Glossa
a journal of general linguistics

# The polarity asymmetry of negative strengthening: dissociating adjectival polarity from face-threatening potential

RESEARCH

**DIANA MAZZARELLA**
**NICOLE GOTZNER**
**Author affiliations can be found in the back matter of this article*

]u[ ubiquity press

## Abstract

The interpretation of negated antonyms is characterised by a polarity asymmetry: the negation of a positive polarity antonym (*X is not interesting*) is more likely to be strengthened to convey its opposite ('X is uninteresting') than the negation of a negative polarity antonym (*X is not uninteresting* to convey that 'X is interesting') is. A classical explanation of this asymmetry relies on face-management. Since the predication of a negative polarity antonym (*X is uninteresting*) is potentially face-threatening in most contexts, the negation of the corresponding positive polarity antonym (*X is not interesting*) is more likely to be interpreted as an indirect strategy to minimise face-threat while getting the message across. We present two experimental studies in which we test the predictions of this explanation. In contrast with it, our results show that adjectival polarity, but not face-threatening potential, appears to be responsible for the asymmetric interpretation of negated antonyms.

CORRESPONDING AUTHOR:
**Diana Mazzarella**
Université de Neuchâtel
Rue Pierre-à-Mazel 7, 2000
Neuchâtel, CH
*diana.mazzarella@unine.ch*

KEYWORDS:
negation; polarity; antonyms; negative strengthening; politeness; face

TO CITE THIS ARTICLE:
Mazzarella, Diana and 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. 1–17. DOI: *https://doi.org/10.5334/gjgl.1342*

# 1 Introduction

## 1.1 The polarity asymmetry of negative strengthening

Opposition is a fundamental relation expressed by all natural languages between propositions or predicates that are incompatible with one another (for seminal work on antonymy in philosophy and linguistics see Vendler 1963; Givón 1970; Lehrer & Lehrer 1982; Cruse 1986; Horn 1989). Since Aristotle, we distinguish between at least two types of opposition: contradictory opposition and contrary opposition. The former, but not the latter, obeys to the Law of the Excluded Middle. Contradictory antonyms, like the pairs of predicates *even* and *odd*, exhaust their semantic space. For any entity in the relevant domain (e.g. integers), it is the case that either it is 'even' or 'odd'. In contrast, contrary antonyms, such as *happy* and *unhappy*, allow for an unexcluded middle (an emotional state that is neither 'happy' nor 'unhappy'). While they cannot be both true of the same relevant entity at the same time, they can be both false.

Many scholars have investigated the interplay between sentential negation, on the one hand, and contrary antonyms, on the other (Horn 1989; 2017; 2020; Krifka 2007; Neuhaus 2016 *inter alia*; for recent experimental investigations, see Ruytenbeek, Verheyen & Spector 2017; Gotzner, Solt & Benz 2018). Given the lexical semantics of contrary antonyms, an utterance like *The professor is not happy with the essay* semantically encodes a meaning which spans from the so-called *zone of indifference* (representing the unexcluded middle; see Sapir 1944) to the opposite 'unhappy'. However, when this utterance is produced in the right context, it can license an inference to 'The professor is unhappy with the essay', which corresponds to the affirmation of the contrary (see ***Figure 1***). This pragmatic inference is referred to as *negative strengthening* (Horn 1989), *middle-excluding inference* (Horn 1989) or *inference to the antonym* (Ruytenbeek et al. 2017), and delivers a stronger interpretation of sentential negation.

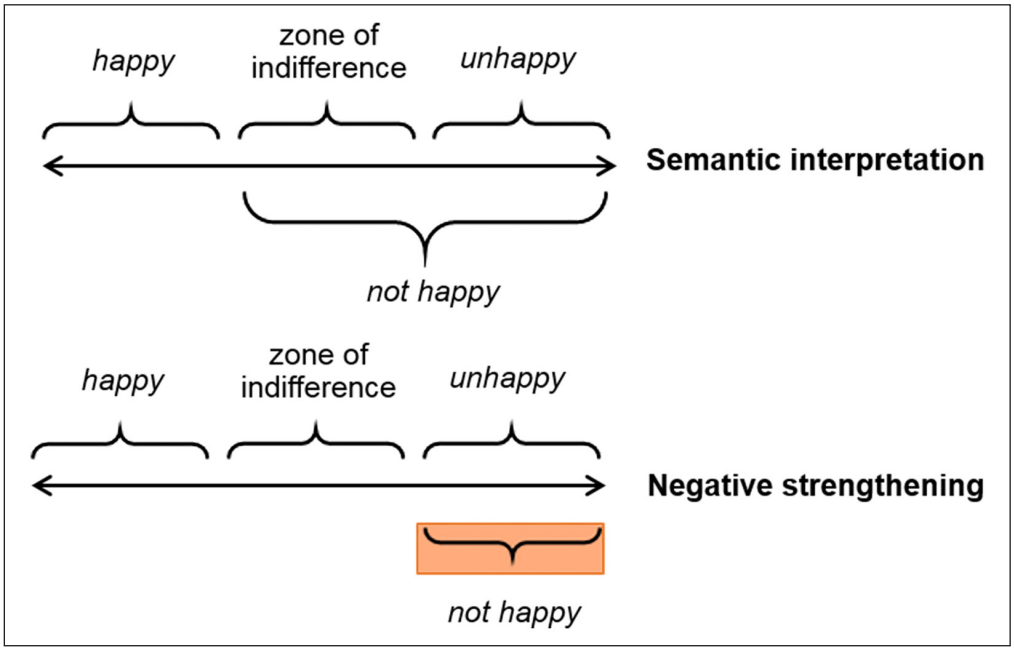


**Figure 1** Semantic interpretation and negative strengthening of 'not happy'. The range of the pragmatically strengthened interpretation of negation is highlighted with colour.

Horn (1989) defines negative strengthening as an inference based on the R-principle ("Say no more than you must") via which the addressee strengthens the negation of an antonym to a more specific, contrary understanding. In contexts licensing negative strengthening, the speaker counts on the addressee's willingness to fill in the intended R-strengthened (contrary) interpretation, rather than taking the formally contradictory negation as such. As we will explore in the next section, Horn maintains that negative strengthening, *qua* R-based inference, is socially – rather than linguistically – motivated (Horn 1989: 358).

Negative strengthening is a particularly interesting phenomenon at the interface between semantics and pragmatics since it appears to be modulated by adjectival polarity. Each antonymic pair consists of two opposites: a positive polarity predicate and a negative polarity one. Polarity is typically defined based on three distinct – often, but not always, converging – criteria (see Cruse 1986 and Ruytenbeek et al. 2017 for discussion). The first one, *evaluative*

*polarity*, is based on judgements of desirability. The positive antonym is desirable, the negative one is undesirable: for instance, *happy* is desirable, thus positive; *unhappy* is undesirable, thus negative. The second one, *dimensional polarity*, is based on the measurement scale associated with the antonymic pair. For each antonymic pair, the positive antonym is the one mapping onto the relevant dimension for the measurement on the scale. For instance, *interesting* is positive and *uninteresting* is negative as they are associated with the scale of interest. The third criterion is markedness, often linked to morphological markedness: the positive antonym is unmarked, the negative one is marked.[1] For instance, as both *unhappy* and *uninteresting* are morphologically marked by the negative prefix *un-*, they represent the negative antonym. Interestingly, Ruytenbeek et al. (2017) have operationalised the notion of polarity as a function of the application of two linguistic tests: i) the *pas très* ('not very') test, which collects acceptability judgement on sentences of the form *X n'est pas très P* and *X n'est pas très Q* (e.g. *Jean n'est pas très grand/petit*, i.e. 'Jean is not very tall/short'); and ii) the exclamation task, which collects acceptability judgements on sentences of the form *C'est étonnant à quel point X n'est pas P* and *C'est étonnant à quel point X n'est pas Q* (e.g. *C'est étonnant à quel point Jean n'est pas grand/petit*, i.e. 'It is surprising to what point Jean is tall/short'). The two tests show high reliability, are highly correlated with each other and are compatible with a priori polarity classifications based on traditional criteria (evaluative, dimensional and markedness). For this reason, they were taken as an empirically-based operationalisation of the notion of polarity (for discussion, see Ruytenbeek et al. 2017: 10–11).

