIVE *on some occasions and* NOT ADJECTIVE *on some other occasions*. To avoid that, we also relied on the following criterion:

Criterion 4: we excluded any stage-level adjective from our list.

It should be noted that we used these tests as a heuristics to build our stimuli. Our intention was not to test for those properties, nor to rigorously control for them. However, we think that having taken those into consideration participated in the success of our experiment.

The 8 adjectives that we used are divided in two types: 4 adjectives qualified human beings (*tall*, *rich*, *heavy*, *old*) and 4 adjectives qualified concrete or abstract objects (*loud*, *fast*, *large*, *wide*). We presented half of the participants with the four adjectives of the former type *first* and the four adjectives of the latter type *second*; the other half of the participants saw the reverse order. This parameter was crossed with the type of negation (syntactic vs. lexical) and each participant responded to *syntactic* descriptions built with either human-oriented or object-oriented adjectives exclusively.

3.3 Participants

We recruited 148 participants online and anonymously via the Amazon Mechanical Turk platform. There a link would redirect them to the Ibex servers, on which the experiment was developed and hosted. Before going through the actual trials, participants first had to complete a pre-questionnaire consisting of seven simple questions. They were also presented with a post-questionnaire that was

used for the study on the syntactic vs antonymic negation. Those forms are reported in Appendix B. Accuracy on the pre-questionnaire and on controls was very good and no participant was excluded.[7]

3.4 Results

The graph in Figure 3.4 reports the average acceptance of each of the four types of descriptions exemplified in Fig. 3. The graph collapses the results along two dimensions:

1. **The block order:** we collapsed the responses from the participants who saw the syntactic descriptions *first* and those from the participants who saw them *second*

   A regression model including block order as a main factor revealed a significant decrease in acceptance of syntactic "and"-descriptions for participants who judged them *after the antonymic block*, dropping from $\approx 50\%$ to $\approx 35\%$

2. **The type of adjective:** we collapsed the responses for the two types of adjectives (human-oriented vs object-oriented)

   A regression model including type of adjective as a main factor revealed a significant increase in acceptance of syntactic "and"-descriptions *for object-oriented adjectives* when judged *before human-oriented adjectives*, jumping from $\approx 37\%$ to $\approx 66\%$

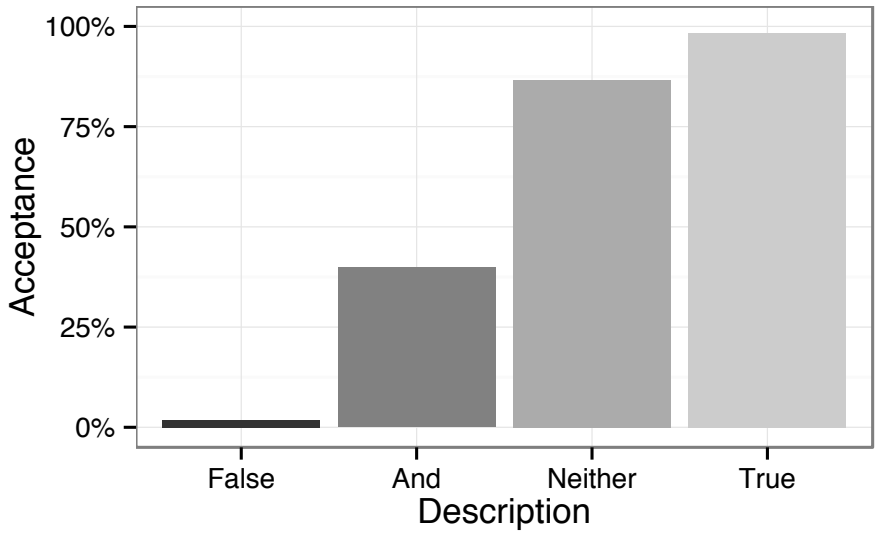

Figure 4: Mean acceptance by description type

Table 3.4 reports the output of three mixed effects logistic regression models that we fitted on a subset of the data, including only the answers to the syntactic trials. They differ in the level used as a baseline for Description. The first report corresponds to a model where *Neither* served as

[7] In this case, exclusion of participants would lead the regression models that we ran to not converge, because of a too low variability on the controls.

the baseline, the second report corresponds to a model where (control) *True* served as the baseline and the third report corresponds to a model where (control) *False* served as the baseline. These models included two parameters as random factors interacting with Description: i) the particular adjective used in the description; and ii) whether the participant judged the antonymic descriptions before the syntactic ones. This ensures that the effects reported in Table 3.4 are exclusively due to Description.[8]

