

Negative Strengthening and Speaker Adaptation in Metalinguistic Disagreements

Introduction: Metalinguistic disagreements are used to debate the usage of underspecified terms, rather than facts about the world [1]. Prior work shows that listeners revise their interpretations of vague or imprecise predicates when challenged by a metalinguistic denial (e.g., S1: *Many of the kids were late.*; S2: *I disagree, I don't think many of the kids were late.*) [2,3]. Additionally, listeners have been shown to track interspeaker differences in the usage of these predicates [4,5]. However, the work examining speaker adaptation only investigates positive assertions, so it remains unclear how listeners track speaker meaning in metalinguistic denials. This is non-trivial, as the interpretation of some negated positive vague predicates strengthens to the antonym (*not big* \rightsquigarrow *small*), while their negated antonym does not (*not small* $\not\rightsquigarrow$ *big*; [6,7]). This is a phenomenon known as asymmetric negative strengthening (ANS). The strength of this asymmetry varies even across gradable adjectives, and prior work has only looked at descriptive negation, so it remains open whether *many/few* undergo ANS and if metalinguistic, as opposed to descriptive, negation modulates this effect. In two experiments, we investigate speaker adaptation to vague predicates within metalinguistic disagreements and listeners' interpretations of negated vague quantifiers in metalinguistic and descriptive contexts. Specifically, we ask if listeners can track these preferences presented in disagreements as a positive assertion (**Q1**), and a negative assertion (**Q2**) and if interpretations between a positive/negative assertion differ (**Q3**). We find that participants can track preferences presented in a disagreement, both as positive and negative assertions, but negative assertions are interpreted as symmetrically weaker than their positive counterparts.

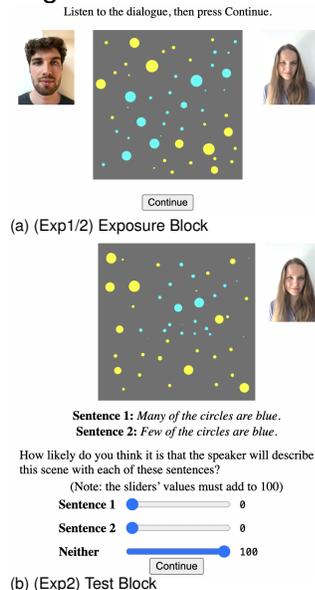


Fig. 1: Example trials in Exp1/2

Results: Q2: See Fig.3. Following [4,5] we analyzed *many/few* ratings by fitting natural splines to participant ratings (excluding *neither* as we expect no difference across S1 biases). We then calculated the AUC difference ($AUC_{Many} - AUC_{Few}$) for each participant. A Welch's two-sample t-test across S1 bias found that participants exposed to the *many*-biased S1 gave significantly higher probabilities that S2 would use *few* compared to those exposed to the *few*-biased S1, both when S2 responded with the antonym ($t = 2.17, df = 18.91, p = 0.043$)—supporting the results of Exp1—and when S2 responded with metalinguistic negation ($t = 2.41, df = 23.47, p = 0.024$), demonstrating that listeners can track speaker preferences presented as a negative assertion.

Exp1: (n=60) In Exp1 (addressing Q1), participants viewed displays of 50 blue/yellow circles. In the exposure trials, Speaker 1 (S1) was to the left, and Speaker 2 (S2) was to the right of the stimuli (Fig.1a). Alongside the visuals were auditory stimuli of metalinguistic disagreements where S1 asserted {*Many/Few*} of the circles are blue and S2 disagreed with *I don't think {many/few} of the circles are blue*. Participants were assigned to a *many*- or *few*-biased condition based on S1's preference at 50% target color (Table 1). In subsequent test trials, participants heard S1's original utterances with the quantifier masked by static and indicated whether S1 had used *many* or *few* to describe the images, ranging from 20-80% of either color. **Results/Discussion:** (Fig.2) Using a spline-based AUC analysis (R package MESS), we found participants exposed to the *many*-biased speaker accepted *many* at a significantly higher rate than the *few*-biased group ($t = 2.56, df = 56.65, p < 0.001$), demonstrating that listeners can track speaker preferences despite conflicting inputs.

Exp2: (n=60) To address Q2/Q3, Exp2 extended Exp1's exposure design with two between-subject groups alongside Negation Disagreement (Neg): Antonym Disagreement (Ant; S2 responds w/antonym) and Question Response (QR; S1 asks question, S2 responds w/negation) (see Table 2 for example dialogues). Test trials presented the original images but asked about S2 instead of S1. Participants used sliders to allocate probability (summing to 100%) that S2 would use *many*, *few* or neither option to describe the image (Fig.1b).

S1 Quantifier:	Many		Few		All		None	
	<i>n</i>	<i>P</i>	<i>n</i>	<i>P</i>	<i>n</i>	<i>P</i>	<i>n</i>	<i>P</i>
<i>Many</i> -biased	10	50%	5	20%	5	100%	—	—
<i>Few</i> -biased	5	80%	10	50%	—	—	5	0%

Table 1: (Exp 1/2)# exposure trials (*n*) per quantifier associated % target color circles (*P*) in *many/few* biased speaker conditions. Critical trials bold.