Polarity appears to play a crucial role in the interpretation of negated adjectives. Linguists since Ducrot (1973) have observed that the negation of a positive polarity antonym is more likely to be strengthened than the negation of a negative polarity antonym is. That is, while the utterance *The professor is not happy with the essay* is likely to be interpreted as conveying that 'The professor is unhappy with the essay', *The professor is not unhappy with the essay* is less likely to receive the strengthened interpretation 'The professor is happy with the essay' (see also Horn 1989: 334–337 for discussion).

In the last twenty years, the asymmetric interpretation of positive and negative adjectives has been investigated in a few experimental studies. Colston (1999) reported evidence of this asymmetry in contexts inducing positive expectations (but not in contexts inducing negative ones). When a positive outcome was expected, participants interpreted negated positives (*It's not exciting*) as expressing an equally negative evaluation as direct negatives (*It's boring*), while their interpretation of negated negatives was not strengthened to the same extent (*It's not boring* was considered less positively than the direct positive *It's exciting*). In contrast with this finding, though, Giora et al. (2005) found that the negation of both evaluatively positive and negative adjectives was taken to convey a weaker, "tinged", evaluation than the affirmation of the antonym, with no evidence of a polarity asymmetry.[2] Furthermore, Fraenkel and Schul (2008) provided evidence of a polarity asymmetry based on adjectival markedness. Crucially, these studies relied on different measures of negative strengthening (respectively, contextualised explicit inferential judgments, decontextualised explicit inferential judgments, and judgements of meaning similarity), as well as on a priori categorisations of polarity based on alternative criteria, such as evaluative polarity and markedness (see Ruytenbeek et al. 2017 for discussion). For this reason, it is not surprising that the evidence provided is at times not consistent.

Recently, the polarity asymmetry of negative strengthening has received strong empirical support from a study by Ruytenbeek et al. (2017), whose main findings have also been confirmed by Gotzner and Mazzarella (2020). In what follows, we discuss Ruytenbeek et al.'s (2017) study in more detail as it provides the starting point for our experimental investigation. Ruytenbeek and colleagues measured the degree of negative strengthening (or, in their terminology, the strength

1 For a characterisation of markedness that does not rely on morphology, see Vendler (1963), Cruse (1986) and Rett (2015). Rett's (2015) markedness test relies on the presuppositional behaviour of adjectives in equative comparisons. For instance, an adjective like *short* is marked as an utterance of *Sally is as short as Mary* presupposes that both Sally and Mary are short (see the difference with the unmarked adjective *tall* in *Sally is as tall as Mary*).

2 It is worth noting that in Giora et al.'s study participants were presented with pairs of target sentences (e.g. *Sari's dress was ugly/Sari's dress was not pretty*) and had to rate them on a *single* 7-point scale (anchored at the two antonyms, e.g. *ugly* and *pretty*). This feature of their design might have prompted participants to provide different ratings for these target sentences, thus reducing the likelihood of negative strengthening (e.g. *not pretty* to be rated as 'rather ugly').

of the inference towards the antonym) in the interpretation of the sentential negation of positive and negative antonyms by employing two different measures. In their first experiment, they collected an indirect measure of negative strengthening that is based on acceptability judgments of sentences of the form *X is not P. Y is Q too*, where *P* and *Q* represent an antonymic pair (e.g. *Paul n'est pas grand. Pierre aussi est petit.*, i.e. 'Paul is not tall. Pierre is also short'). In their second experiment, Ruytenbeek et al. (2017) collected explicit inferential judgements in which they asked participants to judge the subject of the sentence on a continuous scale anchored at *P* and *Q* (e.g., *Paul n'est pas grand* judged on a scale from *grand* to *petit*).

Across both studies, participants were significantly more likely to strengthen the negation of a positive antonym than the negation of a negative antonym, thus confirming the main effect of polarity on the degree of negative strengthening. Furthermore, Ruytenbeek et al.'s (2017) results show an effect of morphological complexity on the degree of negative strengthening: the asymmetry between positive and negative negated adjectives was stronger for morphological pairs than for non-morphological pairs. In sum, Ruytenbeek and colleagues' study has offered robust empirical evidence of the polarity asymmetry of negative strengthening, as well as an effective operationalisation of adjectival polarity. Building on their work, we move one step further and address the question of *why* the polarity asymmetry holds.

Before turning to this, though, it is worth specifying that the linguistic observations and empirical findings reported above concern *weak* positive adjectives (*happy*) but not their stronger scale mates, that are often referred to as *strong* positive adjectives (*ecstatic*). Horn (1989: 337) first observed that the negation of a strong positive scalar is typically interpreted literally, and is less likely to license negative strengthening. The following examples are taken from Israel (2004: 708):

(1) a. He's not mean. (≠ He's nice)
b. She's not sad. (≠ She's happy)
c. She's not ecstatic. (≠ She's miserable)

Horn (1989) explains this difference by observing that the negation of strong positive adjectives occurs more naturally in linguistic contexts containing a previous mention of the adjective (which is then explicitly denied) than as an evaluation initiating an exchange. According to Horn, "[i]n such discourse frames, there is no functional motivation for moving beyond the straightforward (contradictory) assigned by the syntax" (1989: 360).[3] For this reason, any reference to the polarity asymmetry of negative strengthening in this paper should be taken as referring to the contrast between *weak* positive adjectives and negative ones. In the following section, we present a well-known explanation of this pragmatic phenomenon, which will be the object of the experimental investigation presented here.

## 1.2 Negative strengthening and face-management

Many scholars explain the polarity asymmetry observed in the interpretation of negated adjectives as originating from the fact that negation can pragmatically function as a tool for face-management.

To begin with, it is thus worth defining some of the key concepts related to face-management that will play a role in our discussion of negative strengthening. The core notion is that of *face*, a notion introduced by the sociologist Erving Goffman, who defines it as "the positive social value a person effectively claims for himself by the line others assume he has taken during a particular contact", where a *line* corresponds to the "verbal and nonverbal acts by which he expresses his view of the situation and through this his evaluation of the participants, especially himself" (Goffman 1967: 5). According to Goffman, participants in a conversation tend to act in ways compatible with the maintenance of face, be this their own face or their co-participants' face. When their objectives pose a potential threat to face – for instance, when the speaker aims at "giving free expressions

3 These observations are in line with the results of an experiment by Gotzner and Kiziltan (forthcoming). The experiment used a grading scenario, asking participants to associate different negated and non-negated terms on a scale. In this scenario, participants used distinct portions of a scale when interpreting statements involving non-negated weak and strong adjectives (e.g., *large* and *gigantic*). When the same terms appear under negation, participants distinguish positive weak terms (e.g., *not large*) from their stronger scale-mates (e.g., *not gigantic*) but not the corresponding negative antonyms (e.g., *not tiny* and *not small*). Weak and strong negative terms as well as strong positive terms tended to receive a middling interpretation. Thus, the polarity asymmetry of negative strengthening was only present for weak adjectives but not for stronger scale-mates.

to one's true beliefs, introducing deprecating information about the others" (Goffman 1967: 11) – speakers will attempt to employ some face-saving practice. Crucially, Goffman recognises that the maintenance of one's face and that of others' face are closely intertwined and are pursued in parallel during every interaction. For instance, by attempting to save the face of an addressee, a speaker might avoid the hostility that would follow in case of the addressee's face loss, thus contributing to the maintenance of his or her own face. Following Goffman, Brown and Levinson (1987) maintain that any rational agent whose act is potentially face-threatening (FTA) "will take into consideration the relative weightings of (at least) three wants: (a) the want to communicate the content of the FTA *x*, (b) the want to be efficient or urgent, and (c) the want to maintain H[earer]'s face to any degree. Unless (b) is greater than (c), S[peaker] will want to minimise the threat of his FTA." (Brown & Levinson 1987: 68). Crucially, among the linguistic practices that serve this interactional function, they introduce *off-record indirectness*. Off-record utterances allow speakers to convey face-threatening content indirectly (implicitly) while keeping open the possibility of disavowing this content if openly confronted (see also Holtgraves 1986; 1994).