Neither	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.5973	0.3390	10.612	< 2e-16 ***
True	4.2420	1.6383	2.589	0.00962 **
And	-4.3488	0.4842	-8.982	< 2e-16 ***
False	-10.5244	0.7229	-14.559	< 2e-16 ***
True	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	7.839	1.115	7.027	2.10e-12 ***
Neither	-4.242	1.158	-3.664	0.000248 ***
And	-8.591	1.418	-6.057	1.39e-09 ***
False	-14.767	1.193	-12.380	< 2e-16 ***
False	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-6.9274	0.5983	-11.578	< 2e-16 ***
And	6.1755	0.8122	7.603	2.89e-14 ***
Neither	10.5246	0.7023	14.987	< 2e-16 ***
True	14.7677	1.5525	9.512	< 2e-16 ***

Table 3: Outputs of three mixed effects logistic regression models differing in their baseline for *Description*

As is manifest from Table 3.4, the rates of acceptance of each type of description are significantly different from one another. In particular, "and"-descriptions were accepted significantly more often than *control false* descriptions but still significantly less often that "neither"-descriptions, which themselves were not accepted as often as *control true* descriptions.

3.5 Discussion

The results confirm the previous observations from the experimental literature that we reported, according to which speakers do accept contradictory "neither"- and "and"-descriptions to describe borderline cases, given that those were accepted significantly more often than the control false descriptions. However, they provide us with a more nuanced picture of these judgments: first of all, it seems that none of the two types of contradictory descriptions are as acceptable as plainly true (control) descriptions (a point that was left open in the existing experiments); and secondly, they suggest that "neither"-descriptions are preferred to "and"-descriptions, eliciting a difference that was descriptively hinted at in [Alxatib and Pelletier, 2011a] but invisible in [Ripley, 2011]. As a consequence, on the one hand, these observations support the project of developing a system

[8]We used the R package LME4 and its function *glmer*. The specific formula that we used and the outputs corresponding to the random effects are reported in Appendix C.

accounting for the acceptance of both kinds of descriptions but on the other hand, such a system should also account for the preference of "neither"-descriptions over "and"-descriptions.

4 General discussion

Our two main predictions in undertaking this experiment were confirmed: first, we see that both "and"-descriptions and "neither"-descriptions are accepted significantly more than false control sentences using the same predicates, and secondly we see a marked preference for "and"-descriptions over "neither"-descriptions. This finding raises two main questions: the first is whether we can expect it to be robust across other kinds of adjectives than the ones we considered. The second is how we might explain the preference for "neither"-descriptions over "and"-descriptions, even as restricted to the class we considered in our study.

4.1 Absolute vs. relative adjectives

Regarding the first point: because we deliberately included only relative gradable adjectives in our study, we cannot be sure that the same pattern would show with absolute adjectives. When designing the experiment, we simply neglected to treat the relative-absolute distinction as a factor, for the reasons explained in section 3. This is a limitation of our current study.

Indeed, as emphasized already, absolute adjectives also admit borderline cases. The difference with relative adjectives is that the borderline region of an absolute adjective ought to be close to the maximum point denoted by the adjective (in the case of maximum-point adjectives like *full* or *dry*).[9] Intuitively, a glass that is borderline full is a glass that is almost full, it certainly isn't a glass that is half-full and half-empty. Likewise, a borderline case of a *dry* towel would appear to be a towel that is close enough to a towel that is completely dry (such as a towel with a few drops of water on it, or one that has almost dried up). If we were to ask participants to imagine such borderline cases, it is not clear to us whether the same preference for "neither" descriptions over "and" descriptions would be observed. Our own intuition is actually the opposite. Consider the following pair, of a glass that is almost full:

(13) a. This glass is full and not full.
b. This glass is neither full nor not full.

Our sense is that it might be easier and more natural to use (13-a) to refer to such an *almost full* glass. (13-b), on the other hand, might be easier to use to refer to a glass that is actually *half-full*. Not all speakers appear to share our intuition, however, and we are now in the process of designing a follow-up experiment to test for absolute adjectives.

