



Alexandropoulou, Stavroula & Gotzner, Nicole. 2024. Gradable adjective interpretation under negation: The role of competition. *Glossa: a journal of general linguistics* 9(1). pp. 1–33. DOI: <https://doi.org/10.16995/glossa.9919>



## Gradable adjective interpretation under negation: The role of competition

**Stavroula Alexandropoulou**, University College London, Wakefield Street 2, WC1N 1PF London, UK, [stavroulaale@gmail.com](mailto:stavroulaale@gmail.com)

**Nicole Gotzner**, Osnabrück University, Wachsbleiche 27, 49090 Osnabrück, DE, [nicole.gotzner@googlemail.com](mailto:nicole.gotzner@googlemail.com)

Gradable adjectives have recently drawn a lot of attention in experimental semantics/pragmatics regarding the different interpretations they trigger depending on scale structure, polarity, extremeness, and presence of negation (e.g., Gotzner et al. 2018a; b; Leffel et al. 2019). The current study investigates the interpretation of relative and absolute gradable adjectives in the scope of negation (e.g., *not large* and *not clean*, respectively) capitalizing on the role of competition between alternative expressions in adjective interpretation. Our experimental results show that contextual competition between adjective expressions affects the interpretative asymmetry characteristic of positive and negative relative adjectives in the scope of negation (*not large* vs. *not small*; see Horn 1989). However, we do not find evidence that contextual competition affects the symmetric interpretation patterns of positive (*not clean*) and negative absolute adjectives (*not dirty*) under negation, or that potential relative-like interpretations of negated absolute adjectives hinge on the availability of overt contextual competition. The attested polarity asymmetry of negated relative adjectives is captured by Horn's pragmatic theory. The apparent availability of relative-like interpretations of negated absolute adjectives are hypothesized to result from a reasoning akin to Horn's division of pragmatic labor.

*Glossa: a journal of general linguistics* is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 OPEN ACCESS



## 1 Introduction

Gradable adjectives like *large* have mostly attracted semanticists' interest regarding their lexical-semantic properties while less consideration has been given to the impact of these lexical-semantic properties of gradable adjectives and the associated entailment patterns on their pragmatic interpretation (but see Doran et al. 2009; 2012; Gotzner et al. 2018a; b; Leffel et al. 2019; Alexandropoulou & Gotzner accepted). Speakers may use different alternative gradable expressions to communicate varying degrees to which a property holds of an entity and the choice of a given expression is often motivated pragmatically.

A phenomenon that has received considerable attention in pragmatics relates to the asymmetric interpretations of positive and negative gradable predicates in the scope of negation (see for example Horn 1989, and experimental evidence in Ruytenbeek et al. 2017; Gotzner & Mazzarella 2021). Negated positive adjectives, such as *not large*, are often strengthened to express the meaning of their simple antonym ('small'), a phenomenon known as negative strengthening. For instance, (1) may be uttered to convey that the apartment is small. Such a communicative strategy has been observed for cases where the implicated meaning is evaluatively negative (e.g., Brown & Levinson 1987; Horn 1989; 1993b; Terkourafi et al. 2020; Mazzarella & Gotzner 2021). On the other hand, the use of double negatives, such as *not small* in (2), does not seem to give rise to negative strengthening, as their relevant implicated meaning ('large') would be evaluatively positive (but see the discussion of *not bad* in Horn 1989 for a counter example).

(1) My apartment is not large  $\rightsquigarrow$  'My apartment is small'

(2) My apartment is not small  $\not\rightsquigarrow$  'My apartment is large'

Previous experimental work (Gotzner & Kiziltan 2022; Alexandropoulou & Gotzner accepted) has shown that this interpretative asymmetry of gradable adjectives in the scope of negation pertains to informationally weak relative adjectives only (e.g., *not large* vs. *not small*), and not to informationally strong relative adjectives (e.g., *not gigantic* vs. *not tiny*) or absolute gradable adjectives like *not clean* vs. *not dirty*.<sup>1</sup> However, although positive and negative absolute adjectives are expected to be interpreted symmetrically under negation in line with their entailment patterns (see (3) and (4)), some studies have shown that negated absolute adjectives might exhibit relative-like interpretation patterns. Relative-like interpretation patterns are those patterns whereby a negated absolute adjective is interpreted differently from the corresponding

---

<sup>1</sup> When not in the scope of negation, *gigantic* is informationally stronger than *large* because *My apartment is gigantic* entails *My apartment is large* while the reverse does not hold (Horn 1972). Likewise, *tiny* is informationally stronger than *small* as *My apartment is tiny* entails *My apartment is small* but not vice versa.

simple antonym (e.g., *not clean* ≠ ‘dirty’; see, e.g., Paradis & Willners 2006; Albu 2020) receiving a weaker interpretation than that of the simple antonym.

(3) The apartment is not clean  $\Rightarrow$  ‘The apartment is dirty’

(4) The apartment is not dirty  $\Rightarrow$  ‘The apartment is clean’

Such relative-like uses of negated absolute adjectives appear to be inconsistent with standard semantic accounts (such as Kennedy & McNally’s (2005)), which predict symmetric interpretative patterns for positive and negative absolute adjectives in the scope of negation.

Previous findings on the interpretation of negated adjectives may in part be explained by means of pragmatic accounts (Brown & Levinson 1987; Horn 1989; Blutner 2000; Krifka 2007) and by the competition between expressions as induced by presenting different adjective expressions concurrently in an experimental set-up (see Tessler & Franke 2018). The notion of competition among potentially equivalent expressions features in fact prominently in Horn’s (1989) and in complexity-based accounts (as in Krifka 2007) of negated antonymic adjectives: more complex expressions tend to receive weaker and/or more marked meanings compared to their equivalent simple expressions. On these accounts, equivalent expressions have been proposed to compete in terms of sociological considerations (cf. Horn 1989) and/or complexity-based considerations (Horn 1989; Krifka 2007). This has given rise to a theoretical controversy over the precise nature of the underlying interpretation mechanism of antonymic adjectives in the scope of negation. Alexandropoulou & Gotzner’s (accepted) findings of an asymmetric interpretation pattern for negated antonymic relative adjectives and of symmetric interpretation patterns for negated antonymic absolute adjectives successfully settle this controversy. Their results provide evidence in favor of an interpretation mechanism like that proposed by Horn (1989), governed primarily by sociological factors. They moreover show that the interpretation mechanism of negated adjectives requires the availability of a semantic extension gap between antonymic adjectives.

In Alexandropoulou & Gotzner’s (accepted) experiments, alternative adjective expressions were presented concurrently in the experimental set-up, presumably creating some form of competition. Since competition between alternative expressions is at the core of the aforementioned theoretical proposals on adjective interpretation under negation, in the present paper, we aim to investigate to what extent contextual competition between alternative expressions might have contributed to the attested interpretation patterns of negated relative and absolute adjectives. More specifically, we investigate whether the same distinct interpretations patterns obtain for negated relative and absolute adjectives even when there is no overt contextual competition. The specific question we ask is to what extent the absence of overt contextual competition affects the distinct interpretation patterns of negated antonymic relative and absolute gradable

adjectives differing in informational strength. We carry out two experiments where there is no overt competition between alternative adjective expressions, and only one adjective statement is presented at a time (single-statement presentation mode). We discuss the role of competition in adjective interpretation and derive the attested interpretation patterns of negated adjectives in Horn’s Neo-Gricean pragmatic framework.

This paper is structured as follows: The next section presents the relevant theoretical background and previous experimental findings that relate to our investigation. Section 3 presents two experiments we conducted on relative (Experiment 1) and absolute adjectives (Experiment 2), respectively, with a single-statement presentation mode, in order to obtain insights into the role of competition between expressions in adjective interpretation under negation. Section 4 discusses the relevance of the findings of our analyses to the role of competition in adjective interpretation and Section 5 concludes the paper.

## 2 Theoretical and experimental landscape

### 2.1 Semantics and pragmatics of gradable adjectives

Relative adjectives like *large* exhibit the hallmark of vagueness: they give rise to borderline cases. For example, it is felicitous to say *The apartment is neither small nor large*, especially in a situation where the size of the apartment does not clearly exceed the standard of comparison for apartments in a given context of evaluation. Absolute adjectives like *dirty*, on the other hand, do not typically give rise to borderline cases (but see Burnett 2014; Solt 2015 for counter examples). This is tightly linked to the entailment patterns of the different types of adjectives (see especially Kennedy 2007).

The negation of an absolute adjective entails its antonym (Cruse 1986; Rotstein & Winter 2004; Kennedy 2007), as in (5), repeating (3)–(4) from above. This is because “a minimal positive degree corresponds to a maximal negative degree” on the relevant measurement scale (Kennedy 2007: 27). The same does not hold for relative adjectives, which do not give rise to such an antonymy entailment relation, as shown in (6), (Kennedy 2007). Antonymic pairs of relative adjective terms, such as *large* and *small*, share the same dimension and degrees, imposing inverse ordering relations, and allow for a middle ground between their extensions (extension gap). This semantic extension gap arises as the relevant standards (standard of largeness and standard of smallness) may be different (Kennedy 2007). Thus, a negated relative adjective, e.g., *not large*, implicates any degree (interval) below the associated standard (of largeness), be it in the extension gap or in the extension of the antonym (*small*). Hence, *not large* may implicate but not entail its antonym *small*.

- (5)    a. The apartment is not clean  $\Rightarrow$  The apartment is dirty  
       b. The apartment is not dirty  $\Rightarrow$  The apartment is clean

- (6) a. The apartment is not large  $\nRightarrow$  The apartment is small  
 b. The apartment is not small  $\nRightarrow$  The apartment is large

Furthermore, as shown in (7), the entailment to the antonym does not hold for modified absolute adjectives (see (7a)) or stronger scale-mates (see (7b)) either; an observation that, to our knowledge, has not been clearly made so far. The sentences in (7) are in fact compatible with several different possibilities (the apartment being clean, dirty or filthy), possibly recreating a middle ground between positive and negative antonyms as in the case of relative adjectives.

- (7) a. The apartment is not very clean  $\nRightarrow$  The apartment is dirty  
 b. The apartment is not pristine  $\nRightarrow$  The apartment is filthy/dirty

Relatedly, Rotstein & Winter (2004) note that, in some contexts, modified absolute terms by *almost/slightly* break the complementarity of antonymic absolute terms (e.g., *dirty* vs. *clean*) manifested by their entailment patterns in (5). This is because, in some contexts, such expressions (like *almost dirty* according to Rotstein & Winter 2004) can be used to describe a middling situation like being ‘neither clean nor dirty’. One such context is when attention is drawn to the presence of a small amount of dirt, as in (8), and different adjective expressions are under consideration, creating a competition between forms and subtle distinctions in meaning.