In their discussion on the interpretation of negated antonyms, Brown and Levinson (1987) explain the polarity asymmetry in terms of face-management considerations. They suggest that when one performs a face-threatening act such as criticising, "there is a good social motive for saying much less than you mean" (Brown & Levinson 1987: 264). For this reason, the negation of a positive polarity antonym – as in *John is not a friend* – will be likely to be used (and interpreted) as an understatement, an off-record strategy motivated by face-maintenance concerns. That is, *John is not a friend* will often convey the affirmation of the corresponding negative antonym, 'John is an enemy', as a defeasible implicature. In contrast with this, because there is typically no good social motive that prevents speakers from saying that *John is a friend* directly, an utterance like *John is not an enemy* will not be interpreted as conveying that 'John is a friend'. The polarity asymmetry in the interpretation of negated antonyms is thus traced back to the different face-threatening potentials of utterances involving the predication of a negative antonym (typically face-threatening) or a positive antonym (typically not face-threatening) (see also Ducrot 1973).

This face-management based explanation of the polarity asymmetry of negative strengthening has been embraced and developed by many scholars. For instance, Horn (1989: 360) argues that negative strengthening is "motivated by the goal of avoiding the direct assertion of some negative proposition in a context in which it would tend to offend the addressee, overcommit the speaker, or otherwise count as inappropriate".[4] As a result, negative strengthening is more likely to occur in contexts that make these social motives particularly relevant, such as those characterised by "gradable predications involving desirable properties, […] whose denial would reflect undesirably on the subject, speaker and/or addressee" (Horn 1989: 334). In these contexts, the speaker will partially conceal his or her disapproval by adopting a weaker-seeming formulation (e.g. *He is not nice*) which encourages the addressee to recover via pragmatic inference the stronger negative judgement ('He is nasty') that remains implicit. Crucially, though, this socially motivated inference will be less likely to be licensed in contexts in which the speaker's formulation (e.g., *He is not nasty*) does not seem to be prompted by the desire to avoid the direct expression (*He is nice*). For this reason, we can conclude that "[t]he asymmetry is attributable to politeness […] yielding the practical maxim 'If you have something negative to say, don't say it directly'" (Horn 2017: 161). In the same vein, Israel (2004: 709) suggests that negation can serve the pragmatic function of managing face in that it "provides an oblique way of delivering the loaded content" of a speaker's negative judgement.

In sum, according to these authors, the asymmetry in the interpretation of negated positive and negative antonyms is ultimately due to their relative tendency to be employed in the performance of a face-threatening act (for an overview of the relevant considerations, see Leech 2014: 192–193). Given that the predication of a negative antonym is more likely to raise a face-threat than the predication of a positive antonym, speakers will tend to replace the former, but not the latter, with the weaker formulation corresponding to the negation of the antonym. As a consequence, addressees will be more likely to strengthen the negation of a positive antonym than the negation of a negative one.

4 Note that the avoidance of overcommitment is also a face-saving strategy, targeted at preserving the speaker's reputation as a trustworthy source of information (see, e.g., Vullioud, Clément, Scott-Phillips & Mercier 2017; Mazzarella, Reinecke, Noveck & Mercier 2018).

In Gotzner and Mazzarella (2020), we investigated the role of face management in interpreting negated adjectives by manipulating the power relation of dialogue partners and their social distance. We found that face management considerations had an impact on the degree of negative strengthening of both positive and negative adjectives. For instance, the results showed that the greater the power of the hearer over the speaker, the stronger was the degree of negative strengthening. The results also showed a stronger polarity asymmetry for female than male participants and different weightings of power and social distance across gender. In the current study, we investigate the role of face-management in negative strengthening by taking an alternative approach, which we introduce in the next section.

### 1.3 Negative strengthening in non-ordinary contexts

Brown and Levinson (1987: 264–265) discuss the following *prima facie* counterexamples to their face-management explanation of the polarity asymmetry of negative strengthening. First, imagine that John has just been arrested as a Commie spy. By stating that *John is not an enemy of mine*, the speaker could implicate that John *is* a friend of theirs. In this context, being John's friend could endanger the speaker's face. This provides a social motive to merely implicate, rather than assert, that John is a friend. Second, imagine a speaker uttering *She's not bad* in a context in which complimenting the female individual at issue might represent a face-threatening act towards the addressee (e.g., she is a competitor of the addressee). In this case, the speaker uses this formulation to implicate that she is very good. According to Brown and Levinson, despite the negation of *enemy* and *bad* is strengthened to convey 'friend' and 'good' (thus escaping the mitigation or middling effect that is typical of double negatives), these examples are only "apparent exceptions [that] in fact support our argument" (Brown & Levinson 1987: 264). Indeed, they confirm the relevance of face-management considerations in negative strengthening: while in most contexts the speaker would not perform a face-threatening act by using a positive antonym (e.g., *X is happy*), this is not the case in the examples at issue. In both cases, to preserve their face or the face of the addressee, speakers would be less likely to convey a message with an explicit predication of the positive antonym. They would rather opt for an indirect strategy: e.g., asserting *X is not unhappy* to implicate that 'X is happy'. This suggests that, independently of adjectival polarity, negative strengthening is motivated by face management considerations and that the observed polarity asymmetry is the result of the fact that negative polarity antonyms are more likely than positive polarity antonyms to be face-threatening (all else being equal).

In what follows, we elaborate on Brown and Levinson's discussion, which we take as the starting point of our experimental investigation. Based on Brown and Levinson's examples, we define a context as *ordinary* relative to a given antonymic pair if the predication of the negative polarity antonym represents a face-threatening act towards the speaker's/hearer's face. Conversely, a context is defined as *non-ordinary* – relatively to the same antonymic pair – if it is the predication of the positive polarity antonym to represent a face-threatening act towards the speaker's/hearer's face. In other words, non-ordinary contexts reverse the frequent association between negative polarity, on the one hand, and face-threatening potential, on the other hand. Consider the antonymic pair *good/bad*. Relative to this pair, we can compare an utterance of *The CV is not good/bad* in an ordinary context (as in 2) and a non-ordinary context (as in 3):

(2) Ordinary context for *good/bad*
You want to join a prestigious company and you are competing with one of your current colleagues. After looking at your CV, your officemate tells you: *The CV is not good/bad.*

(3) Non-ordinary context for *good/bad*
You want to join a prestigious company and you are competing with one of your current colleagues. After looking at your competitor's CV, your officemate tells you: *The CV is not good/bad.*

In the ordinary context (2), the utterance *The CV is bad* would represent a face-threatening act towards the face of the addressee, whereas in the non-ordinary context (3) the utterance *The CV is good* would be face-threatening towards the face of the addressee. For this reason, speakers might be more likely to employ an indirect formulation that minimises the face-threat. This suggests that while in an ordinary context speakers might be more likely to utter *The CV is not*

*good* to implicate that 'The CV is bad', in a non-ordinary context they will be more likely to utter *The CV is not bad* to implicate that 'The CV is good'. With this in mind, addressees will thus be more inclined to strengthen the negation in *The CV is not good* than in *The CV is not bad* in an ordinary context, but to do the opposite in a non-ordinary one.

The face-management explanation of the polarity asymmetry of negative strengthening predicts that this asymmetry should be reversed in non-ordinary contexts. This is because, according to Brown and Levinson (1987), this asymmetry is ultimately due to considerations about the face-threatening potential of the formulation containing the bare adjective over the indirect formulation in which the corresponding antonym is negated. For this reason, we should expect that the negation of a positive polarity antonym would be more likely to be strengthened than the negation of a negative polarity antonym in ordinary contexts, while the negation of a negative polarity antonym would be more likely to be strengthened than the negation of a positive polarity antonym in non-ordinary contexts. Evidence of this reversal of the polarity asymmetry across ordinary and non-ordinary contexts would thus represent crucial evidence in favour of the face-management explanation of negative strengthening. In the following experiment, we investigate the polarity asymmetry of negative strengthening in non-ordinary contexts.

# 2 Experiment 1

## 2.1 Method

### 2.1.1 Participants

We recruited 60 participants with US IP addresses on Mechanical Turk (across two experimental lists). Participants were screened for native language and only included in the analysis if their self-reported native language was English. 24 women and 36 men participated in the study. Their mean age was 35.56, with a standard deviation of 10.61 (age range 23 to 72). The experiment lasted about 10 minutes and participants were paid $0.80 in compensation.