Pending the results of such an experiment, we can only speculate. If our intuition about the last pair proved correct, on the other hand, this would suggest that "neither"-descriptions are preferred to "and"-descriptions provided the borderline case of the adjective is not located too close to one of the endpoints on the relevant scale. In the experiment we presented in the last section, this hypothesis is automatically met, since the item is always described as lying "in the middle between" two extreme regions of a population distribution. One possibility, therefore, might simply be to say that the interpretation of the adjective and its negation, in the case of relative adjectives like "tall", is primarily determined by the ends of the distribution that participants are asked to consider

[9] The latter are known as *total* absolute adjectives, see [Burnett, 2012] for a recent exposition.

("people with a very high height" for "tall", and "people with a very low height for "not tall"). This interpretation is not implausible, but then it leaves us with the task of explaining why "and"-descriptions are acceptable at all, rather than mostly rejected as false. Some other mechanism must be considered then.

4.2 Omission bias and Consistency Bias

Let us focus on our current sample of relative adjectives from now on. One way in which one might be tempted to explain the preference for "neither"-sentences over "and"-sentences to describe a borderline case is as an instance of a domain-general bias toward omission rather than commission (see [Spranca et al., 1991], [Bonini et al., 1999]). The explanation might go like this: a borderline case of a tall person is one for which participants feel uncertain whether to apply the predicate "tall" as opposed to "not tall". Participants may feel uncertain because, as postulated by [Bonini et al., 1999], they may think that as a matter of fact only one of those descriptions is correct, but both descriptions compete on their mind. Intuitively, "neither"-descriptions can be taken to adequately express the participants' reluctance to ascribe either "tall" or "not tall". Instead of committing themselves to either the predicate or its negation, the participants express a preference for omitting both.

That explanation has a ring of truth, but it can't be quite adequate. One assumption that appears inadequate is that if participants felt that only one of the predicates "tall" and "not tall" ought to apply, then they should massively check "No" for descriptions like "x is tall and not tall". Unlike [Alxatib and Pelletier, 2011a], we did not leave participants with the option to say "Can't tell" in response to the sentences. But as our data indicate, however, more than 37% of participants checked "Yes" to "and"-descriptions. This confirms that an explanation in the style of Bonini et al. cannot be right: if we follow that explanation we can no longer explain why participants accept glutty descriptions to the extent that they do.

In light of out data, a different way of articulating the omission bias hypothesis may be as follows: participants do not feel that there is a fact of the matter dictating that only one description in terms of "tall" or "not tall" should be the correct one, but they may feel that that once you commit yourself to one description, you should avoid using the other. In other words, participants may simply have a bias toward consistency. Although they feel that "tall" and "not tall" are equally applicable of a borderline case, they find more adequate to say that neither description applies, to avoid an inconsistency, rather than to say that both descriptions apply. In other words, the participants' behavior would simply reflect a preference for incompleteness over inconsistency.

That explanation is a bit more convincing, but it still strikes us as *ad hoc*. It assumes that participants would interpret "tall" in a fixed way across both occurrences of "tall" and "not tall". What if participants were contextualists instead? They may very well understand a sentence like "Betty is tall and not tall" to mean: "Betty is tall is some sense, and he is not tall in some other sense". If really participants think that no single way of drawing a line between "tall" and "not tall" people is correct, then this would be a very rational interpretation for them to use. It also appears to be a natural interpretation. But if participants have that interpretation in mind, then they should actually show a preference for "and"-descriptions over "neither"-descriptions. The reason is that, from a logical point of view, "Betty is neither tall nor not tall" should be taken to mean "Betty is neither [tall in some sense of "tall"] nor [not tall in some sense of "tall"]". But this time the latter interpretation would be a contradiction. Another way to cash out that argument is in subvaluationist terms. "Betty is tall in some sense of "tall" and Betty is not tall in some other sense of "tall"" expresses the idea that the sentence "Betty is tall" is true in some precisifications, and

false in others. But under that interpretation "Betty is neither tall nor not tall" should mean that there is no precisification where "Betty is tall" is true, and likewise that there is no precisification where "Betty is not tall" is true: this is a contradiction if indeed participants accept that "Betty is tall" and "Betty is not tall" are true according to some ways of delineating the meaning of "tall".

In summary, we think that an explanation in terms of an omission bias can't be adequate, because it should predict that "and"-descriptions are not accepted at all. And an explanation in terms of a preference for incompleteness over inconsistency seems to us *ad hoc* by assuming that participants would interpret "tall" rigidly, in a way that does not seem supported by the relative acceptance of sentences of the form "Betty is tall and not tall". This leaves us with a puzzle similar to the kind of puzzle that [Alxatib and Pelletier, 2011a] have set forth: intuitively, "Betty is tall and not tall" is easily understandable to mean that "Betty is tall" in some acceptable sense of "tall", and also "not tall" in some other acceptable sense of "tall". But then we cannot explain why "Betty is neither tall nor not tall" is acceptable at all. Some other explanation is therefore needed.