- (8) It is certainly not clean, since it has some small spots on it, but it is not really dirty...  
 (from Rotstein & Winter 2004)

A well-known case where competition between forms leads to distinctions in meaning is the so-called “division of pragmatic labor” featuring in Horn’s Neo-Gricean pragmatic theory (Horn 1989). In the following, we present Horn’s theory, which comprehensively discusses and accounts for the interpretation of negated adjectives.

Horn proposes two opposed conversational principles that interlocutors observe during communication: The **Q** Principle (**Q** for Quantity): *Say as much as you can (modulo Quality and R)* and the **R** Principle (**R** for Relation): *Say no more than you must (modulo Q)*. Exploiting the **Q** Principle, the listener infers that anything beyond what is being said does not hold (*upper-bounding implicature*). Conversely, by exploiting the **R** Principle, the listener infers that what has been said is the least that holds (*lower-bounding implicature*, also known as *strengthening*). These two opposed pragmatic principles may interact dividing up the pragmatic labor, which has been established as “the division of pragmatic labor” (e.g., Horn 1993a):

“... given two co-extensive expressions, the more specialized form — briefer and/or more lexicalized — will tend to become **R**-associated with a particular unmarked, stereotypical meaning, use, or situation, while the use of the periphrastic or less lexicalized expression, typically (but not always) linguistically more complex or prolix, will tend to be **Q**-restricted

to those situations outside the stereotype, for which the unmarked expression could not have been used appropriately.” (Horn 1993a: 41)

Let us now consider how Horn’s theory can be applied to adjective interpretation. The Q principle is employed to capture upper-bounding inferences of expressions that differ in informational strength (also referred to as scalar implicatures) such as the adjectives *large* and *gigantic*, whereby the latter asymmetrically entails and is informationally stronger than the former (see also footnote 1). As (9) illustrates, for example, the apartment is large can Q-implicate that the apartment is large but not gigantic, hence, imposing an upper bound on the interpretation of *large*.

(9) The apartment is large  $\rightsquigarrow$  ‘The apartment is large but not gigantic’

Horn’s framework further captures inferences triggered by negated adjectives: e.g., the negative strengthening interpretation in (10) as well as the so-called middling interpretations that double negatives tend to trigger under negation, see (11).

(10) My apartment is not large.  $\rightsquigarrow$  ‘My apartment is small’ (*negative strengthening*)

(11) My apartment is not small.  $\rightsquigarrow$  ‘My apartment is neither large nor small’  
(*middling interpretation*)

More specifically, Horn (1989) derives negative strengthening (in (10)) as an R-based implicature: According to the hearer’s reasoning, the speaker chooses the more complex yet still less informative expression *not large* to conceal the stronger negative meaning expressed by the simple antonym *small*. Namely, given the R principle, the hearer reasons that what has been said (*not large*) is the least that holds true (lower bound on the interpretation of *not large*). The hearer in (10) thus strengthens the literal interpretation ‘less than large’ (that includes both medium-size and small degrees) to ‘small’. In turn, middling interpretations of double negatives, such as *not small* in (11), are derived via reasoning about an interplay between Horn’s Q and R principles:

“... double negation violates not only Grice’s Brevity maxim, i.e. the R-based principle of syntagmatic economy, but also the Q-based informativeness criterion: a double negative is longer and typically weaker than its simpler affirmative counterpart.” (Horn 1993a: 59)

In other words, the hearer in (11) reasons that the speaker utters a prolix expression such as a double negative (*not small*) because the corresponding brief and more informative simple antonym (*large*) is not precisely applicable. The hearer concludes that a weaker situation must hold instead, one that corresponds to the range of degrees in the extension gap between the two antonymic adjectives, arriving, thus, at the interpretation ‘neither large nor small’.

## 2.2 Previous experimental findings

Using a variety of experimental paradigms and tasks, previous experimental studies have attested an asymmetric interpretation for positive and negative relative adjectives in the scope of negation (see, e.g., Colston 1999; Fraenkel & Schul 2008; Ruytenbeek et al. 2017; Gotzner & Mazzarella 2021), such that positive adjectives (*not large*) are more likely to trigger negative strengthening ('rather small') than negative adjectives (*not small*) in the scope of negation.

As already mentioned, such a polarity asymmetry in adjective interpretation is not expected to arise for absolute adjectives in the scope of negation. Under standard semantic accounts of absolute adjectives (see, e.g., Kennedy 2007), negated absolute adjectives entail their simple antonym (*not clean*  $\Rightarrow$  'dirty', *not dirty*  $\Rightarrow$  'clean'; see also Section 2.1), hence, no extension gap between antonymic absolute adjectives, resulting in a symmetric interpretation pattern for negated positive (*not clean*) and negative terms (*not dirty*). However, Paradis & Willners (2006) report asymmetric interpretation patterns for negated positive and negative absolute adjectives, and importantly interpretations different and weaker than that of the corresponding simple antonym, namely relative-like interpretations (e.g., *not empty*  $\nRightarrow$  'full' and *not full*  $\nRightarrow$  'empty'; see also Fraenkel & Schul 2008 and Albu 2020). In this study, Paradis & Willners tested a larger set of antonymic adjective pairs, which besides absolute adjectives included also non-gradable adjectives such as *dead* vs. *alive*. This set of adjectives was subsumed by the category of contradictory antonyms, where one term must be true and the other one false of a given individual (e.g., one must be either dead or alive). Since this set of adjectives was rather heterogeneous, the type of adjective, namely, relative vs. absolute, has not been systematically controlled for in this study.

Alexandropoulou & Gotzner (accepted) set out to systematically investigate the interpretation of gradable adjectives in the scope of negation while incorporating the well-known semantic differences between absolute and relative adjectives. This study tested the competing predictions of Horn's (1989) and Krifka's (2007) accounts regarding the interpretation of negated relative and absolute adjectives in the scope of negation. Alexandropoulou & Gotzner (accepted) conducted one experiment on absolute adjectives and one on relative adjectives using a reward task (building on a study by Gotzner & Kiziltan (2022)). In this task, participants assigned ratings from a 5-point rating scale to co-occurring adjective predication statements differing in the presence or absence of negation, the informational strength (weak vs. strong) and the evaluative polarity of the adjectival predicate (positive vs. negative; see an example for each adjective type in **Tables 1** and **2**, respectively).

Alexandropoulou & Gotzner (accepted) found that informationally weak relative adjectives are interpreted asymmetrically with respect to evaluative polarity in the scope of negation (*not large* vs. *not small*): positive adjectives were partly interpreted similarly to their simple antonyms (*not large*  $\approx$  *small*), which was not the case for negative adjectives (*not small*). The

corresponding informationally strong antonymic terms (*not gigantic* and *not tiny*) exhibited a similar interpretation pattern to that of informationally weak negatives (*not small*; see Section 1.2 in Appendix for statistical analysis). Positive (*not clean*) and negative absolute adjectives (*not dirty*), on the other hand, presented symmetric interpretation patterns under negation (see Section 1.2 in Appendix for statistical analysis), presumably in line with their entailment patterns (see Section 2.1), with weak and strong conditions largely overlapping. Interestingly, negated absolute adjective conditions also received ratings evincing that these conditions were not strictly interpreted in the same way as the corresponding simple antonyms (cf. ratings of 3 in negated conditions vs. non-negated simple antonyms in **Figure 4**, left panel). Alexandropoulou & Gotzner propose that this could point to the apparent availability of relative-like interpretations for negated absolute adjectives, as in the case of Paradis & Willners (2006).

Alexandropoulou & Gotzner's (accepted) findings for weak antonymic relative and absolute adjectives confirm the predictions stemming from Horn's (1989) account of negated predicates. According to this account, negated weak relative (*not large* vs. *not small*) and absolute adjectives (*not clean* vs. *not dirty*) present distinct interpretation patterns, with the former exhibiting a polarity asymmetry. Furthermore, it is concluded that the existence of a semantic extension gap for the involved predicates is an essential property of the interpretation mechanism of negated predicates.

In this previous study, Alexandropoulou & Gotzner (accepted) used a concurrent presentation mode (multiple-statement presentation mode) in their task allowing for a competition and direct comparison among all different adjective statements. Crucially, this may have encouraged participants to draw more fine-grained (meaning) distinctions between competing adjective expressions. This is reminiscent of Horn's notion of competition as expressed by the division of pragmatic labor (i.e., that competition between a simple and a more complex form leads to distinctions in meaning).

The role of competition in meaning distinctions has also been tested experimentally via a manipulation of presentation mode in a study by Tessler & Franke (2019). Tessler & Franke (2019) showed that morphologically negative adjectives like *unhappy* are interpreted differently from the corresponding negated positive counterparts (*not happy*) in a multiple-statement presentation mode as opposed to a single-statement presentation mode, namely, when adjective statements are presented one at a time in a given context and there is no overt competition among different expressions. This essentially highlights the importance of overt contextual competition among alternative expressions in adjective interpretation, which may bring about meaning distinctions that are less likely to arise in the absence of such competition.

In light of Tessler & Franke's (2019) findings and Horn's notion of competition as expressed by the division of pragmatic labor, the current study sets out to investigate the influence of contextual competition among alternative adjective expressions on distinct interpretation patterns



of antonymic relative and absolute adjectives under negation. We will explore to what extent overt contextual competition might have contributed to Alexandropoulou & Gotzner's (accepted) attested distinct interpretation patterns of the two gradable adjective classes in the scope of negation. To that end, we ran experiments on relative (Experiment 1) and absolute gradable adjectives (Experiment 2) similar to those in Alexandropoulou & Gotzner accepted but with a single-statement presentation mode. Conducting these experiments will provide insight into: (i) whether the contextual competition between different adjective expressions can influence the distinct interpretation patterns of negated relative and absolute adjectives in general, and (ii) whether the apparent availability of relative-like interpretations of negated absolute adjectives, in particular, may hinge on contextual competition.

### **3 Current Study**

#### **3.1 Research objectives and questions**

We have seen that overt competition between a simple and a more complex form may bring about meaning distinctions between the two forms (cf. also Tessler & Franke's (2019) experimental findings). Drawing inspiration from Tessler & Franke (2019), the current study is concerned with the effect of contextual competition among alternative adjective expressions on adjective interpretation under negation. The specific research question we address is to what extent the absence of overt contextual competition affects the distinct interpretation patterns of negated antonymic relative and absolute adjectives differing in informational strength. The sub-questions we address in our study are the following: (i) Does the absence of overt contextual competition affect the symmetric and asymmetric interpretation patterns of antonymic absolute and relative adjectives under negation? We ask this question mainly for weak adjective terms, and also explore their interpretation in relation to strong adjective terms. (ii) Do apparent relative-like interpretations of negated absolute adjectives hinge on overt contextual competition?