### 2.1.2 Materials

We selected 20 antonymic pairs from Ruytenbeek et al. (2017), which displayed consistent polarity across different criteria (evaluativity, dimensionality and markedness). As the original items from Ruytenbeek et al. (2017) were in French, we verified that English translation equivalents had the same polarity based on our intuitions concerning the output of the relevant criteria (see Appendix A). The list contained 11 morphological pairs and 9 non-morphological ones. Target sentences displayed the positive and negative members of each antonym pair in a negated statement, thus totalling 40 critical items. Target sentences were preceded by a context, which described the conversational situation. The context was always non-ordinary, that is, it was designed in such a way that the predication of the bare positive polarity antonyms represented a potential face-threat to the addressee. The critical utterance either contained the negated positive adjective or its negated antonym. ***Table 1*** shows an item as an example. The complete list of stimuli is available in Appendix B.

| |
|---|
| Context: You have decided to quit your current job and change career path. You are not popular in your team. A colleague from another department, tells you: |
| *Your team is not sad.* |
| According to your colleague, the team is: |
| sad 1 2 3 4 5 6 7 happy |

**Table 1** Example item for the adjective *sad* in Experiment 1 (negative polarity, non-ordinary context).

The task of the participants was to indicate what the speaker wanted to communicate on a scale ranging from the negated adjective to its antonym.[5] For instance, in the example item, participants judged the extent to which – according to the speaker – the team is 'happy'/'sad'. Judgments were given on a 7-point Likert scale anchored at the negated adjective (1) and its

5 In contrast to Ruytenbeek et al.'s (2017) study, our test question involved the explicit attribution of the intended implication to the speaker. This was meant to ensure that participants provided answers which reflected their interpretation of the speaker's utterance and not their own beliefs (see Dulcinati 2018 for an illustration of how 'epistemic questions' and 'meaning questions' can elicit different patterns of response).

antonym (7). Hence, we measured the degree of negative strengthening as a function of the likelihood with which the antonym is taken to be conveyed by the speaker's utterance.

The main manipulation was adjectival polarity (positive vs. negative), which was administered in a within-subject but between-item design. That is, one participant would only see either the positive or negative adjective of a given antonymic pair. Hence, each participant saw 10 statements with positive and 10 statements with negative adjectives, resulting in 20 critical trials and an overall number of 1200 critical observations. In addition to the critical items, participants were presented with 10 filler statements not involving negation, such as 'John is gorgeous' (where the response scale was anchored at the adjectives 'gorgeous' and 'ugly'). The filler sentences also served as attention checks.

The experiment was programmed in HTML and run via MTurk's built-in environment. The experimental procedures and predictions were pre-registered with the as.predicted.org template. The pre-registration is available on the Open Science Framework at the following link: *osf.io/zrwmg*.

#### 2.1.3 Procedure

Participants read an instruction explaining the task with an example. The running example was an adjective not used in the stimulus set (i.e., You ask your friend John: *How do I look?* and John responds: *You are not gorgeous*). For each stimulus, the 1–7 point scale was anchored to the negated adjective used by the speaker (1) and its antonym (7). The instructions told participants to judge what the speaker wanted to convey in each dialogue. Experimental trials and filler trials were randomised for each participant using a built-in randomisation function.

### 2.2 Predictions

The prediction based on the face-management explanation of the polarity asymmetry of negative strengthening is that the negation of negative polarity antonyms will be more likely to be strengthened than the negation of positive polarity antonyms given the 'non-ordinary' nature of the contexts. That is, the prediction is a reversal of the traditional polarity asymmetry (as pre-registered at *osf.io/zrwmg*).

### 2.3 Results

The data were analysed using R (version 3.6.2). We excluded the data of four participants based on inconsistent responses in the filler trials (more than 50% responses not in line with the adjective used in the filler statements). ***Figure 2*** shows the mean responses by adjectival polarity.

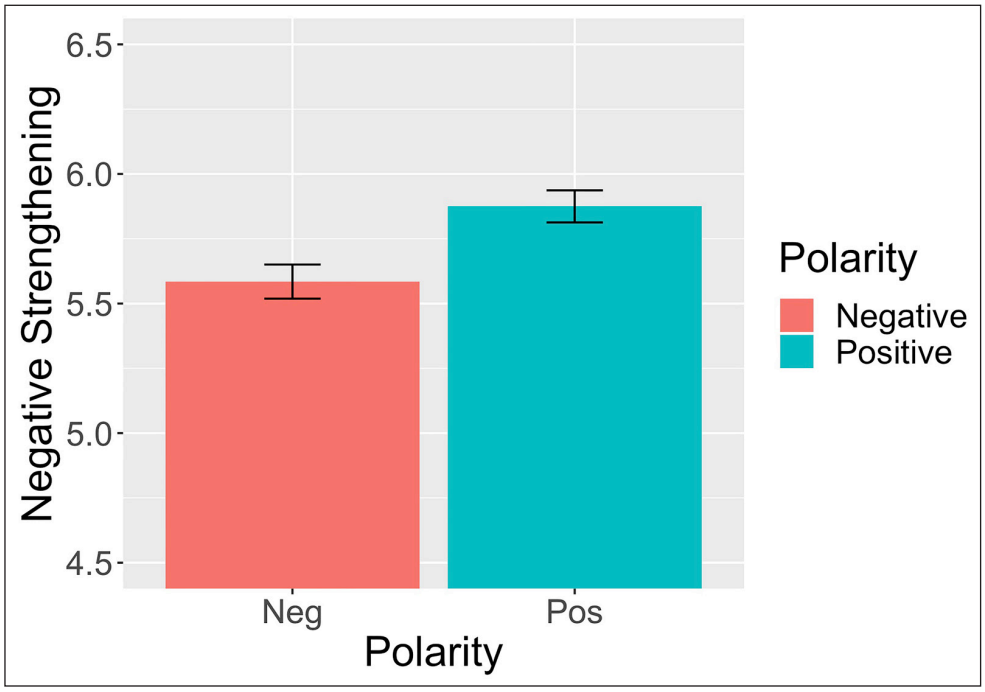


**Figure 2** Mean degree of negative strengthening by polarity (Experiment 1). Error bars represent +/– 1 SEM.

All results were analysed with cumulative link mixed-effects models using the function clmm() in the ordinal package (Christensen 2018), as they are more appropriate than linear mixed models for Likert scales.[6] We included the fixed factor polarity as well as random intercepts for items and participants. The factor polarity was sum-coded. The results of the model showed a main effect of polarity, with positive adjectives involving a higher degree of negative strengthening than negative ones ($B = -.27$, $SE = .06$, $z = -4.38$, $p < .001$). Thus, participants were significantly more likely to interpret a negated positive adjective such as *not happy* as conveying 'rather sad' than to interpret a negated negative adjective such as *not sad* as conveying 'rather happy'. Against our prediction, this finding replicates the traditional polarity asymmetry in non-ordinary contexts, in which the bare negative antonym (e.g. *sad*) does not pose a potential face-threat. A summary of the model is presented in ***Table 2***.

| | Estimate | SE | z-value | p-value |
|---|---|---|---|---|
| Polarity | −0.26985 | 0.06164 | −4.378 | 0.000012 |

**Table 2** Summary of cumulative link mixed-effects model including the sum-coded fixed effect polarity (Experiment 1).

Following Ruytenbeek et al. (2017), we also analysed whether the polarity asymmetry was greater for morphologically-marked antonyms (e.g., *happy/unhappy*, where the negative antonym contains a negative morpheme) compared to lexical, non-morphological, antonyms (*happy/sad*). The model included the sum-coded fixed factors polarity, morphological complexity and their interaction, as well as random intercepts for participants and items. In line with Ruytenbeck et al. (2017), we found an interaction between polarity and morphological complexity. This interaction shows that the difference between the negative strengthening of positive and negative terms is greater in morphologically-marked antonymic pairs such as *happy/unhappy* than in lexical antonyms such as *happy/sad* ($B = -.15$, $SE = .06$, $z = -2.4$, $p < .05$). In addition, there was a main effect of morphological complexity showing that morphologically-marked antonymic pairs displayed a greater degree of negative strengthening compared to lexical antonyms ($B = .13$, $SE = .06$, $z = 2.2$, $p < .05$). The details of this analysis are presented in Appendix C.