4.3 Strict meaning vs. Tolerant meaning

In order to explain the findings of our study, we therefore turn to a distinct set of assumptions, and basically adopt the working assumption of [Alxatib and Pelletier, 2011a] and [Cobreros et al., 2012] according to which vague adjectives are pragmatically ambiguous between two interpretations, a *tolerant* and a *strict* interpretation.[10]

We shall not review all the extant evidence for the distinction between strict and tolerant interpretations, but we only point out one of the earlier findings by [Alxatib and Pelletier, 2011a], which is that a significant proportion of their participants who checked True to the description of a borderline tall man as "tall and not tall" also checked False to the separate descriptions "tall" and "not tall".[11] The way that particular finding is explained by [Alxatib and Pelletier, 2011a] as well as [Cobreros et al., 2012] is by appeal to the Strongest Meaning Hypothesis (SMH), namely the hypothesis that among two ambiguous meanings, the default should be to choose the strongest non-trivial meaning. In the case in question, the strongest non-trivial meaning that can be given to a sentence like "Betty is tall and not tall" is the conjunction of the tolerant meanings of "Betty is tall" and "Betty is not tall" respectively, whereas the strongest non-trivial meaning that can be given to the conjuncts separately is the strict meaning of "Betty is tall" and "Betty is not tall". For a speaker obeying the SMH, it is therefore consistent to accept "Betty is tall and not tall", and to reject "Betty is tall" and "Betty is not tall", because the interpretation of "tall" and "not tall" switches from tolerant to strict from the conjunction to the conjuncts.

What the SMH encapsulates, therefore, is not a domain-general bias (like the omission bias), or an overall logical bias (such as a preference for incompleteness over inconsistency), but rather a bias toward selecting the most informative meaning. We think the SMH can also be used to account for the preference of "neither"-descriptions over "and"-descriptions in our experiment. Our basic idea is simple: the idea is that when asked to decide whether a sentence containing a positive adjective or its negation is true or false, participants should have a bias toward selecting the strict meaning of the adjective and of its negation first. Only secondarily will they consider the tolerant meaning. To

[10] Alxatib and Pelletier talk of sub- and super-interpretation. As pointed out by [Cobreros et al., 2012], we can talk of sub- and super-interpretations, but provided we do not mistake the resulting logic of vague predicates for the subvaluationist and supervaluationist logics respectively, which are not compositional, unlike the strict-tolerant logic used in [Cobreros et al., 2012]. See [Ripley, 2013] and [Alxatib et al., 2013] for discussion and comparison.

[11] See also [Egré et al., 2013] for discussion, where a related phenomenon is discussed under the name "Hump Effect".

get the details right, however, we need to spell out some assumptions.

4.4 Strict first

We adopt the three-valued presentation of strict and tolerant meanings given in [Cobreros et al., 2015a]. Given a propositional or first-order language, and a three-valued model, we call a sentence strictly true if it takes the value 1, and tolerantly true if it takes a value at least $1/2$ in that model, and we adopt the strong Kleene rules for the connectives. Given a vague predicate P, we represent the fact that a is a borderline case of P by assigning the value $1/2$ to Pa in that model. We now state our specific assumptions in order to explain the preference of "neither"-descriptions over "and"-descriptions.

(i) Local strengthening. Our first assumption is that predicates are *locally* interpreted as tolerant or as strict at the subsentential rather than the sentential level. We basically posit strict and tolerant operators S and T whose semantics is defined as below in table 4.4:

ϕ	$T\phi$	$S\phi$
0	0	0
#	1	0
1	1	1

Table 4: Truth-tables for the S and T operators

Local strengthening is a natural idea since strict and tolerant meanings seem to have adverbial reflects:

(14) Betty is (somehow) tall and (somehow) not tall

(15) Betty is neither (clearly) tall nor (clearly) not tall

We do not say that *somehow* and *clearly* correspond to the strict and tolerant operators that we just introduced. What (14) and (15) show is that local modifications of the adjectives are a productive linguistic operation. In the rest of the discussion, we will assume the existence of two covert linguistic operators *strictly* and *tolerantly* that can appear in place of *somehow* and *clearly* and that respectively have the effects of T and S.