We assume that overt competition is induced when multiple adjective statements are presented and interpreted concurrently (i.e., in a multiple-statement presentation mode), whereas no overt competition is available in a single-statement presentation mode, where one adjective statement is presented and judged upon at a time. Alexandropoulou & Gotzner's (accepted) experiments on relative and absolute adjectives employed a multiple-statement presentation mode. They reported distinct interpretation patterns for antonymic relative and absolute adjectives in the scope of negation. In the present study, we carried out the corresponding experiments with a single-statement presentation mode, to find insights into whether the induced overt contextual competition might have contributed to the attested interpretation patterns. In the following, we tackle (i) and (ii) in turn: we first investigate whether the same distinct interpretation patterns obtain in the absence of overt competition, and next whether the seemingly relative-like interpretations of negated absolute adjectives vanish in the absence of overt competition.

In both experiments, we primarily focus on weak adjectives. Regardless of competition, weak relative and absolute adjectives are predicted to have clearly distinct interpretations, unlike their strong counterparts. Such interpretation patterns were attested with overt competition between alternative expressions (Alexandropoulou & Gotzner accepted) and here we probe the extent to which they obtain even in the absence of overt competition. As regards strong adjective expressions, we will explore how those are interpreted in comparison to weak terms, bearing in mind that strong terms of relative and absolute adjective scales have similar interpretations. This is so, because antonymic pairs of strong relative (*gigantic* vs. *tiny*) and strong absolute adjectives (*pristine* vs. *filthy*) both represent contraries, namely, such antonymic terms might be false at the same time, but they cannot be true at the same time. We conducted Experiment 1 on relative adjectives and Experiment 2 on absolute adjectives.

## 3.2 Methods

In this section, we present the specifics of Experiments 1 and 2 together, as those were identical except for the type of adjectives tested (relative vs. absolute).

### 3.2.1 Participants

We recruited 240 participants in Experiment 1 (142 female, 95 male, 2 diverse, 1 that preferred not to say, mean age: 37.28, age range: 18–79) and an additional 240 in Experiment 2 (161 female, 76 male, 3 diverse, mean age: 36.75, age range: 19–81), on the Prolific platform.<sup>2</sup> Participants were residents of the United States and underwent screening for their native language. Their inclusion in the data analysis was contingent upon self-reporting English as their first native language. Following this criterion, we excluded the data of two participants in Experiment 1 (N = 238) and of three participants in Experiment 2 (N = 237), who all reported a native language other than English. Both experiments had a duration of up to 4 minutes and participants received compensation of 0.58 USD in Experiment 1 and 0.44 USD in Experiment 2.

### 3.2.2 Materials

The materials of Experiments 1 and 2 are adapted from Alexandropoulou & Gotzner (accepted). Specifically, Experiment 1 tested 8 quadruples of relative adjectives and Experiment 2 tested 8 quadruples of absolute adjectives, listed in Tables A and B, respectively, in Appendix. Each quadruple represented a scale and consisted of an informationally weak (*large/clean*) and an informationally strong positive term (*gigantic/pristine*)<sup>3</sup>—henceforth, weak and strong,

---

<sup>2</sup> <https://www.prolific.co>.

<sup>3</sup> The set of strong adjectives used in our experiments consists of what Morzycki (2012) has defined as extreme adjectives (both lexically and contextually extreme adjectives). Extreme adjectives like *gigantic* are argued not to share the same kind of scale as informationally weak adjectives (*large*). Weak adjectives involve salient degrees from a (contextually-provided) sub-part of the associated measurement scale, and strong/extreme adjectives involve degrees that lie off this contextually-defined sub-scale.

respectively—and the corresponding weak and strong negative antonyms in terms of evaluative polarity (*small/dirty* and *tiny/filthy*).<sup>4</sup> (See Section 1.2 in Appendix on the selection process and criteria for each adjective quadruple.)

Thus, we manipulated the factors: Scalar strength, Evaluative polarity, as well as Negation in a 2 (weak vs. strong)  $\times$  2 (evaluatively positive vs. evaluatively negative, henceforth, positive and negative, respectively)  $\times$  2 (non-negated vs. negated, henceforth  $\emptyset$  and negated, respectively) design, as each adjective of a given quadruple was embedded in a simple predication statement ( $x$  is ADJ), either with or without negation. Hence, we created 8 adjectival expressions for a given scale in both experiments, namely 8 conditions per adjective scale/item, resulting in 64 experimental trials in total (8 scales/items  $\times$  8 conditions). These trials were rotated through 8 lists in a (repeated-measures) Latin Square design, and every list had all 8 items, each in a separate condition (8 trials in total).

Experiments 1 and 2 made use of the same experimental task and set-up, and a single-statement presentation mode, which are illustrated in **Tables 1** and **2** with an example relative adjective item and an example absolute adjective item, respectively. **Tables 1** and **2** additionally illustrate the corresponding example items from Alexandropoulou & Gotzner’s (accepted) experiments using a multiple-statement presentation mode. As becomes obvious from the tables, in the current Experiments 1 and 2 and their Latin Square design, participants were presented with a single predication statement in a single trial containing an adjective expression in one of the 8 conditions of the relevant adjective scale, hence the name single-statement presentation mode. In Alexandropoulou & Gotzner’s (accepted) experiments, on the other hand, participants saw multiple statements ( $N = 8$ ) concurrently in a given trial (in all 8 conditions), hence the name multiple-statement presentation mode (cf. Tessler & Franke’s (2018) similar types of presentation mode). In the following, we present the task, set-up and presentation mode of Experiments 1 and 2 of the current study in juxtaposition to those of Alexandropoulou & Gotzner’s (accepted) experiments.

In both presentation modes, there was a unique context for each adjective quadruple/scale. In each context, participants had to give a rating for an adjectival predication statement (see Tessler & Franke 2018 for a similar paradigm). In every context, there was a speaker with complete knowledge (cf. Tim in **Table 1** and an examiner in **Table 2**). This speaker uttered a predication statement about a specific individual (person, object or activity) in the single-statement presentation mode, whereas in the multiple-statement presentation mode they uttered similar statements about a set of individuals ( $N = 8$ ). In both presentation modes, participants were instructed to specify the rating an individual (e.g., a room in **Table 1** and a hospital in **Table 2**) would receive concerning a specific aspect (e.g., size, hygiene standards) (refer to

---

<sup>4</sup> Evaluative polarity refers to the type of polarity defined based on judgements of desirability. In this context, the positive term, such as *large*, *clean*, is connotationally linked with a desirable property while the negative one, such as *small*, *dirty*, is associated with an undesirable property.

Alexandropoulou & Gotzner’s (accepted) supplementary materials for the prompts used for the individual contexts). While in the single-statement presentation mode participants had to rate a single individual, in the multiple-statement presentation mode they had to rate every individual out of eight referred to by the speaker.

Single-statement presentation mode
<p><b>Context:</b></p> <p>A group of friends goes on vacation. One friend named Tim writes a review for each person’s room on booking.com.</p> <p><i>Please decide which rating the room receives in terms of its size based on Tim’s statement.</i></p> <p><b>1 = tiny; 5 = gigantic</b></p> <p>Tim writes:</p> <p>Anna’s room was not tiny.</p> <p>1    2    3    4    5</p>
Multiple-statement presentation mode
<p><b>Context:</b></p> <p>A group of friends goes on vacation. One friend named Tim writes a review for each person’s room on booking.com.</p> <p><i>Please decide which rating the room receives in terms of its size based on Tim’s statement.</i></p> <p><b>1 = tiny; 5 = gigantic</b></p> <p>Tim writes:</p> <p>Anna’s room was not tiny.</p> <p>1    2    3    4    5</p> <p>David’s room was small.</p> <p>1    2    3    4    5</p> <p>Brian’s room was gigantic.</p> <p>1    2    3    4    5</p> <p>...</p>

**Table 1:** Example item *gigantic* from experiment on relative adjectives in single-statement presentation mode (Experiment 1) and in multiple-statement presentation mode (from Alexandropoulou & Gotzner accepted). In the single-statement presentation mode, the item appears in the negated negative strong condition (*not tiny*). In the multiple-statement presentation mode, three out of eight conditions of the item are presented. Although the rest are omitted here for presentation reasons, in the actual experiment participants saw all 8 adjective statements concurrently.

Single-statement presentation mode	
<b>Context:</b>	
	The government examines the hospitals of a big city for their hygiene standards. The examiner writes a review.
	<i>Please decide which rating the hospital gets for its hygiene standards based on the examiner's statement.</i>
	<b>1 = filthy; 5 = pristine</b>
	The examiner says:
	The Saint Anthony's Hospital is not filthy.
	1    2    3    4    5

Multiple-statement presentation mode	
<b>Context:</b>	
	The government examines the hospitals of a big city for their hygiene standards. The examiner writes a review.
	<i>Please decide which rating each hospital gets for its hygiene standards based on the examiner's statements.</i>
	<b>1 = filthy; 5 = pristine</b>
	The examiner says:
	The Saint Anthony's Hospital is not filthy.
	1    2    3    4    5
	The Saint Joseph Hospital is not dirty.
	1    2    3    4    5
	The Saint's Mary's Hospital is pristine.
	1    2    3    4    5
	...

**Table 2:** Example item *pristine* from experiment on absolute adjectives in single-statement presentation mode (Experiment 2) and in multiple-statement presentation mode (from Alexandropoulou & Gotzner accepted). In the single-statement presentation mode, the item appears in the negated negative strong condition (*not filthy*). In the multiple-statement presentation mode, three out of eight conditions of the item are presented. Although the rest are omitted here for presentation reasons, in the actual experiment participants saw all 8 adjective statements concurrently.

Participants' judgments were given on a 5-point rating scale (see Katsos & Bishop (2011) on use of gradient vs. binary response option) below the predication statement, where 1 stood for the negative strong adjective (e.g., *tiny/filthy*) and 5 for its positive counterpart (*gigantic/*

*pristine*).<sup>5</sup> Therefore, participants in this paradigm provided a rating for a specific individual (e.g., Anna's room in **Table 1**, and the Saint Anthony's Hospital in **Table 2**) they had read an evaluative statement about. In other words, this task offers a rating tool of a given individual satisfying the property expressed by the relevant adjectival predication given a specific context, and a specific response scale and endpoints. The end result achieved by employing this rating tool for all different adjective conditions is the ranking of the respective adjective expressions by virtue of the width of their communicated ranges. This is conjectured to reflect the meaning differences among the different adjective expressions. Crucially, notice that in the multiple-statement presentation mode there was a response scale below each of the 8 predication statements with the same endpoints and participants rated all 8 individuals within a specific context and technically given the same response scale and endpoints. This enables a direct comparison among the different adjective expressions with respect to a specific aspect/dimension and a direct ranking of the different expressions, facilitating meaning distinctions among them.