Finally, as previous research suggests the existence of gender differences in negative strengthening (Gotzner & Mazzarella 2020), we conducted an analysis with participant gender as a binary treatment-coded factor. We computed a model with polarity and participant gender. The model again revealed a main effect of Polarity but no main effect of participant gender ($p = .49$) and no interaction between polarity and participant gender ($p = .48$) (see Appendix C for detailed results).

## 2.4 Discussion

The results of Experiment 1 disconfirmed the prediction, based on the face management explanation of the polarity asymmetry of negative strengthening, that the negation of negative antonyms would be more likely than the negation of positive antonyms to be strengthened in non-ordinary contexts. Despite the nature of the context, the results revealed a robust polarity asymmetry in the usual direction, with negated positive antonyms being strengthened to a higher degree than negated negative antonyms. This finding leads us to conclude that, when adjectival polarity and face-threatening potential compete with each other, adjectival polarity wins over face-threatening potential. The asymmetric interpretation of antonyms does not seem to be driven by their respective face-threatening potential, but adjectival polarity *per se* appears to be responsible for the observed asymmetry. Crucially, our results replicate the main finding of Ruytenbeek et al. (2017), which was obtained for decontextualised sentences involving negated antonyms. Furthermore, they also replicate the findings on morphological complexity and support their conclusion that the polarity asymmetry in the interpretation of negated antonyms is stronger for morphological than non-morphological antonymic pairs (see also Gotzner & Mazzarella 2020). However, the asymmetry between positive and negative adjectives was only present for morphologically complex antonyms and not for lexical antonyms.[7]

6 The function clmm() is the more recent variant of clmm2(), allowing for the implementation of multiple random effects. However, since the status of random slopes for ordinal models is debated, we only included random intercepts.

7 We note that this experiment contained half the number of critical observations compared to the Gotzner and Mazzarella's (2020) study.

Further, as Experiment 1 only included non-ordinary contexts (that is, context type was not experimentally manipulated), it is not possible to directly compare the strength of the polarity asymmetry across ordinary and non-ordinary contexts. In light of the results of Experiment 1, though, this comparison would be useful to determine whether face-management contributes to – even if it does not explain – the polarity asymmetry of negative strengthening. This question is addressed in Experiment 2.

# 3 Experiment 2

## 3.1 Methods

### 3.1.1 Participants

We recruited 90 participants with US IP addresses on Mechanical Turk (45 participants across two experimental lists).[8] Participants were screened for native language and only included in the analysis if their self-reported native language was English. One participant's native language was not English and was therefore excluded from further analyses. 53 men and 36 women participated in the study (one participant did not respond to the gender question). Their mean age was 36.01, with a standard deviation of 10.1 (age range 21 to 60). The experiment lasted 15 to 20 minutes and participants were paid $1 in compensation.

### 3.1.2 Materials

We used the same list of adjectives from Experiment 1 (see Appendix A). The experimental stimuli were constructed as in Experiment 1, for a total of 40 critical items. Target utterances, containing the negated antonym, were embedded in ordinary contexts, in which the bare negative poses a potential face-threat, or non-ordinary contexts, in which the bare positive poses a potential face-threat. For each antonymic pair, we constructed a minimal pair of contexts in which the manipulation of minimal linguistic material would turn the context from an ordinary to a non-ordinary one. ***Table 3*** shows an example stimulus. The complete list of stimuli is available in Appendix D.

| |
|---|
| Context: After thinking of a career change for a long time, you have decided to ***stay in/quit*** your current job. A colleague from another department tells you: |
| *Your team is not happy.* |
| According to your colleague, the team is: |
| happy 1 2 3 4 5 6 7 sad |

**Table 3** Example item for the adjective *happy* (positive polarity) in Experiment 2. The minimal manipulation of the context is displayed in ***bold:*** *stay in* (ordinary context), *quit* (non-ordinary context).

The task of the participants was the same as in Experiment 1. Our two factors, Polarity and Context, were all within-subject but spread across two different item lists in a Latin square design. Each participant saw 20 statements with positive and 20 statements with negative adjectives, rotated over context conditions. That is, each participant completed 40 critical trials. The resulting overall number of critical observations was 3560, hence the sample size was three times as large as that of Experiment 1. In addition to the critical items, participants were presented with 10 filler statements not involving negation, which also served as attention checks.

The experiment was programmed in HTML and run via MTurk's built-in environment. The pre-registration form of the second experiment is available at the following link: *osf.io/z963u*.

### 3.1.3 Procedure

The procedure was the same as in Experiment 1.

## 3.2 Predictions

Experiment 2 tests the contribution of face-management considerations to the asymmetric interpretation of negated antonyms. In what follows, we outline three alternative hypotheses on the relative contribution of face-threatening potential and adjectival polarity and their respective predictions (pre-registered at *osf.io/z963u*).

8 In Experiment 2 we recruited 90 rather than 60 participants in order to maximize the possibility of observing differences across contexts and polarity. The choice of this bigger sample size was pre-registered.

*Hypothesis 1*: The polarity asymmetry of negative strengthening is explained by face-management considerations, hence it is the result of the differential face-threatening potential of positive and negative antonyms in context (Brown & Levinson 1987). This hypothesis, which was tested and disconfirmed in Experiment 1, gives rise to the following prediction: in ordinary contexts, negation is more likely to be strengthened for positive antonyms than for negative ones, but the opposite is true in non-ordinary contexts. Given the findings of Experiment 1, though, we expected to replicate the usual polarity asymmetry across both ordinary and non-ordinary contexts. For this reason, Experiment 2 was designed to disentangle the following two hypotheses.

*Hypothesis 2*: Adjectival polarity and face-management considerations both contribute to the asymmetric interpretation of polar opposites. This hypothesis predicts that context type modulates the strength of the polarity asymmetry. Specifically, it is expected that the polarity asymmetry will be stronger in the ordinary context condition than in the non-ordinary context condition. Hypothesis 2 thus predicts a main effect of polarity and an interaction between polarity and context.

*Hypothesis 3*: Adjectival polarity alone is responsible for the polarity asymmetry of negative strengthening, and face-management does not play a role. This hypothesis predicts a main effect of Polarity, but no interaction between polarity and context. That is, the strength of the polarity asymmetry is not modulated by the face-threatening potential of the antonyms in context.

These three alternative hypotheses were put forth to establish *a priori* how any possible pattern of results would be interpreted (thus avoiding HARKing – *Hypothesising After Results are Known*; see Kerr 1998). While *Hypothesis 1* is the only hypothesis which is directly derived from a specific theoretical framework (Brown & Levinson 1987), *Hypothesis 2* and *Hypothesis 3* are compatible with any accounts of the polarity asymmetry of negative strengthening that admit a context-independent contribution of adjectival polarity.

### 3.3 Results

The data were analysed using R (version 3.6.2). We excluded four participants based on inconsistent responses in the filler trials (more than 50% responses not in line with the adjective used in the filler statements). ***Figure 3*** shows the mean responses by adjective polarity and context condition.