The algorithm must be such as to output the following interpretations for the above sentences:[12]

(16) $T(tall(a)) \wedge T(\neg tall(a))$

(17) $\neg(S(tall(a)) \vee S(\neg tall(a)) \equiv \neg S(tall(a) \wedge \neg S(\neg tall(a))$

The assumption of local strengthening is a significant departure from most accounts we can think of, where the pragmatic meaning of a sentence is computed globally for the whole sentence (see [Cobreros et al., 2012], [Alxatib et al., 2013], [Cobreros et al., 2015b]). We discuss that assumption in greater detail below.

[12]Since the algorithm we propose operates on high-level linguistic representations, it equally derives the expected interpretation regardless of the form of the logical translation of the *neither* descriptions.

(ii) Predicate negation. To flesh out assumption (i), we supplement it with another one, which concerns the behavior of negation. We take it that the operators do not get embedded inside a negated predicate (as in *not tall*) except if specifically marked (as in *not* $\underline{tall}_{Focus}$ where the short break between *not* and *tall* might signal the presence of the *strictly* operator).

(iii) Bottom-up strengthening. Our third assumption is that strongest meanings are computed incrementally in the course of building the syntactic representation of a sentence. The idea is that first, the leaves of a syntactic tree are given the strongest meanings. Then, given two subconstituents of a larger constituent, their meaning composition gets a check for nontriviality. We call a meaning trivial if it is necessarily empty or necessarily tautological. If it is nontrivial, the algorithm proceeds according to classical rules in order to deliver a semantic verdict relative to the model at hand. If a triviality is reached, then one needs to backtrack and reassign the leaves of the tree the next strongest meaning available in order to reiterate the algorithm, until the algorithm ends, before giving a semantic verdict.

(iv) Least effort In order to explain the preference for "neither"-descriptions of borderline cases over "and"-descriptions, we need a final assumption, which is that the simplest of two computational procedures should be generally preferred to the more complex. Or to put it differently, if a run of the algorithm involves backtracking, it will involve a more costly representation of meaning, and participants will be less likely to compute it. That is, if a triviality is reached in the course of building the meaning of a sentence, a participant can always be lazy and deliver a verdict according to the interpretation reached, instead of repairing to get a nontrivial meaning.

Illustration Let us see how our algorithm works on the non-adverbial versions of sentences (14) and (15). The sentence (14) should first be interpreted as *Betty is strictly tall and strictly not tall* ($S(tall(a)) \wedge S(\neg tall(a))$) in virtue of (i), (ii) and (iii). That sentence is a contradiction, hence a trivial sentence. The next nontrivial interpretation we can obtain after backtracking is *Betty is tolerantly tall and tolerantly not tall* ($T(tall(a)) \wedge T(\neg tall(a))$) in virtue of (iii). Note that according to (iv), given the model we assumed, participants will either be lazy and judge $S(tall(a)) \wedge S(\neg tall(a))$ to have value 0 relative to their model of the situation, or they will have worked out the meaning of the sentence to be $T(tall(a)) \wedge T(\neg tall(a))$, and they will give it the value 1. The case of sentence (15) on the other hand involves no backtracking at all. The meaning we get for the sentence is *Betty is neither strictly tall nor strictly not tall* ($\neg S(tall(a)) \wedge \neg S(\neg tall(a))$), which is nontrivial and gets the value 1 in the model.

4.5 Assessment and comparisons

In order to establish whether our algorithm is plausible, we first need to compare it with extant algorithm, and then to see whether it makes further adequate predictions. To the best of our knowledge the closest kin to our algorithm is sketched in remarks made by [Alxatib and Pelletier, 2011a] about the computation of sub- and super-interpretation. Although Alxatib and Pelletier do not outline a general algorithm, they make some suggestive remarks, for instance concerning the meaning of double negations of gradable adjectives. For example, a sentence like:

(18) Betty is not not tall.

can be used to convey that Betty is not short, but also that she is not tall, in other words that she is borderline tall. Our algorithm can derive that meaning, since we basically get *Betty is not strictly not tall* ($\neg S(\neg tall(a))$) as the strengthened meaning.[13] This means exactly that Betty is tolerantly tall. As already mentioned, a globalist algorithm like that in [Cobreros et al., 2012] cannot derive that interpretation, since $\neg\neg tall(a)$ is necessarily equivalent to $tall(a)$ both under its strict and under its tolerant interpretation. The same holds of [Cobreros et al., 2015b]'s modified algorithm, which introduces no pragmatic difference between a sentence and its double negation. Our assumption that predicate negation and sentential negation should be treated differently plays a crucial role here.