Lastly, this task capitalizes on the notion of evaluative polarity, where participants grade statements within a context in terms of a specific aspect every time, giving higher ratings to those statements predicating a desirable property and lower ratings to those predicating a less desirable property in the context.

In both Experiments 1 and 2 of the current study, every participant was randomly assigned to exactly one of the 8 lists and saw the 8 experimental trials of the list in a randomized order.

### 3.2.3 Procedure

Both experiments began with a couple demographic questions, the instructions, and a practice phase. After completing the demographic questions, participants read the instructions, which also provided an example illustrating the experimental task. In this example, there was a context and two predication statements involving an antonymic pair of adjectives that was not part of the experiments' selected sets of adjectives (*terrible* vs. *fantastic* in Experiment 1, and *false* vs. *true* in Experiment 2). The example said explicitly that the statement with the negative antonym should be assigned a rating of 1 and the statement with the positive antonym should be assigned a rating of 5 from the relevant response scales. Participants had to select these exact ratings in order to move on to the practice phase. If they selected a different rating from that explicitly mentioned in the example, they received feedback on which rating to assign to each of the two adjective statements. After selecting these ratings, participants were free to proceed to the practice phase.

---

<sup>5</sup> In a pilot study with multiple-statement presentation mode, we had 1 represent the worst rating and 5 the best rating. The resulting pilot patterns were the same as in a pilot version where participants were explicitly given, e.g., the information that 1 = *tiny* and 5 = *gigantic*. Since there was no difference between the two, we decided to go with the latter version in all subsequent experiments.

The practice phase included two additional contexts/items, each featuring two statements. These statements involved a pair of antonymic adjectives that was not part of the experiments' selected set of adjectives. These two practice items were similar to the example item; however there was no explicit instruction given on how to rate each adjective statement. Only upon selecting the rating 1 for the negative antonym and the rating 5 for the positive antonym in both practice items, were participants free to advance to the main experiment. If they selected a different rating in either case, they received feedback and were instructed to choose the appropriate rating. The introduction of the practice phase made sure that participants understood the experimental task and familiarized themselves with using the 1–5 response scale. Finally, both experiments were programmed using PCIBex (Zehr & Schwarz 2018).

### 3.3 Results

#### 3.3.1 Overview of analyses

We performed an analysis for each adjective type separately in order to tackle our first sub-question, namely whether the absence of overt contextual competition affects the symmetric and asymmetric interpretation patterns of negated antonymic absolute and relative adjectives differing in informational strength. We have three manipulations: (i) Informational/Scalar Strength (weak vs. strong), (ii) Evaluative Polarity (positive vs. negative), (iii) Negation (non-negated ( $\emptyset$ ) vs. negated).

In both analyses, we would like to test whether the attested distinct interpretation patterns of negated antonymic relative and absolute adjectives in terms of evaluative polarity also obtain in a single-statement presentation mode. To that end, for each adjective type, we built a cumulative link mixed-effects model using the ordinal R package (Christensen 2019). We included the treatment-coded fixed effects of Negation (non-negated vs. negated; with negated as the reference level) and Scalar Strength (weak vs. strong; with weak as the reference level), and the sum-coded fixed effect of Evaluative polarity (positive vs. negative). We also included the maximal converging random-effect structure justified by the experimental design (including random by-participant and by-item intercepts and slopes).

In such a model, the effect of Negation basically indicates whether positive (*not large/not clean*) and negative weak adjectives (*not small/not dirty*) are interpreted symmetrically or not under negation in comparison to their base forms without negation. A significant effect of Negation indicates an asymmetric interpretation of positive and negative weak adjectives while a non-significant effect of Negation suggests that there is no evidence of such an asymmetric interpretation under negation. Hence, this effect in the model with a contrast-coding scheme is of primary importance for the interpretation patterns we set out to investigate.

Next, in a third analysis, we tackle our second sub-question, namely, whether the apparent availability of relative-like interpretations of negated weak absolute adjectives (*not clean/not dirty*) obtains even when there is no overt contextual competition. In other words, whether a negated weak absolute adjective (*not dirty/not clean*) is judged to be different and weaker than the corresponding simple antonyms (*clean/dirty*). In order to address our second sub-question, we built a cumulative link mixed-effects model with the factor Antonym type as a fixed effect with four levels: negative weak (*dirty*), negated positive weak (*not clean*), negated negative weak (*not dirty*), positive weak (*clean*). We also included the maximal converging random-effect structure justified by the experimental design.

Antonym type was treatment-coded with negative weak as the reference level in order to get at the critical comparison of negative weak (*dirty*) vs. negated positive weak (*not clean*). The reference level of Antonym type was recoded (with positive weak as the reference level) in order to get at the second critical comparison of negated negative weak (*not dirty*) vs. positive weak (*clean*). We ran this analysis on weak absolute adjectives only. Recall that antonymic pairs of strong absolute adjectives like *pristine* and *filthy* constitute contraries. The semantic interpretation of contrary antonyms leaves an extension gap between the positive and the negative term that is. By virtue of that, it may not be entirely unexpected if strong absolute terms are found to trigger interpretations weaker than their corresponding antonym. By contrast, the semantic interpretation of weak absolute adjectives does not leave an extension gap between the positive and negative term under standard semantic accounts, hence no weak, relative-like interpretation is expected to arise for those.

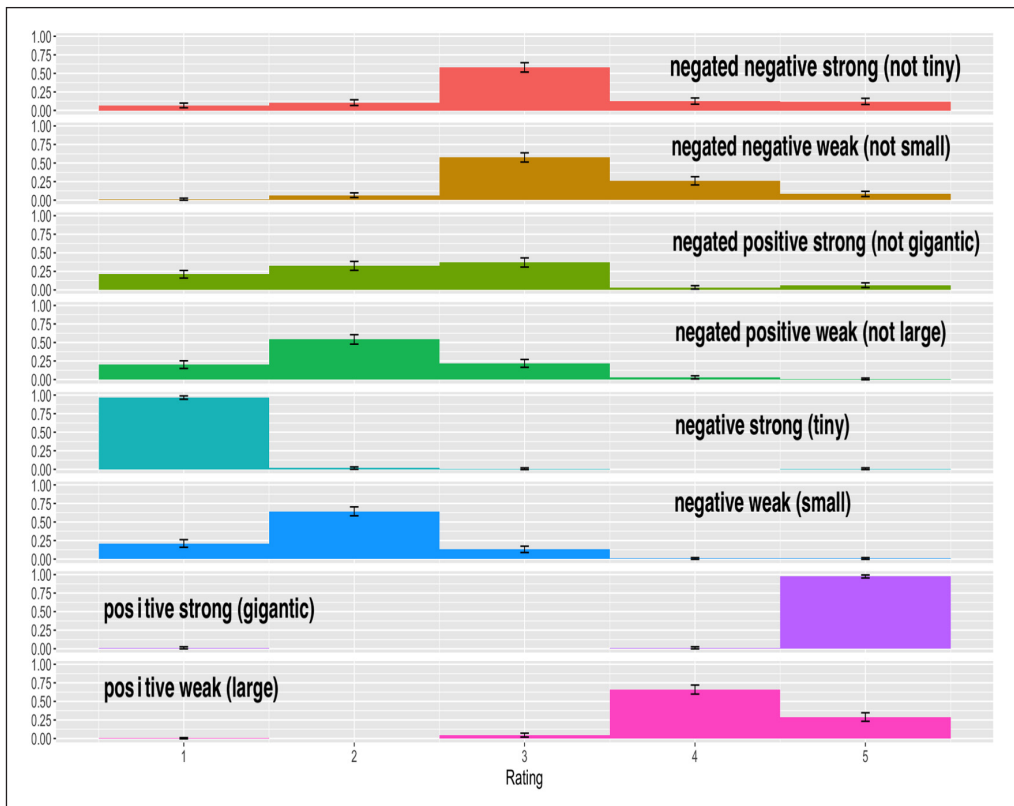
In the next three sections (Sections 3.3.2, 3.3.3, and 3.3.4), we present the output of the three analyses detailed above.

### 3.3.2 Results of Experiment 1

For the single-statement presentation experiment, we applied the same exclusion criterion on participants' data as that applied by Alexandropoulou & Gotzner (accepted): Should a participant placed the two strong adjectives (without negation) of the list at opposite ends of the 1–5 response scale (for instance, assigning a rating of 1 to *gigantic*, or of 5 to *tiny*, contrary to the specified scale endpoints in the Context; cf. also **Table 1**), all data from this participant were removed from subsequent analyses. No data were removed on the basis of this criterion in the single-statement presentation experiment on relative adjectives.

A visual inspection of the response patterns in **Figure 1** reveals an asymmetry between the two negated weak conditions (*not large* vs. *not small*), with positive terms (*not large*) receiving a high proportion of ratings of 2 – rather similarly to the corresponding simple antonymic condition (*small*) – and negative ones (*not small*) mainly receiving ratings of 3. This was generally observed across different contexts/items (see Figure D in Appendix).





**Figure 1:** Proportions per rating per relative adjective condition in Experiment 1 of current study. Error bars represent 95% Confidence Intervals.

The statistical analysis, detailed in **Table 3**, revealed a highly significant simple effect of Negation ( $\beta = 0.89$ ,  $SE = 0.27$ ,  $z = 3.30$ ,  $p < .0001$ ) such that the difference of the positive weak condition minus the negated positive weak condition (*large-not large*) does not cancel out the difference between negated negative weak and negative weak conditions (*not small-small*), as the former is smaller than the latter.<sup>6</sup> Hence, it follows that there is a polarity asymmetry in the interpretation of weak antonymic relative terms in the scope of negation as reported by Alexandropoulou & Gotzner (accepted) for the multiple-statement presentation mode. Additionally, the interaction of Negation\*ScalarStrength did not reach significance ( $p = 0.61$ ), thus there is no evidence that the aforesaid interpretation pattern is different between weak and strong antonymic adjectives. This may suggest the presence of a polarity asymmetry in the case of negated strong adjectives, too.