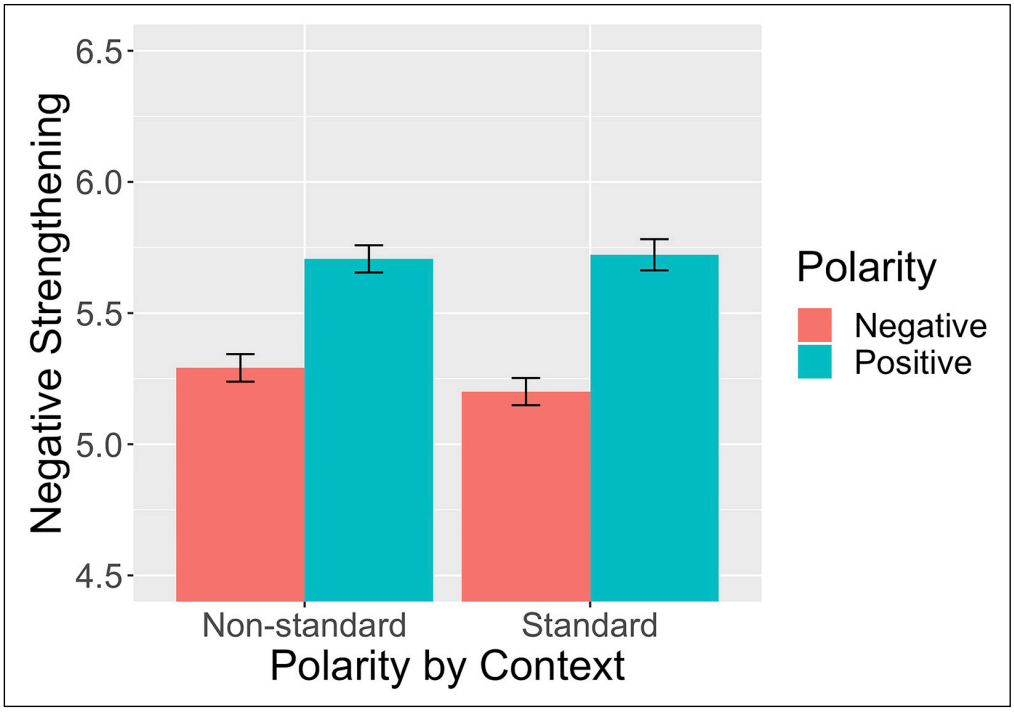


**Figure 3** Mean degree of negative strengthening by Polarity and Context (Experiment 2). Error bars represent +/– 1 SEM.

All results were analysed with cumulative link mixed-effects models using the function clmm() in the ordinal package. We included the sum-coded factors polarity, context, their interactions as well as random intercepts for items and participants. The results of the model showed a main

effect of polarity with positive adjectives involving a higher degree of negative strengthening than negative ones ($B = -.47$, $SE = .035$, $z = -13.32$, $p < .001$). This finding replicates the polarity asymmetry discussed in previous work (e.g., Ruytenbeek et al., 2017): the negation of positive adjectives such as *happy* is more likely to be strengthened than the negation of negative adjectives like *sad*. There was no main effect of context, and the interaction between polarity and context was also not significant ($p = .79$). A summary of the model is presented in ***Table 4***.[9]

| | Estimate | SE | z-value | p-value |
|---|---|---|---|---|
| Polarity | −0.469497 | 0.035259 | −13.316 | 0.000000001 |
| Context | −0.006586 | 0.034092 | −0.193 | 0.847 |
| Polarity:Context | 0.014234 | 0.052891 | 0.269 | 0.788 |

**Table 4** Summary of cumulative link mixed-effects model including the sum-coded fixed effects polarity and context (Experiment 2).

Following Ruytenbeek et al. (2017), we also investigated the effect of morphological complexity on negative strengthening. In line with Ruytenbeek et al. (2017), and with the results of Experiment 1, we found a significant interaction between polarity and morphological complexity. This interaction shows that the difference between positive and negative terms is greater in morphologically-marked antonym pairs such as *happy/unhappy* than in lexical antonyms such as *happy/sad* ($B = -.15$, $SE = .03$, $z = -4.54$, $p < .001$). The details of these analyses are presented in Appendix E.

Finally, we computed a model with polarity and participant gender, as we had found such an interaction in Gotzner and Mazzarella (2020). The model again revealed a main effect of polarity. Furthermore, there was an interaction between participant gender and polarity ($B = .25$, $SE = .07$, $z = 3.7$, $p < .001$). This interaction indicates that female participants were particularly likely to strengthen negated positive adjectives, thus showing a polarity asymmetry to a greater extent. The details of this analysis can be found in Appendix E.

## 3.4 Discussion

Experiment 2 aimed at establishing the relative contribution of face-management considerations and adjectival polarity to the asymmetric interpretation of negated antonyms. The results confirmed the prediction of Hypothesis 3, that is, the hypothesis that this asymmetry is due to the polarity of the adjective, independently of its face-threatening potential. Not only does the face-management explanation of the asymmetric interpretation of negated antonyms not explain this asymmetry, but it appears not to contribute to it.

First, the results of Experiment 2 replicated those of Experiment 1. In the non-ordinary context condition, the traditional polarity asymmetry stands: negated positive antonyms were more likely to be strengthened than negated negative antonyms. Second, they confirm the robustness of Ruytenbeek et al.'s (2017) findings: the polarity asymmetry is confirmed, and the interaction between polarity and morphological complexity is replicated. Finally, they indicate a role of the sociological variable of gender in the modulation of the strength of the polarity asymmetry: this asymmetry is even stronger when we look at the interpretation provided by female participants in comparison to that provided by male ones.[10] This gender effect is consistent with results from Gotzner and Mazzarella (2020), who showed that female participants displayed a greater polarity asymmetry compared to male participants. While in Gotzner and Mazzarella's (2019) study, target utterances were embedded in ordinary contexts, the present findings suggest that this gendered-interpretative pattern may generalise across contexts of different nature.

9 To further verify our experimental manipulation, we asked an annotator (trained linguist, native English speaker and naive to the purpose of our study) to decide which contexts he considered ordinary vs. non-ordinary for the experimental material of Experiment 2. The interrater agreement between his annotation and ours was $\varkappa = 0.5$. We then did a further analysis excluding the 6 contexts for which there was a divergence between our coding and the annotator's coding. The analysis with the resulting subset of the data showed the same results as the main analysis, that is, a main effect of polarity ($p < .0001$) but no effect of context ($p = .81$), and no interaction between polarity and context ($p = 0.71$).

10 In contrast to this, Experiment 1 did not reveal any effects related to gender. Note that the sample size in Experiment in 1 was smaller compared to Experiment 2 (both in terms of the number of items and participants), which might explain the lack of significant gender effects.

## 4 General discussion

Taken together, the results of Experiment 1 and Experiment 2 provide strong evidence of an asymmetric interpretation of negated positive and negative antonyms. Crucially, this asymmetry holds across both ordinary and non-ordinary contexts. Hence, our results disconfirm the prediction of the face-management explanation of the polarity asymmetry that this asymmetry should be reversed in non-ordinary contexts (Brown & Levinson 1987: 264–265)

Shall we conclude that face-management considerations are irrelevant when interpreting negated antonyms? This conclusion appears to be too strong. In Gotzner and Mazzarella (2020), we demonstrated that the social context influences negative strengthening. In a series of studies, we showed that negative strengthening is affected by the interaction of the sociological variables of gender, power and social distance. More specifically, we found evidence of a stronger polarity asymmetry for female than male participants and different weightings of power and social distance across gender. While these results were not predicted by the traditional account of the polarity asymmetry of negative strengthening (based on Horn 1989 and Brown & Levinson 1987), they do suggest that negative strengthening, including the strength of its polarity asymmetry, is modulated by features of the social context and they are thus indicative of the potential relevance of face-management to the interpretation of negated antonyms.