Another problem that has been raised by [Alxatib et al., 2013] for the original globalist account in [Cobreros et al., 2012] concerns sentences such as:

(19) Betty is tall and Betty is not tall, or Betty is rich.

When interpreted strictly, the sentence says no more than that Betty is strictly rich. But intuitively, it in fact says that Betty is borderline tall and clearly rich, which is stronger. [Cobreros et al., 2015b] propose a more elaborate algorithm capable of deriving that meaning. Our algorithm also derives it straightforwardly: the incremental processing of (19) will treat the constituent subsentence "Betty is tall and Betty is not tall", but as illustrated earlier embedding the strict operator would yield a trivial constituent so backtracking reinterprets to mean that Betty is tolerantly tall and tolerantly not tall, and "Betty is rich" will be interpreted strictly, and their union will be nontrivial.

We note finally that none of the other algorithms of pragmatic meaning we cited would predict a preference for "x is neither P nor not P" over "x is P and not P". [Alxatib et al., 2013]'s algorithm based on fuzzy logic, for example, would predict (14) and (15) to be equally acceptable, like [Cobreros et al., 2012] or [Cobreros et al., 2015b]. The difference lies mostly in the fact that pragmatic strengthening in those cases is done globally, and not by the insertion of local operators.

5 Conclusions and Perspectives

We reported on two main findings in this paper. The first is a new confirmation of the fact that classical contradictions both of the "neither" type and of the "and" type are accepted and used by naive participants to describe borderline cases. This confirms that both gaps and gluts are operational in the representation of borderline cases (see [Egré et al., 2013] for a similar point). The second and more interesting finding is that for a representative class of relative gradable adjectives, "gappy" descriptions of the form "neither P nor not P" are preferred to "glutty" descriptions of the form "P and not P". The sense in which "neither"-descriptions are "gappy" is that they express that a particular case falls in the underlap between two strict extensions, and the sense in which "and"-descriptions are "glutty" is that they express that the same case falls in the overlap between two tolerant extensions. In agreement with the earlier accounts of the pragmatic meaning of vague predicates presented in [Alxatib and Pelletier, 2011a] and [Cobreros et al., 2012, Egré et al., 2013, Cobreros et al., 2015b], we have argued that we can account for that preference if indeed there is a bias toward selecting the strict meaning first, but moreover, and more centrally regarding the implications of our the theory, if pragmatic strengthening is done locally rather than globally.

Two main questions remain. The first is: does our algorithm predict the same preference for

[13] Note that here again, *strictly* would appear under *not* and typically trigger a short break in the prosodic contour.

"neither"-descriptions over "and"-descriptions for absolute adjectives like "empty"? Consequently, the second is: what are the *facts* for absolute adjectives? To answer the first question, we need a representation of absolute adjectives in our model. [Burnett, 2012, Burnett, 2014] argues that the strict denotation of "empty" coincides with the classical denotation of "empty" (that is, it should denote the zero degree of being filled on the relevant scale), though the tolerant denotation of "empty" can include recipients with a tiny bit of stuff in them. Symmetrically, Burnett takes the tolerant extension of "not empty" to be identical to the classical extension, but she assumes the strict extension to be a proper subset of the classical extension (hence "not empty", read strictly, does not mean "strictly speaking, not empty", but something like: "clearly empty", as Burnett stresses). The strict/tolerant duality that we observe for relative adjectives is therefore preserved for absolute adjectives: the strict extension of "not empty" is the complement of the tolerant extension of "empty" and conversely. As a result, and as for relative adjectives, the gap defined by the strict extensions of "empty" and "not empty" corresponds to the glut defined by the tolerant extensions of "empty" and "not empty". Using our algorithm, Burnett's account would therefore predict the same preference for a description of the form "neither empty nor not empty" over a description of the form "empty and not empty" when describing a glass with a tiny bit of liquid in it. Although both are pragmatically non trivial and true, the former should be preferred, by the principle of using strict meaning first.

While the second question calls for a new experiment, the intuitions expressed above in section 4.1 suggest to us that "and"-descriptions could possibly be preferred to "neither"-descriptions for absolute adjectives, contrary to the predictions just discussed. If these intuitions were to be confirmed by experimental data, one could question the interpretation of the second member of the "neither"-descriptions. It seems to us that a description of the form "neither empty nor not empty" has a natural interpretation paraphrasable as "neither empty nor, *strictly speaking*, not empty". Using our *strictly* operator, this reading amounts to "neither *strictly* empty nor not *strictly* empty," which is a manifest contradiction.