<sup>6</sup> Since we treat the obtained data as ordinal and analyse them with cumulative link models, we shall clarify that the term “difference” concerns differences between distributions of ratings.

	Estimate	SE	z-value	p-value
Negationnon-negated	0.8915	0.2699	3.303	0.000957***
ScalarStrengthstrong	0.1050	0.1663	0.632	0.527662
Polarity1	-1.5384	0.1837	-8.373	< 2e-16***
Negationnon-negated:ScalarStrengthstrong	-0.2252	0.4405	-0.511	0.609125
Negationnon-negated:Polarity1	4.3110	0.2332	18.490	< 2e-16 ***
ScalarStrengthstrong:Polarity1	0.5794	0.1397	4.147	3.37e-05 ***
Negationnon-negated:ScalarStrengthstrong:Polarity1	3.9844	0.4326	9.211	< 2e-16 ***

**Table 3:** Experiment 1 (relative adjectives) with single-statement presentation mode: Output of cumulative link model. `clmm(Rating ~ Negation * ScalarStrength * Polarity + (Negation + ScalarStrength + Polarity | Participant) + (Negation * ScalarStrength * Polarity | Item), data = data)`.

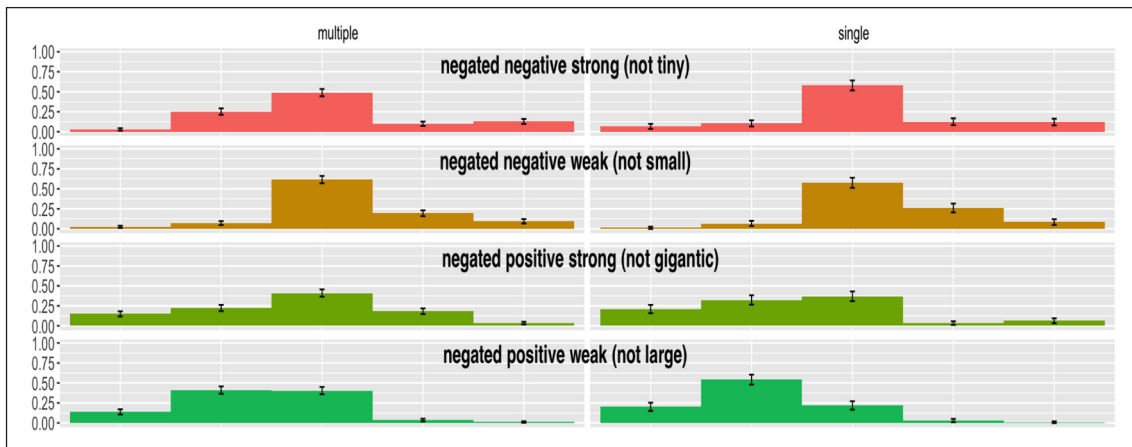
If one compares the two presentation modes (**Figure 2**), it becomes evident that the asymmetry of negated weak antonymic terms is more pronounced in the single-statement presentation mode. An exploratory cross-experiment analysis we conducted with Presentation mode as an additional treatment-coded fixed effect (multiple vs. single; with multiple as the reference level; see **Table 5**) confirmed this observation.<sup>7</sup>

The interaction of Negation\*Presentationmode was significant ( $\beta = 0.34$ ,  $SE = 0.15$ ,  $z = 2.23$ ,  $p < .05$ ), with the effect of Negation being larger in single- than in multiple-statement presentation mode. The interaction of Negation\*ScalarStrength\*Presentationmode was not significant ( $p = 0.592$ ).

We conclude that the interpretation asymmetry of positive and negative weak relative adjectives in the scope of negation still obtains in the absence of overt competition. Moreover, there are indications that Presentation mode affects this interpretation asymmetry (cf. significant

---

<sup>7</sup> The results of such a cross-experiment analysis should be taken with caution, as the two experiments under comparison differ in a couple other aspects besides presentation mode: Firstly, the dataset of the multiple-statement presentation mode is larger than that of the single-statement presentation mode (8 ratings per condition per participant vs. 1 rating per condition per participant). Secondly, the distribution of the ratings per item in the 8 conditions represents within-participants variation in the multiple-statement presentation mode and between-participants variation in the single-statement presentation mode. We thank an anonymous reviewer for raising this issue. We further acknowledge that an alternative approach for conducting a cross-experiment analysis would involve using a subset of the data from the multiple-statement presentation experiment. However, such an analysis is not possible given the randomized trial structure of our design.



**Figure 2:** Proportions per rating per negated relative adjective condition in Alexandropoulou & Gotzner’s (accepted) experiment on relative adjectives (left panel: multiple) and in Experiment 1 of current study (right panel: single). Error bars represent 95% Confidence Intervals.

Negation\*Presentationmode): negated positive weak terms (*not large*) are interpreted more similarly to the corresponding simple antonyms (*small*) in the single- than in the multiple-statement presentation experiment, contrary to negated negative weak terms (see also **Figure 1** above and Figure A in Appendix). Thus, it appears that the interpretation asymmetry may be more strongly available when there is no overt contextual competition between alternative adjective expressions.

### 3.3.3 Results of Experiment 2

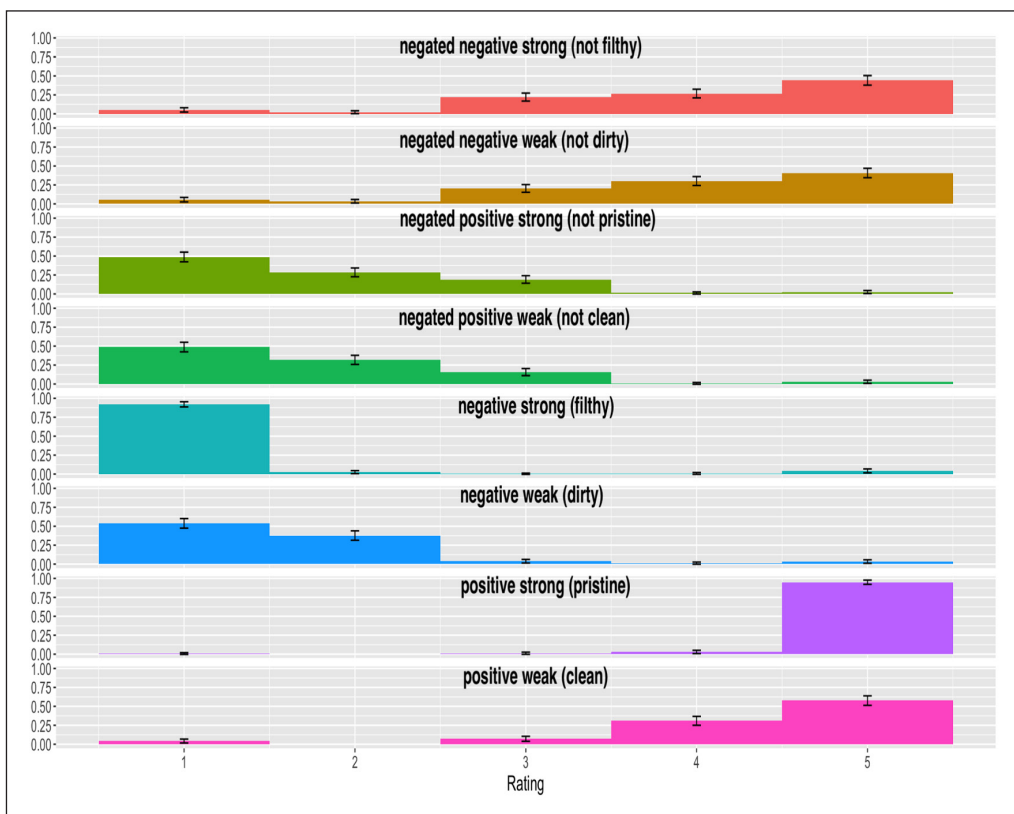
We removed two participants after applying the same participant exclusion criterion as in Experiment 1 (remaining participants’  $N = 235$ ).

One can observe in **Figure 3** that weak positive (*not clean*) and negative terms (*not dirty*) exhibit symmetric response patterns under negation. This is also observed in most of the individual items tested (see Figure F in Appendix). Moreover, although the response patterns in negated weak absolute conditions (*not clean/not dirty*) and those of the corresponding simple antonyms (*dirty/clean*) appear to be rather similar, there still is some amount of mid-scale ratings in negated conditions. Last, negated strong absolute terms (*not pristine/not filthy*) display response patterns that largely overlap with those of weak terms (*not clean/not dirty*).

The statistical analysis (see **Table 4**) revealed a marginal simple effect of Negation ( $p = 0.071$ ), hence there is no robust evidence of an asymmetric interpretation pattern of positive (*not clean*) and negative weak absolute adjectives (*not dirty*) in the scope of negation in the

absence of overt competition. Moreover, the interaction of Negation\*ScalarStrength did not reach significance either ( $p = 0.61$ ). Thus, there is no evidence that weak and strong antonymic adjectives are interpreted differently in the scope of negation.

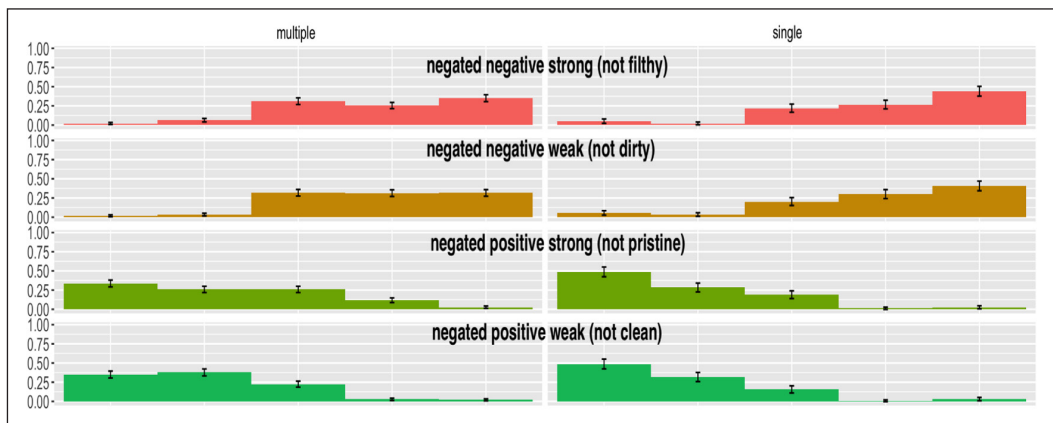
In addition, comparing these data with Alexandropoulou & Gotzner's (accepted) multiple-statement presentation mode data (see **Figure 4**), it is noticeable that there are relatively fewer mid-scale ratings in the single- than in the multiple-statement presentation mode. However, an exploratory cross-experiment analysis we conducted, with Presentation mode as an additional treatment-coded fixed effect (with multiple as the reference level; see **Table 6**), revealed no significant interaction effect of Negation\*Presentationmode ( $p = 0.37$ ) or of Negation\*ScalarStrength\*Presentationmode ( $p = 0.74$ ). Thus, there are no indications that Presentation mode affects the interpretation patterns of weak antonymic absolute adjectives in the scope of negation or their interpretation relative to their strong counterparts. Symmetric interpretation patterns of antonymic absolute adjectives do not seem to hinge on overt contextual competition.