In light of the results of the present study, proponents of the face-management explanation of the polarity asymmetry of negative strengthening could appeal to a *diachronic* perspective.[11] It could be argued, for instance, that face-management concerns may well be the motivator of the polarity asymmetry in ordinary contexts, without being operative in non-ordinary ones. One could imagine face-management concerns to be at the origin of the asymmetric interpretation of negated positive and negative antonyms: the face-threatening potential of negative adjectives in usual, everyday, contexts would make speakers likely to avoid them in favour of weaker, more polite, formulations (e.g., to use *X is not interesting* to convey that 'X is uninteresting'), and hearers to recover the intended strengthened meaning when interpreting negated positive antonyms (e.g., to derive 'X is uninteresting' from *X is not interesting)*. Crucially, though, given the *ordinary* nature of these contexts, this pragmatically motivated pattern of interpretation could become *conventionally* associated with the use of negated positive adjectives. The existence of such a pragmatic convention would result in this interpretative pattern carrying over to non-ordinary contexts, regardless of the actual face-threatening potential of the adjective in context. That is, as there are not many contexts in which positive adjectives are face-threatening, interpreters could overgeneralise the pragmatic convention of strengthening negation, which would be reinforced by the regularity and frequency of ordinary contexts. This version of the face-management explanation of the polarity asymmetry of negative strengthening is in principle compatible with our data: first, it is consistent with the main effect of adjectival polarity; second, it is compatible with the gender effect observed in Experiment 2. Indeed, Horn maintains that when pragmatic conventions are in place, "we expect to find differences between speakers and between languages as to just which conventions of usage are operative" (Horn 1989: 344). Variation across gender would thus be captured by the relative weight accorded to this convention (or its salience in a given social context) by female or male interlocutors. Previous research on face-management and gender shows that women are often more prone to produce and interpret language in line with normative expectations about polite interactions (for a review, see Chalupnik, Christie & Mullany 2017; Eckert & McConnell-Ginet 2013). It follows that a pragmatic convention that is rooted in face-management concerns might be more likely to be displayed by female than male participants. It is worth noting, however, that while Horn recognises that there are "pockets of conventionalisation" (Horn 1989: 353) related to the interpretation of negated adjectives, the nature of the inference leading to a contrary reading is assumed to be that of an "online-pragmatic strengthening" (Horn 2017: 153). Future research should address the question of the nature of negative strengthening by appealing to psycholinguistic methods suitable to investigate on-line processing. Crucially, though, even if face-management considerations might motivate negative strengthening within a diachronic perspective, the present study provides clear evidence that adjectival polarity, and not face-management, is responsible for the asymmetric interpretation of negated antonyms.

11 We thank an anonymous reviewer for drawing our attention to this possibility.

In the following, we consider two alternative explanations of this polarity asymmetry that we believe could be fruitfully explored in future research. A first alternative explanation, outlined by Ruytenbeek et al. (2017: 7), focuses on the relative complexity of positive and negative adjectives and relies on the hypothesis that negative adjectives are intrinsically more complex than their positive antonyms (*Negative Adjective Complexity Hypothesis*, or *NACH;* psycholinguistic studies also indicate that positive adjectives are processed and encoded more readily compared to their negative counterparts, see especially Clark 1969; 1976). According to NACH (Büring 2007a; b), negative adjectives combine a negative morpheme with the corresponding positive antonym (whether or not this morpheme is realised, as in *unhappy*, or only abstract, as in *sad*). For this reason, the negation of a negative antonym is always more complex than the negation of a positive antonym (as complexity is introduced not only by sentential negation but also by adjectival complexity). Crucially for our purposes, it follows that the difference in complexity between the negation of a negative antonym (*not sad*) and its corresponding antonym (*happy*) is greater than the difference in complexity between the negation of the positive antonym (*not happy*) and its corresponding antonym (*sad*). The difference in complexity is a crucial factor in the *division of pragmatic labour* advocated by Horn (1991: 85): "There is an R-motivated correlation between the stylistic naturalness of a given form, its relative brevity and simplicity, and its use in stereotypic situations. The corresponding periphrastic forms, stylistically more complex, are correspondingly Q-restricted to those situations outside the stereotype, for which the unmarked expression could have been used appropriately". Based on the division of pragmatic labour, if we consider a given antonymic pair (*happy* and *sad*) and its euphemistic alternatives (respectively, *not sad* and *not happy*), and we assume NACH, it follows that *not sad* should be mapped onto instances of happiness that are even less stereotypical than the instances of sadness that are targeted by *not happy* (see Ruytenbeek et al. 2017 for an application of the same reasoning to Krifka's 2007 account of negated antonyms). The division of pragmatic labour, coupled with NACH, thus provides an alternative explanation of the polarity asymmetry in the interpretation of negated adjectives.

A second alternative explanation focuses on the valence of the adjectival lexical meaning. This explanation, whose development is inspired by Terkourafi, Weissman and Roy (2020), appeals to the positivity bias of human languages. Since the seminal work of Boucher and Osgood (1969), the *Pollyanna Hypothesis*, that is, the hypothesis of a "universal human tendency to use evaluatively positive (E+) words more frequently, diversely and facilely than evaluatively negative (E–) words", has received robust confirmation via big data studies (see Dodds et al. 2015). The Pollyanna Hypothesis might explain the polarity asymmetry in the interpretation of negated antonyms in the following way. Speakers may be more likely to employ evaluatively positive words, rather than their antonyms, to express all sorts of judgements (judgements of the form *X is E+* but also judgements of the form *X is not E+*). For this reason, speakers may be more likely to express a negative judgement by *X is not E+* rather than *X is E–* even when they target the semantic space covered by *E–*. By recognising this tendency, addressees may thus be more likely to strengthen the negation of a positive antonym (*X is not E+*) to convey the affirmation of the corresponding negative antonym, as the formulation *X is E–* is dispreferred. Insofar as this interpretative pattern is based on a positivity bias for evaluatively positive *words*, it may carry on independently from the words' occasion-specific face-threatening potential. That is, even if in non-ordinary contexts the predication of a negative antonym (e.g., *He is mean* as referred to the addressee's competitor) is not face-threatening, the negative valence of the lexical meaning may make it a dispreferred option as compared to the negation of the positive antonym (*He is not kind*), thus encouraging the negative strengthening of *X is not E+*. According to this hypothesis, the positive valence of the lexical meaning would thus override the effect of social context, and make the polarity asymmetry arise across both ordinary and non-ordinary contexts.[12]

## 5 Conclusion

The present study investigated the face-management explanation of the asymmetric interpretation of negated antonyms by testing one of its direct predictions (Brown &

12 An anonymous reviewer rightly pointed out that the Pollyanna hypothesis alone cannot explain why weak positive adjectives are more likely to be strengthened than strong positive adjectives. To do this, it would need to combine with considerations about adjectival markedness and/or frequency.

Levinson 1987: 264–265). According to this explanation, the negation of positive antonyms should be more likely to be strengthened than the negation of negative antonyms, but only in *ordinary contexts*, that is, in contexts in which the predication of the negative adjective represents a face-threat to the addressee. If the face-threatening potential of polar opposites is reversed, so should be the polarity asymmetry of negative strengthening. For this reason, we set out to investigate the interpretation of negated antonyms in *non-ordinary contexts*, in which positive adjectives display a face-threatening potential (Experiment 1). Our results disconfirm the face-management explanation of the polarity asymmetry of negative strengthening: the asymmetry does not appear to be due to the face-threatening potential of the adjectives in context (above and beyond their polarity). Furthermore, Experiment 2 suggests that face-management considerations do not even contribute to the strength of the polarity asymmetry. Taken together, Experiments 1 and 2 strongly suggest that the asymmetric interpretation of negated antonyms is due to adjectival polarity *per se*, independently of the face-threatening potential of the members of the antonymic pair in a given context.

We have discussed three further explanations of the polarity asymmetry of negative strengthening. The first explanation appeals to the existence of a pragmatic convention, originally based on face-management considerations, which would apply to the interpretation of negated adjectives independently of their actual face-threatening potential in context. The second explanation is based on the complexity of alternative expressions (Krifka 2007; Ruytenbeek et al. 2017). The third explanation relies on the Pollyanna Hypothesis – the view that evaluatively positive words are used more frequently than evaluatively negative words (Boucher & Osgood 1969; Terkourafi et al. 2020). Developing and teasing apart these (or further) alternative explanations of the asymmetric interpretation of positive and negative adjectives is a promising (certainly not uninteresting) question for future experimental research.

## Additional file

The additional file for this article can be found as follows:

- **Supplementary file 1.** Appendices. DOI: *https://doi.org/10.5334/gjgl.1342.s1*

## Funding information

This work was supported by the German Research Foundation (DFG) as part of the *Xprag.de* Initiative (Grant Nr. BE 4348/4-1) as well as an Emmy Noether grant awarded to NG (Grant Nr. GO 3378/1-1).

## Acknowledgements

We thank the audience of the DegPol workshop at the Leibniz-ZAS (Berlin) for useful discussion of this work. We are grateful to Rick Nouwen for providing us with the idea underlying Experiment 2. We also thank Stanley Donahoo for annotating our stimuli and the three anonymous reviewers of *Glossa* for their precious feedback.

## Competing interests

The authors have no competing interests to declare.