Our final remarks here are based entirely on conjectures and introspective data, and we need to confront them to actual data to make progress. We are in the process of running a separate study on absolute adjectives in order to gain further insights about such examples. Nevertheless, we think our results on relative adjectives already give us a compelling argument in favor of some form of local pragmatic strengthening in the computation of the meaning of sentences involving vague vocabulary.

Appendix A

Scenarios for human-oriented adjectives

Rich

A survey on wealth has been conducted in your country. In the population there are people with a very high degree of wealth, and people with a very low degree of wealth. Then there are people who lie in the middle between these two areas.

Imagine that Sam is one of the people in the middle range. Comparing Sam to other people in the population, is it true to say the following?

Tall

A survey on heights has been conducted in your country. In the population there are people of a very high height, and people of a very low height. Then there are people who lie in the middle between these two areas.

Imagine that Sam is one of the people in the middle range. Comparing Sam to other people in the population, is it true to say the following?

Old

A survey on age has been conducted in your country. In the population there are people whose age is very high, and people whose age is very low. Then there are people who lie in the middle between these two areas.

Imagine that Sam is one of the people in the middle range. Comparing Sam to other people in the population, is it true to say the following?

Heavy

A survey on weight has been conducted in your country. In the population, there are people of a very high weight, and people of a very low weight. Then there are people who lie in the middle between these two areas.

Imagine that Sam is one of the people in the middle range. Comparing Sam to other people in the population, is it true to say the following?

Scenarios for object-oriented adjectives

Fast

A survey on people's cars has been conducted in your country. In the population, there are people who own very high speed cars, and people who own very low speed cars. Then there are people who own cars that lie in the middle between these two areas.

Imagine that Sam is one of the people owning a car in the middle range. Comparing Sam's car to the cars of other people in the population, is it true to say the following?

Large

A survey on people's houses has been conducted in your country. In the population, there are people who own houses with a lot of space, and people who own houses with very little space. Then there are people who own houses that lie in the middle between these two areas.

Imagine that Sam is one of the people owning a house in the middle range. Comparing Sam's house to the houses of other people in the population, is it true to say the following?

Loud

A survey on people's voice has been conducted in your country. In the population, there are people whose voice has a very high intensity, and people whose voice has a very low intensity. Then there are people whose voice lie in the middle between these two areas.

Imagine that Sam is one of the people whose voice lie in the middle range. Comparing Sam's voice to the voices of other people in the population, is it true to say the following?

Wide

A survey on people's feet has been conducted in your country. In the population, there are people with a very high foot breadth, and people with a very low foot breadth. Then there are people whose foot breadth lie in the middle between these two areas.

Imagine that Sam is one of the people with a foot breadth in the middle range. Comparing Sam's feet to the feet of other people in the population, is it true to say the following?

Appendix B

Pre-questionnaire

Before proceeding to the actual experiment, please answer these simple questions.

There are 7 days in a week	❍ Yes	❍ No
Barack Obama is the current President of the USA	❍ Yes	❍ No
Abraham Lincoln was born in 2003	❍ Yes	❍ No
Marilyn Monroe died in 1780	❍ Yes	❍ No
Nicolas Sarkozy was one of the Presidents of the United States	❍ Yes	❍ No
California is part of the USA	❍ Yes	❍ No

Post-questionnaire

Please answer these few questions about the expriment. We are interested in what you actually remember at this point, so please do not reload the previous pages. Thank you.

Was there a scenario describing a population of pregnant women? ❍ Yes ❍ No

Were the descriptions that you saw in the first half of the experiment of a different form from those in the second half? ❍ Yes ❍ No

[Can you give an example of the type of descriptions in the first half?

|

Can you give an example of the type of descriptions in the second half?

|]

The last two questions and their input fields would appear only if the participants reported a difference between the descriptions in the first and in the second halves of the experiment.