**Figure 3:** Proportions per rating per absolute adjective condition in Experiment 2 of current study. Error bars represent 95% Confidence Intervals.

	Estimate	SE	z-value	p-value
Negationnon-negated	0.2684	0.1486	1.806	0.0709.
ScalarStrengthstrong	0.1460	0.2031	0.719	0.4722
Polarity1	-1.8768	0.1383	-13.566	< 2e-16 ***
Negationnon-negated: ScalarStrengthstrong	0.1623	0.3193	0.508	0.6111
Negationnon-negated:Polarity1	4.2956	0.1916	22.420	< 2e-16 ***
ScalarStrengthstrong:Polarity1	-0.1025	0.2379	-0.431	0.6665
Negationnon-negated: ScalarStrengthstrong:Polarity1	2.7077	0.3247	8.338	< 2e-16 ***

**Table 4:** Experiment 2 (absolute adjectives) with single-statement presentation mode: Output of cumulative link model. clmm (Rating ~ Negation \* ScalarStrength \* Polarity + (Negation + ScalarStrength \* Polarity | Participant) + (Negation \* ScalarStrength \* Polarity | Item), data = data).



**Figure 4:** Proportions per rating per negated absolute adjective condition in Alexandropoulou & Gotzner's (accepted) experiment on absolute adjectives (left panel: multiple) and in Experiment 2 of current study (right panel: single). Error bars represent 95% Confidence Intervals.

### 3.3.4 Results: Apparent relative-like interpretations of absolute adjectives

In this analysis, we will examine whether there is still a difference in the interpretation patterns between negated antonymic weak absolute adjectives (*not dirty/not clean*) and their truth-conditionally equivalent simple weak adjectives (*clean/dirty*) when there is no overt contextual competition. This will address the second sub-question of the present study.

	Estimate	SE	z-value	p-value
Negationnon-negated	0.57771	0.22481	2.570	0.01018 *
ScalarStrengthstrong	0.13780	0.09461	1.456	0.14529
Polarity1	-1.15757	0.15334	-7.549	4.39e-14 ***
Presentationmodesingle	-0.27953	0.11171	-2.502	0.01234 *
Negationnon-negated: ScalarStrengthstrong	-0.45321	0.25547	-1.774	0.07606 .
Negationnon-negated:Polarity1	3.66915	0.16845	21.782	< 2e-16 ***
Scalarstrengthstrong:Polarity1	0.74821	0.09570	7.819	5.34e-15 ***
Negationnon-negated: Presentationmodesingle	0.34018	0.15253	2.230	0.02573 *
ScalarStrengthstrong: Presentationmodesingle	-0.02523	0.15475	-0.163	0.87050
Polarity1:Presentationmodesingle	-0.35966	0.13757	-2.614	0.00894 **
Negationnon-negated: ScalarStrengthstrong:Polarity1	3.30333	0.23085	14.309	< 2e-16 ***
Negationnon-negated: ScalarStrengthstrong: Presentationmodesingle	0.19626	0.36650	0.535	0.59231
Negationnon-negated: Polarity1:Presentationmodesingle	0.58716	0.14817	3.963	7.41e-05 ***
ScalarStrengthstrong: Polarity1:Presentationmodesingle	-0.16588	0.15316	-1.083	0.27879
Negationnon-negated:Scal- arStrengthstrong:Polarity1: Presentationmodesingle	0.60015	0.36621	1.639	0.10125

**Table 5:** Cross-experiment analysis (relative adjectives): Output of cumulative link model `clmm(Rating ~ Negation * ScalarStrength * Polarity * Presentationmode + (Negation + ScalarStrength * Polarity | Participant) + (Negation * ScalarStrength * Polarity | Item), data = data)`.

The analysis revealed that there is a statistically significant difference between the negated negative weak adjectives (*not dirty*) and the truth-conditionally equivalent positive weak adjectives (*clean*) in the single-statement presentation mode ( $\beta = 0.94$ ,  $SE = 0.21$ ,  $z = 4.45$ ,  $p < 0.0001$ ; see **Table 7**). The difference between the other pair of truth-conditionally equivalent adjective expressions, viz., negated positive weak adjectives (*not clean*) and negative weak adjectives (*dirty*), was also significant ( $\beta = -0.44$ ,  $SE = 0.21$ ,  $z = -2.09$ ,  $p < 0.05$ ). We also built cumulative link mixed-effects models with Presentation mode as an additional treatment-coded fixed effect (with single as the reference level). Presentation mode did not interact significantly with either of the aforesaid critical comparisons (respective  $p$ 's = 0.946 and 0.124).

	Estimate	SE	z-value	p-value
Negationnon-negated	0.157713	0.109622	1.439	0.15023
ScalarStrengthstrong	0.196014	0.154730	1.267	0.20522
Polarity1	-1.854414	0.143614	-12.912	< 2e-16 ***
Presentationmodesingle	-0.163429	0.109654	-1.490	0.13612
Negationnon-negated: ScalarStrengthstrong	0.221104	0.293550	0.753	0.45133
Negationnon-negated:Polarity1	4.433570	0.186421	23.783	< 2e-16 ***
ScalarStrengthstrong:Polarity1	0.009189	0.278631	0.033	0.97369
Negationnon-negated: Presentationmodesingle	0.139019	0.155519	0.894	0.37137
ScalarStrengthstrong: Presentationmodesingle	-0.104241	0.159191	-0.655	0.51258
Polarity1:Presentationmodesingle	-0.479241	0.110995	-4.318	1.58e-05 ***
Negationnon-negated: ScalarStrengthstrong:Polarity1	4.292782	0.265051	16.196	< 2e-16 ***
Negationnon-negated: ScalarStrengthstrong: Presentationmodesingle	0.112287	0.335688	0.334	0.73800
Negationnon-negated:Polarity1: Presentationmodesingle	0.826399	0.154150	5.361	8.28e-08 ***
ScalarStrengthstrong:Polarity1: Presentationmodesingle	-0.181755	0.244072	-0.745	0.45647
Negationnon-negated:Scal- arStrengthstrong:Polarity1: Presentationmodesingle	-1.035654	0.351130	-2.949	0.00318 **

**Table 6:** Cross-experiment analysis (absolute adjectives): Output of cumulative link model  $\text{clmm}(\text{Rating} \sim \text{Negation} * \text{ScalarStrength} * \text{Polarity} * \text{Presentationmode} + (\text{Negation} + \text{ScalarStrength} * \text{Polarity} | \text{Participant}) + (\text{Negation} * \text{ScalarStrength} * \text{Polarity} | \text{Item}), \text{data} = \text{data})$ .

	Estimate	SE	z-value	p-value
Negated negative weak vs. simple positive weak	0.9346	0.2099	4.453	8.46e-06 ***
Negated positive weak vs. simple negative weak	-0.4368	0.2095	-2.085	0.037 *

**Table 7:** Analysis of weak absolute adjectives in single-statement presentation mode: Cumulative link model  $\text{clmm}(\text{Rating} \sim \text{AntonymType} + (\text{AntonymType} | \text{Participant}) + (\text{Item}), \text{data} = \text{data})$ .

Our findings suggest that there is evidence of a different interpretation of negated weak adjectives (*not dirty/not clean*) and corresponding simple antonyms (*clean/dirty*) in single-statement presentation mode. An exploratory cross-experiment investigation showed that these differences may not be affected differentially by Presentation mode. Taken together, our results suggest that the apparent availability of relative-like interpretations of negated weak absolute adjectives (*not clean/not dirty*) does not seem to hinge on Presentation mode, as manipulated and implemented in the present study.

## 4 General discussion

### 4.1 The role of competition in adjective interpretation

In this paper, we tackled the role of competition between alternative expressions in adjective interpretation, zooming in on the interpretation of adjectives under negation in relation to corresponding simple antonyms (e.g., *not small* vs. *large*). Such competition occupies a vital position in Horn's (1989) and in Krifka's (2007) accounts, where complex forms (e.g., *not small*) are assigned a different interpretation from the corresponding simple form (*large*) on the basis of complexity-based considerations. Specifically, we were concerned with the contextually overt competition among alternative adjective expressions and investigated to what extent the absence of such competition affects the distinct interpretation patterns of relative and absolute adjectives in the scope of negation.

We addressed two sub-questions: (i) Does the absence of overt contextual competition affect the symmetric and asymmetric interpretation patterns of antonymic absolute and relative adjectives differing in informational strength in the scope of negation? (ii) Do apparent relative-like interpretations of negated absolute adjectives hinge exclusively on overt contextual competition? We assumed that overt contextual competition is induced by a multiple-statement presentation mode, as conducted by Alexandropoulou & Gotzner (accepted), whereas the absence of an overt contextual competition is achieved via a single-statement presentation mode. To address our two sub-questions, we conducted two experiments with a single-statement presentation mode: one on relative and one on absolute adjectives.

Our analysis showed that the asymmetric interpretation patterns of antonymic weak relative adjectives (*not large* vs. *not small*) attested by Alexandropoulou & Gotzner (accepted) also obtain when there is no overt competition. A further exploratory analysis suggested that the relevant polarity asymmetry is stronger in the single-statement presentation mode than in the multiple-statement presentation mode, with positive relative terms (*not large*) being interpreted more similarly to the corresponding simple antonym (*small*) in the former than in the latter version while negated negative terms (*not small*) behaved consistently across versions. From this, we may conclude that contextual competition does affect the asymmetric interpretation



of negated antonymic weak relative adjectives. In particular, when there is no overt contextual competition, positive relative terms (*not large*) tend to be negatively strengthened more and interpreted similarly to the corresponding simple antonym (*small*). Hence, distinctions between these two adjective forms are less likely to occur in the absence of overt contextual competition, as predicted. It was further shown that antonymic strong relative adjectives are not interpreted differently than antonymic weak relative adjectives, and this comparison was not modulated by Presentation mode.