## Author affiliations

**Diana Mazzarella** *orcid.org/0000-0002-7650-7196*
Université de Neuchâtel, Rue Pierre-à-Mazel 7, 2000 Neuchâtel, CH

**Nicole Gotzner** *orcid.org/0000-0002-9584-4518*
Universität Potsdam, Karl-Liebknecht-Straße 24-25, 14476 Potsdam, DE

## References

Boucher, Jerry & Charles E. Osgood. 1969. The pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behaviour* 8(1). 1–8. DOI: https://doi.org/10.1016/S0022-5371(69)80002-2

Brown, Penelope & Stephen Levinson. 1987. *Politeness: Some universals in language usage*. Cambridge & New York: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511813085

Büring, Daniel. 2007a. More or less. *Proceedings of the Chicago Linguistics Society (CLS)* 43(2). 3–17.

Büring, Daniel. 2007b. Cross-polar normalies. *Proceedings of Semantics and Linguistic Theory (SALT)* 17. 37–52. DOI: https://doi.org/10.3765/salt.v17i0.2957

Chalupnik, Malgorzata, Christine Christie & Louise Mullany. 2017. (Im)politeness and gender. In Jonathan Culpeper, Michael Haugh & Dániel Z. Kádár (eds.), *The Palgrave Handbook of Linguistic (Im) politeness*, 517–537. London: Palgrave. DOI: https://doi.org/10.1057/978-1-137-37508-7_20

Christensen, Rune Haubo B. 2018. Cumulative link models for ordinal regression with the R package ordinal. Retrieved from: https://rdrr.io/cran/ordinal/f/inst/doc/clm_article.pdf.

Clark, Herbert H. 1969. Linguistic processes in deductive reasoning. *Psychological Review* 76. 387–404. DOI: https://doi.org/10.1037/h0027578

Clark, Herbert H. 1976. *Semantics and Comprehension*. De Gruyter: Mouton. DOI: https://doi.org/10.1515/9783110871029

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). 237–256. DOI: https://doi.org/10.1080/01638539909545083

Cruse, David. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.

Dodds, Peter Sheridan, Eric M. Clark, Suma Desu, Morgan R. Frank, Andrew J. Reagan, Jake Ryland Williams, Lewis Michell, Kameron Decker Harris, Isabel M. Kloumann, James P. Bagrow, Karine Megerdoomian, Matthew T. McMahon, Brian F. Tivnan & Christopher M. Danforth. 2015. Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences* 112(8). 2389–2394. DOI: https://doi.org/10.1073/pnas.1411678112

Ducrot, Oswald. 1973. *La preuve et le dire.* Paris: Maison Mame.

Dulcinati, Giulio. 2018. Cooperation and pragmatic inferences. London, United Kingdom: University College London dissertation.

Eckert, Penelope & Sally McConnell-Ginet. 2013. *Language and gender*. New York: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139245883

Fraenkel, Tamar & Yaacov Schul. 2008. The meaning of negated adjectives. *Intercultural Pragmatics* 5(4). 517–540. DOI: https://doi.org/10.1515/IPRG.2008.025

Giora, Rachel, Noga Balaban, Ofer Fein & Inbar Alkabets. 2005. Negation as positivity in disguise. In Herbert Colston & Albert Katz (eds.), *Figurative language comprehension: Social and cultural influences*, 233–258. Hillsdale, NJ: Erlbaum.

Givón, Talmy. 1970. Notes on the semantic structure of English adjectives. *Language* 46. 816–837. DOI: https://doi.org/10.2307/412258

Goffman, Erving. 1967. *Interaction ritual. Essays on face-to-face behavior*. New York: Pantheon Books.

Gotzner, Nicole & Diana Mazzarella. 2020. Face management and negative strengthening: The role of power relations, social distance and gender. DOI: https://doi.org/10.31234/osf.io/w5fyb

Gotzner, Nicole, Stephanie Solt & Anton Benz. 2018. Scalar diversity, negative strengthening, and adjectival semantics. *Frontiers in psychology* 9. 1659. DOI: https://doi.org/10.3389/fpsyg.2018.01659

Gotzner, Nicole & Sybille Kiziltan. forthcoming. She is brilliant! Distinguishing different readings of relative gradable adjectives.

Holtgraves, Thomas R. 1986. Language Structure in Social Interaction. Perceptions of Direct and Indirect Speech Acts and Interactants Who Use Them. *Journal of Personality and Social Psychology* 51(2). 305–314. DOI: https://doi.org/10.1037/0022-3514.51.2.305

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

Horn, Laurence R. 1991. Duplex negatio affirmat…: the economy of double negation. In Lise M. Dobrin, Lynn Nichols & Rosa M. Rodriguez (eds.), *Regional Meeting of the Chicago Linguistic Society. Part Two*, 27. 80–106.

Horn, Laurence R. 2017. The singular square: Contrariety and double negation from Aristotle to Homer. In Joanna Blochowiak, Cristina Grisot, Stephanie Durrleman & Christopher Laenzliger (eds.), *Formal Models in the Study of Language*, 143–179. Cham: Springer. DOI: https://doi.org/10.1007/978-3-319-48832-5_9

Horn, Laurence R. 2020. Negation and opposition: Contradiction and contrariety in logic and language. In Vivian Déprez & Maria Teresa Espinal (eds.), *The Oxford Handbook of Negation*, 7–25. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780198830528.013.1

Israel, Michael. 2004. The pragmatics of polarity. In Laurence R. Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 701–723. Oxford: Blackwell. DOI: https://doi.org/10.1002/9780470756959.ch31

Kerr, Norbert L. 1998. HARKing: Hypothesising after the results are known. *Personality and social psychology review* 2(3). 196–217. DOI: https://doi.org/10.1207/s15327957pspr0203_4

Krifka, Manfred. 2007. Negated antonyms: creating and filling the gap. In Uli Sauerland (ed.), *Presupposition and implicature in compositional semantics* (Palgrave studies in pragmatics, language and cognition series), 163–177. Basingstoke: Palgrave Macmillan. DOI: https://doi.org/10.1057/9780230210752_6

Leech, Geoffrey. 2014. The Pragmatics of Politeness. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780195341386.001.0001

Lehrer, Adrienne & Keith Lehrer. 1982. Antonymy. *Linguistics and philosophy* 5. 483–501. DOI: https://doi.org/10.1007/BF00355584

Mazzarella, Diana, Robert Reinecke, Ira Noveck & Hugo Mercier. 2018. Saying, presupposing and implicating: How pragmatics modulates commitment. *Journal of Pragmatics* 133. 15–27. DOI: https://doi.org/10.1016/j.pragma.2018.05.009

Neuhaus, Laura. 2016. Four potential meanings of double negation: The pragmatics of nicht un-constructions. *International Review of Pragmatics* 8(1). 55–81. DOI: https://doi.org/10.1163/18773109-00702500

Rett, Jessica. 2015. *The semantics of evaluativity*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199602476.001.0001

Ruytenbeek, Nicolas, Steven Verheyen & Benjamin Spector. 2017. Asymmetric inference towards the antonym: Experiments into the polarity and morphology of negated adjectives. *Glossa: A Journal of General Linguistics* 2(1). 92. DOI: https://doi.org/10.5334/gjgl.151

Sapir, Edward. 1944. Grading: A study in semantics. Reprinted in Pierre Swiggers (ed.), *The Collected Works of Edward Sapir,* 447–70. Berlin: de Gruyter, 2008.

Terkourafi, Marina, Benjamin Weissman & Joseph Roy. 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

Vendler, Zeno. 1963. The transformational grammar of English adjectives. University of Pennsylvania.

Vullioud, Colin, Fabrice Clément, Thom Scott-Phillips & Hugo Mercier. 2017. Confidence as an expression of commitment: Why misplaced expressions of confidence backfire. *Evolution and Human Behavior* 38(1). 9–17. DOI: https://doi.org/10.1016/j.evolhumbehav.2016.06.002

TO CITE THIS ARTICLE:
Mazzarella, Diana and 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. 1–17. DOI: https://doi.org/10.5334/gjgl.1342

Submitted: 10 June 2020
Accepted: 22 November 2020
Published: 12 April 2021



*Glossa: a journal of general linguistics* is a peer-reviewed open access journal published by Ubiquity Press.