Appendix C

As mentioned in footnote 8, we used the R package LME4 and its function *glmer* to fit logistic mixed-effect regression models on our data. The models had to predict each single judgment that we observed (*Yes* or *No*) given the **description** associated wit the judgment (*Control True*, *Control False*, *And*, *Neither*). To better predict the outcomes, some variability was taken into account: variability inherent to the **participant** (some participants were generally more charitable than others),

variability inherent to the **block** for each description (the *and* descriptions were less accepted when judged after the antonymic block, as pointed out in the *Results* section) and variability inherent to each **adjective** for each description (acceptance of *and* descriptions were generally boosted for object-oriented adjectives judged after human-oriented adjectives, as pointed out in the *Results* section). The formula in (20) shows how this was implemented in R. *Answer* has 2 levels: 0 corresponds to a *No* answer, 1 corresponds to a *Yes*. *IP* uniquely identifies each **participant**. ***Adjective*** has 8 levels, one for each single adjective in the set we draw. *Type* has 2 levels (between subjects): *first* indicates answers from the first **block** *second* indicates answers from the second block.

(20) $glmer((Answer == \text{“}1'') \sim Question + (1|IP) + (1 + Question|Adjective) + (1 + Question|Type), data = syntactic, family = binomial, control = glmerControl(optimizer = \text{“}bobyqa''))$

Following the recommendations of [Barr et al., 2013], we used the most complex models that converged. We also used the optimizer *bobyqa* to help the models converge.

References

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

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

[Alxatib and Pelletier, 2011b] Alxatib, S. and Pelletier, J. (2011b). On the psychology of truth-gaps. In Nouwen, R., van Rooij, R., Sauerland, U., and Schmitz, H.-C., editors, *Vagueness in communication*, pages 13–36. Springer, LNAI 6517.

[Barr et al., 2013] Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3):255 – 278.

[Bonini et al., 1999] Bonini, N., Osherson, D., Viale, R., and Williamson, T. (1999). On the psychology of vague predicates. *Mind & language*, 14(4):377–393.

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

[Burnett, 2014] Burnett, H. (2014). Penumbral connections in comparative constructions. *Journal of Applied Non-Classical Logics*, 24(1-2):35–60.

[Carlson, 1977] Carlson, G. N. (1977). *Reference to kinds in English*. PhD Dissertation, UMass / Amherst.

[Cobreros et al., 2015a] Cobreros, P., Egré, P., Ripley, D., and van Rooij, D. (2015a). Vagueness, truth and permissive consequence. In D. Achouriotti, H. Galinon, J. M., editor, *Unifying the Philosophy of Truth*, pages 409–430. Springer.

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

[Cobreros et al., 2015b] Cobreros, P., Egré, P., Ripley, D., and van Rooij, R. (2015b). Pragmatic interpretations of vague expressions: strongest meaning and nonmonotonic consequence. *Journal of Philosophical Logic*, 44(4):375–393.

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

[Fine, 1975] Fine, K. (1975). Vagueness, truth, and logic. *Synthese*, 30:265–300.

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

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

[Peirce, 1902] Peirce, C. S. (1902). Vague. In Baldwin, J. M., editor, *Dictionary of philosophy and psychology*, page 748. New York: Macmillan.

[Rett, 2007] Rett, J. (2007). Antonymy and evaluativity. In *Semantics and Linguistic Theory*, volume 17, pages 210–227.

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

[Ripley, 2013] Ripley, D. (2013). Sorting out the sorites. In Tanaka, K., Berto, F., and Mares, E., editors, *Paraconsistency: Logic and Applications*, pages 329–3. Springer.

[Roche, 2012] Roche, L. (2012). La négation des adjectifs. Master's thesis, under the supervision of B. Spector, ENS/EHESS/Paris V.

[Ruytenbeek, 2013] Ruytenbeek, N. (2013). An experimental approach of negated gradable adjectives. Master's thesis, under the supervision of B. Spector, ENS/EHESS/Paris V.

[Sæbø, 2009] Sæbø, K. J. (2009). Judgment ascriptions. *Linguistics and philosophy*, 32(4):327–352.

[Serchuk et al., 2011] Serchuk, P., Hargreaves, I., and Zach, R. (2011). Vagueness, logic and use: Four experimental studies on vagueness. *Mind & Language*, 26(5):540–573.

[Solt and Gotzner, 2010] Solt, S. and Gotzner, N. (2010). Expensive, not expensive, or cheap? Paper presented at the 11th Szklarska Poreba Workshop.

[Spranca et al., 1991] Spranca, M., Minsk, E., and Baron, J. (1991). Omission and commission in judgment and choice. *Journal of experimental social psychology*, 27(1):76–105.

[Unger, 1975] Unger, P. (1975). *Ignorance: A Case for Scepticism: A Case for Scepticism*. Oxford University Press, USA.

[Zehr, 2014] Zehr, J. (2014). *Vagueness, Presupposition and Truth-Value Judgments*. PHD Thesis, ENS, Paris.