Our analysis for absolute adjectives revealed no robust evidence of an asymmetric interpretation pattern of positive (*not clean*) and negative weak absolute adjectives (*not dirty*) in the scope of negation when there is no overt contextual competition. A further exploratory analysis suggested that the apparent symmetric response patterns of antonymic weak absolute adjectives were not affected differentially by Presentation mode, namely by the presence or absence of overt competition. This was also the case for the symmetric response patterns of negated strong positive (*not pristine*) and negative absolute terms (*not filthy*) in relation to their weak counterparts.

In all, in the absence of overt competition, weak antonymic relative adjectives are interpreted more asymmetrically in the scope of negation while the symmetric interpretation patterns of weak antonymic absolute adjectives remain unaffected. Now, let us critically evaluate the connection between our findings and the fundamental assumption that competition among equivalent expressions leads to more distinctions in meaning (complex forms receive a different interpretation from corresponding simple antonyms (Horn 1989; Krifka 2007)). Given this assumption, we can explain why in the absence of contextual competition participants choose an interpretation of negated positive relative adjectives similar to that of their simple antonyms, resulting in a stronger polarity asymmetry. In particular, such a setting possibly leaves more space for a reasoning that does not appeal to complexity-based considerations encouraged by the simultaneous presence of competing expressions. That is, a negative strengthening reasoning (implication to the antonym) as outlined in Horn's (1989) account is more likely to emerge. On the other hand, it is suggested that the reasoning that distinguishes *not small* from the corresponding simple form *large* appears to be independent of the availability of the competing alternative expression in the particular communicative setting. A complexity-based reasoning may thus be initiated in this case regardless of the presence of competing alternative expressions in the specific communicative setting. We believe that such a reasoning is instead dependent on the existence of an extension gap between the semantic interpretations of antonymic relative adjectives – a core property of relative adjectives. This reasoning, along with the aforementioned negative strengthening one, will be further elucidated in the following section, drawing on Horn's theory.

In order to tackle our second sub-question, relating to the apparent availability of relative-like interpretations of negated absolute adjectives, we performed an analysis for weak absolute adjectives only, with Antonym type being the only predictor (treatment-coded fixed effect). This analysis showed a reliable difference between negated absolute adjectives (*not dirty/not clean*) and corresponding simple antonyms (*clean/dirty*). Hence, relative-like interpretations of negated absolute adjectives seem to arise in the absence of overt contextual competition, too. Interestingly, an additional exploratory analysis with Presentation mode as a second treatment-coded fixed effect revealed no interaction of Presentation mode with either of the two comparisons (*not dirty* vs. *clean*, and *not clean* vs. *dirty*).

When considering all the results for absolute adjectives, the observed symmetric interpretation patterns align with standard semantic accounts (e.g., Kennedy 2007), regardless of the presentational/communicative context. However, fewer meaning distinctions between truth-conditionally equivalent expressions (e.g., *not clean* vs. *dirty*) might have been expected in the absence of overt competition. This anticipation is rooted in the aforesaid assumption that more complex forms (*not clean*) obtain a different interpretation from the competing truth-conditionally equivalent simple forms (*dirty*), resulting in relative-like interpretations for the former. This contrast should be less likely to arise in the absence of overt competition compared to when both competing expressions are present in a specific communicative setting. We strongly believe that there are a couple reasons why we do not find evidence that the apparent availability of relative-like interpretations is exclusively dependent on overt contextual competition as induced by the specific experimental set-up. First, it might be the case that the presence of strong antonymic adjectives (*filthy, pristine*) as indicated scale endpoints in the context of every item (across Presentation modes, see **Table 2**) might evoke different related adjective forms. This may create some competition between different forms, which may in turn encourage drawing finer distinctions in meaning, like those between *not dirty* and *clean*, and between *not clean* and *dirty*. Second, the use of a 5-point rating scale as a response scale in both Presentation modes may further encourage such distinctions (cf. Katsos & Bishop's (2011) findings on graded rating scales bringing out meaning distinctions, unlike binary judgment tasks).<sup>8</sup> In the next section, we claim that a complexity-based reasoning may be at play in the case of negated absolute adjectives similarly to negated negative relatives (*not small*). However, we will argue that this reasoning is prompted by overt or indirect contextual competition, makes reference to the level of precision, and crucially, does not presuppose the existence of an extension gap between the

---

<sup>8</sup> Note that either of these explanations may also be the reason why, besides the ratings of 2 reflecting negative strengthening (*not large*  $\rightsquigarrow$  'small') in Experiment 1, we still observe some amount of mid-scale ratings for negated positive weak relatives (*not large*), indicating the availability of a different and weaker interpretation from the simple antonym (*small*; see Figure 1).

semantic interpretations of antonymic absolute adjectives, unlike in the case of negated relative adjectives.

## 4.2 Explaining adjective interpretation under negation: Horn's theory

The findings of the present study are relevant for the theoretical controversy over the interpretation of negated antonymic adjectives in the scope of negation. Specifically, this controversy is about whether the interpretation of antonymic negated adjectives is governed by sociological factors (as in Horn's face-management-related theory) or by complexity-based considerations (as in Krifka's (2007) account). These two competing views yield distinct interpretation patterns for negated antonymic adjectives: An account assuming face-management considerations derives an asymmetric interpretation pattern only for antonymic relative adjectives in the scope of negation, whereas an account adopting complexity-based considerations derives symmetric interpretation patterns for negated antonymic relative and absolute adjectives. The interpretation asymmetry between negated antonymic weak relative terms with respect to polarity (*not large* vs. *not small*), that persists in the absence of overt contextual competition, and the absence of such an interpretation pattern for negated antonymic weak absolute adjectives are in support of Horn's (1989) as opposed to Krifka's (2007) account.

The attested interpretation pattern of negated positive relative terms (*not large*) is evocative of the so-called negative strengthening interpretation captured by Horn as an R-based implicature conditioned on sociological considerations: e.g., participants may consider that *not large* conceals the stronger negative meaning expressed by the simple antonym *small* and that the former, as opposed to the latter, is presumably preserving the hearer's face.<sup>9</sup> In turn, no similar social pressure applies to negated negative adjectives (*not small*), hence this condition does not receive ratings similar to the corresponding simple antonym (*large*). Rather, the interpretation pattern of negated negative terms receiving a high proportion of middle-scale ratings seems to be indicative of the availability of middling interpretations ('neither large nor small'), derived as a Q- and R-based implicature in Horn's theory (see Section 2.1): participants reason that a given weak and prolix double negative term (*not small*) is used to express a weaker meaning than that expressed by the corresponding simple antonym (*large*), namely the meaning 'neither small nor (quite) large' corresponding to the degrees in the extension gap between *large* and *small*.

---

<sup>9</sup> In a sense, in the present study, sociological considerations (cf. Horn, 1989) are operationalized in terms of evaluative polarity (across items), while we do not manipulate other factors relating to face preserving/threatening, such as the social distance or power relations between speaker and addressee. For such experimental investigations, we refer the reader to Gotzner & Mazzarella (2021), who discovered a pervasive effect of polarity on negative strengthening potential (among other findings of differential interactions with the sociological variables of social distance, power, and gender).

Negated absolute adjectives (*not clean/not dirty*) are interpreted symmetrically and showed no robust indications of the existence of an interpretation asymmetry like that observed for negated antonymic relative adjectives. However, they still allow for seemingly relative-like interpretations: negated weak absolute adjectives are interpreted differently from the corresponding simple antonyms. This is *prima facie* at odds with their entailment patterns (*not dirty*  $\Rightarrow$  ‘clean’, *not clean*  $\Rightarrow$  ‘dirty’) and standard semantic accounts of absolutes (e.g., Kennedy 2007) that do not assume the existence of a semantic extension gap between antonymic absolute adjectives. It would potentially be in line with Krifka’s complexity-based account predicting symmetric response patterns distinct from corresponding simple antonyms, which critically relies on the assumption that the pragmatic interpretation of antonymic adjectives leaves an extension gap in-between. Nevertheless, it is not necessary to subscribe to such an assumption, which would even call for a further explanation of why the two adjective classes under study present distinct interpretation patterns under negation while they both allow for an extension gap. Instead, we adopt an explanation for the apparent “gappy” behavior of negated absolute adjectives that alludes to Horn’s division of pragmatic labor (see Section 2.1). Importantly, this explanation avoids the need to invoke the existence of an extension gap and, thus, we do not have to abandon standard assumptions about the semantics of absolute adjectives.

We propose that the apparent “gappy” behavior of negated absolute adjectives may be the result of a reasoning that assumes some form of overt or covert competition between truth-conditionally equivalent forms appealing to distinctions due to the level of precision, which results in distinctions in meaning. As an illustration, when two truth-conditionally equivalent expressions like *clean* and *not dirty* become contextually salient and in competition, the simple adjective form *clean* may be taken to convey a stereotypical situation like that of being maximally clean (interpretation of *clean* at a high precision level, similar to *pristine*). In turn, the complex and marked form *not dirty* may then be assigned a non-stereotypical, weaker, meaning such as being neither really dirty nor (completely) clean (cf. Rotstein & Winter’s (2004) example (8) from section 2.1: *It is certainly not clean, since it has some small spots on it, but it is not really dirty*). Similarly, in the case of *not clean* vs. *dirty*, the simple form could be understood to convey a good amount of dirt (interpretation of *dirty* at a high precision level, similar to *filthy*), while the complex form might communicate a non-stereotypical situation such as presenting a minimal amount of dirt. This could perfectly account for why participants assign mid-scale ratings to both *not dirty* and *not clean* (vs. *clean/dirty*). It moreover accounts for the overlap of strong (*pristine/filthy*) and weak absolute conditions (*clean/dirty*) in the base environments without negation (cf. **Figure 3**). It is worth noting that one might assume that truth-conditionally equivalent expressions become salient and in competition across presentation modes. This assumption is based on the idea that some form of competition occurs indirectly in the single-statement presentation mode due to the

presence of strong antonymic adjectives as indicated scale endpoints in the context of every item (see previous section).

A further methodological point relating to the setup of our experiments may be relevant in explaining the difference of negated absolute conditions (*not clean/not dirty*) from the corresponding simple antonyms (*dirty/clean*): Participants assign to the simple antonym conditions without negation (*dirty* and *clean*) the ratings 1 and 2, on the one hand, and 4 and 5, on the other. Having done so, they rate the negated weak (*not clean/not dirty*) and strong absolute adjectives (*not pristine/not filthy*) similarly making use of the ratings they had not assigned to the simple antonym (3, 4, and 5 for *not dirty/not filthy*, and 1, 2, 3 for *not clean/not pristine*). Thus, for the negated weak and strong conditions they may select any of the remaining scale points that have not been assigned to the respective simple antonym. This explanation suggests a chronological sequence of assigning ratings from simple, non-negated adjective expressions to complex, negated ones. Crucially, under this explanation too, the observed symmetric response patterns are in line with the entailment patterns of absolute adjectives and standard semantic accounts of absolute adjectives like Kennedy's (2007). However, we believe that such strategic response behavior is less likely to occur in a single-statement, as opposed to a multiple-statement, presentation mode as participants cannot directly compare the responses they gave to different statements.

## 5 Conclusion

In the current study, we investigated the role of competition between alternative expressions in adjective interpretation under negation. An overt competition as induced by a concurrent presentation of alternative adjective expressions has been shown to bring about adjective meaning distinctions, which are less likely to arise otherwise. More specifically, the current study probed the role of competition in the interpretation of relative and absolute gradable adjectives in the scope of negation (e.g., *not large* and *not clean*, respectively), seeking to find out to what extent the absence of an overt contextual competition affects the distinct interpretation patterns of negated relative and absolute adjectives.

Our experimental results show that the absence of overt contextual competition between adjective expressions affects the interpretative asymmetry characteristic of positive and negative relative adjectives in the scope of negation (*not large* vs. *not small*; see Horn 1989): this polarity asymmetry appears to be enhanced without overt contextual competition. By contrast, we did not find any robust evidence that it affects the symmetric response patterns of positive (*not clean*) and negative absolute adjectives (*not dirty*), or that the apparent relative-like interpretations of negated absolute adjectives vanish completely in the absence of overt contextual competition.

In sum, our investigation of the role of competition in adjective interpretation yields indications that the contextual availability of expressions of different complexity that compete in meaning (i) weakens the negative strengthening potential of complex positive expressions like *not large* and promotes distinctions in meaning between simple and complex expressions in line with Horn's (1989) and Krifka's (2007) accounts, (ii) leaves the interpretation of complex negative expressions like *not small* unaffected, as this is grounded on the existence of an extension gap between antonymic adjectives, and (iii) is crucial to the apparent availability of relative-like interpretations for complex absolute adjective expressions (*not clean/not dirty*) when competing expressions are made overtly or indirectly available.

Finally, we propose to capture the apparent availability of relative-like interpretations of negated absolute adjectives by means of a reasoning akin to Horn's division of pragmatic labor.

---

## Supplementary files

**Supplementary file 1: Appendix.** Data related to the experimental materials, the previous experimental findings, and the experiments of the current study. DOI: <https://doi.org/10.16995/glossa.9919.s1>

## Acknowledgements

The authors wish to thank the audience of the Deutschen Gesellschaft für Sprachwissenschaft (DGfS) workshop on Change of State Verbs – Empirical and Theoretical Perspectives in Tübingen as well as Henrik Discher, Georgia Haralampous, Marisha Herb, and Heidi Klockmann for their valuable help and feedback on this work. Our research was supported by the DFG (Emmy Noether Grant GO 3378/1-1 awarded to Nicole Gotzner).

## Competing interests

The authors have no competing interests to declare.

---

## References

- Albu, Elena. 2020. Not known: Anonymous, unknown or non-known? A pilot test on the interpretation of negated absolute adjectives in Romanian. *Glossa: A Journal of General Linguistics* 5(1). DOI: <https://doi.org/10.5334/gjgl.666>
- Alexandropoulou, Stavroula & Gotzner, Nicole. accepted. The interpretation of relative and absolute adjectives under negation. *Journal of Semantics*.
- Blutner, Reinhard. 2000. Some aspects of optimality in natural language interpretation. *Journal of Semantics* 17(3). 189–216. DOI: <https://doi.org/10.1093/jos/17.3.189>
- Brown, Penelope & Levinson, Stephen. 1987. *Politeness: Some universals in language usage*. Cambridge & New York: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511813085>
- Burnett, Heather. 2014. A delineation solution to the puzzles of absolute adjectives. *Linguistics and Philosophy* 37. 1–39. DOI: <https://doi.org/10.1007/s10988-014-9145-9>
- Christensen, Rune H. B. 2019. Regression models for ordinal data. R package version: 2019.12-10.
- Colston, Herbert. 1999. “Not good” is “bad”, but “not bad” is not “good”: An analysis of three accounts of negation asymmetry. *Discourse Processes* 28(3). 237–257. DOI: <https://doi.org/10.1080/01638539909545083>
- Cruse, David. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.
- Doran, Ryan & Baker, Rachel E. & McNabb, Yaron & Larson, Meredith & Ward, Gregory. 2009. On the non-unified nature of scalar implicature: An empirical investigation. *International Review of Pragmatics* 1. 1–38. DOI: <https://doi.org/10.1163/187730909X12538045489854>

- Doran, Ryan & Ward, Gregory & McNabb, Yaron & Larson, Meredith & Baker, Rachel E. 2012. A novel paradigm for distinguishing between what is said and what is implicated. *Language* 88. 124–154. <http://www.jstor.org/stable/41348885>. DOI: <https://doi.org/10.1353/lan.2012.0008>
- Fraenkel, Tamar & Schul, Yaacov. 2008. The meaning of negated adjectives. *Intercultural Pragmatics* 5(4). 517–540. DOI: <https://doi.org/10.1515/IPRG.2008.025>
- Gotzner, Nicole & Kiziltan, Sybille. 2022. She is brilliant! Distinguishing different readings of relative adjectives. In *Measurement, numerals and scales*, 117–134. Palgrave Macmillan. DOI: [https://doi.org/10.1007/978-3-030-73323-0\\_7](https://doi.org/10.1007/978-3-030-73323-0_7)
- Gotzner, Nicole & Mazzarella, Diana. 2021. Face management and negative strengthening: The role of power relations, social distance and gender. *Frontiers in Psychology*. DOI: <https://doi.org/10.3389/fpsyg.2021.602977>
- Gotzner, Nicole & Solt, Stephanie & Benz, Anton. 2018a. Adjectival scales and three types of implicature. *Semantics and Linguistic Theory* 28. 409–432. DOI: <https://doi.org/10.3765/salt.v28i0.4445>
- Gotzner, Nicole & Solt, Stephanie & Benz, Anton. 2018b. Scalar diversity, negative strengthening, and adjectival semantics. *Frontiers in Psychology* 9. 1659. DOI: <https://doi.org/10.3389/fpsyg.2018.01659>
- Horn, Laurence R. 1972. *On the semantic properties of the logical operators in English*: Indiana University dissertation.
- Horn, Laurence R. 1989. *A natural history of negation*. Chicago: University of Chicago Press.
- Horn, Laurence R. 1993a. Economy and redundancy in a dualistic model of natural language. In Shore, Susanna & Vilkkuna, Maria (eds.), *Yearbook of the Linguistic Association of Finland*, 33–72. SKY.
- Horn, Laurence R. 1993b. *Duplex negatio affirmat...*: The economy of double negation. In Dobrin, Lise M. & Nichols, Lynn & Rodriguez, Rosa M. (eds.), *Papers from the 27<sup>th</sup> Regional Meeting of the Chicago Linguistic Society 1991: The parasession on negation.*, vol. 27 (2), 80–106. Chicago Linguistic Society.
- Katsos, Napoleon & Bishop, Dorothy VM. 2011. Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition* 120(1). 67–81. DOI: <https://doi.org/10.1016/j.cognition.2011.02.015>
- Kennedy, Christopher. 2007. Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* 30. 1–45. DOI: <https://doi.org/10.1007/s10988-006-9008-0>
- Kennedy, Christopher & McNally, Louise. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language* 81. 345–381. DOI: <https://doi.org/10.1353/lan.2005.0071>
- Krifka, Manfred. 2007. Negated antonyms: Creating and filling the gap. In Sauerland, Uli & Stateva, Penka (eds.), *Presupposition and implicature in compositional semantics*, 163–177. Houndmills: Palgrave Macmillan. DOI: [https://doi.org/10.1057/9780230210752\\_6](https://doi.org/10.1057/9780230210752_6)



- Leffel, Timothy & Cremers, Alexandre & Gotzner, Nicole & Romoli, Jacopo. 2019. Vagueness in implicature: The case of modified adjectives. *Journal of Semantics* 36(2). 317–348. DOI: <https://doi.org/10.1093/jos/ffy020>
- Mazzarella, Diana & Gotzner, Nicole. 2021. The polarity asymmetry of negative strengthening: Dissociating adjectival polarity from face-threatening potential. *Glossa: A journal of general linguistics* 6(1). 47. DOI: <https://doi.org/10.5334/gjgl.1342>
- Morzycki, Marcin. 2012. Adjectival extremeness: Degree modification and contextually restricted scales. *Natural Language and Linguistic Theory* 30. 567–609. DOI: <https://doi.org/10.1007/s11049-011-9162-0>
- Paradis, Carita & Willners, C. 2006. Antonymy and negation—The boundedness hypothesis. *Journal of Pragmatics* 38(7). 1051–1080. DOI: <https://doi.org/10.1016/j.pragma.2005.11.009>
- Rotstein, Carmen & Winter, Yoad. 2004. Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics* 12. 259–288. DOI: <https://doi.org/10.1023/B:NALS.0000034517.56898.9a>
- Ruytenbeek, Nicolas & Verheyen, Steven & Spector, Benjamin. 2017. Asymmetric inference towards the antonym: Experiments into the polarity and morphology of negated adjectives. *Glossa: A Journal of General Linguistics* 2(1). 1–27. DOI: <https://doi.org/10.5334/gjgl.151>
- Solt, Stephanie. 2015. Measurement scales in natural language. *Language and Linguistics Compass* 9(1). 14–32. DOI: <https://doi.org/10.1111/lnc3.12101>
- Terkourafi, Marina & Weissman, Benjamin & Roy, Joseph. 2020. Different scalar terms are affected by face differently. *International Review of Pragmatics* 12(1). 1–43. DOI: <https://doi.org/10.1163/18773109-01201103>
- Tessler, Michael H. & Franke, Michael. 2018. Not unreasonable: Carving vague dimensions with contraries and contradictions. In Kalish, Charles & Rau, Martina & Zhu, Jerry & Rogers, Timothy (eds.), *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 40, 1108–1113.
- Tessler, Michael H. & Franke, Michael. 2019. Not unreasonable: Why two negatives don't make a positive. DOI: <https://doi.org/10.31234/osf.io/tqjr2>
- Zehr, Jeremy & Schwarz, Florian. 2018. PennController for Internet Based Experiments (IBEX). DOI: <https://doi.org/10.17605/OSF.IO/MD832>