Condition	Dialogue
Negation	S1: { <i>Many/few</i> } of the circles are blue.
Disagreement (Neg)	S2: I disagree, I don't think { <i>many/few</i> } of the circles are blue.
Antonymy	S1: { <i>Many/few</i> } of the circles are blue.
Disagreement (Ant)	S2: I disagree, I think { <i>few/many</i> } of the circles are blue.
Question	S1: Do you think { <i>many/few</i> } of the circles are blue?
Response (QR)	S2: No, I don't think { <i>many/few</i> } of the circles are blue.

Table 2: (Exp2) Critical trial example dialogues. Colors correspond to dialogue pairs (*many/few* S1 bias)

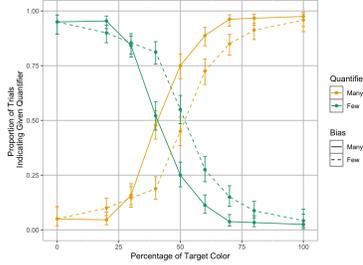


Fig. 2: (Exp1) Prop of trials indicating S1 used *many/few* at the different percentages by Bias

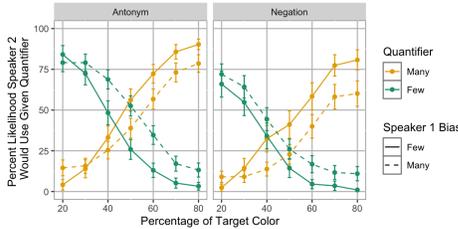


Fig. 3: (Exp2) % likelihood of S2 using *many/few* at different percentages for Disagreement Conditions (controlling for Quantifier and Target Color; random participant intercepts). We found no significant interaction ($t = -0.825, df = 51, p = 0.408$). This indicates that the difference in interpretation of the negative assertion compared to its antonym was consistent, regardless of whether the negated quantifier was *many* or *few*. Finally, we tested whether the symmetric negative strengthening is due to the nature of the quantifiers or the metalinguistic disagreement. We compared response rates for the *many*, *few*, and *neither* responses separately across the Neg and the QR conditions (Fig.5). We found no significant differences between the two contexts across any response category. This indicates that the symmetric interpretations of the negated quantifiers are due to the nature of the quantifiers and not simply the metalinguistic disagreement.

Discussion and Conclusion: From our two experiments, we conclude that listeners can track speaker preferences when expressed as both positive and negative assertions within a disagreement. We also conclude that listeners do not have asymmetric interpretations of negated quantifiers, as both negated *many* and negated *few* were interpreted as equally weak relative to their antonymic counterparts across both the metalinguistic disagreement and the question response condition. Prior literature indicates that there exists diversity in the strength of ANS across adjectival antonym pairs; the current studies show this variability extends to other lexical items, meaning that vagueness alone is not a sufficient requirement for ANS. Future work will examine the factors that influence ANS strength within and across lexical categories.

Q3: Next, we compared Neg vs. Ant conditions (Fig.4). Unlike across S1 biases, we anticipate a difference in the *neither* category across S2 responses. Specifically, we expect higher rates of *neither* and lower rates of *many/few* in the Neg condition, as participants may interpret the negated response to S1's statement as the speaker preferring neither *many* nor *few*. We calculated the AUC for all three response categories, then computed a composite score: $(AUC_{Many} + AUC_{Few}) - AUC_{Neither}$, where lower scores indicate increased reliance on *neither*. A Welch's t-test confirmed that the Neg condition had significantly lower scores than the Ant condition, both when S1 was *many*-biased ($t = 3.81, df = 17.93, p < 0.01$) and when S1 was *few*-biased ($t = 2.15, df = 15.30, p < 0.05$). This indicates participants in the Neg condition relied more heavily on *neither* as a response, meaning the negated propositions were interpreted as distinct from their positive counterparts. Given that the negated propositions were interpreted as weaker than the positive antonyms, we assessed if this difference was symmetric (i.e., if one quantifier strengthened more than the other). To do this, we fit a linear mixed-effects model on just the Neg and Ant conditions with fixed effects for S1 Bias, S2 Response Type, and their interaction (controlling for Quantifier and Target Color; random participant intercepts). We found no significant interaction ($t = -0.825, df = 51, p = 0.408$). This indicates that the difference in interpretation of the negative assertion compared to its antonym was consistent, regardless of whether the negated quantifier was *many* or *few*. Finally, we tested whether the symmetric negative strengthening is due to the nature of the quantifiers or the metalinguistic disagreement. We compared response rates for the *many*, *few*, and *neither* responses separately across the Neg and the QR conditions (Fig.5). We found no significant differences between the two contexts across any response category. This indicates that the symmetric interpretations of the negated quantifiers are due to the nature of the quantifiers and not simply the metalinguistic disagreement.

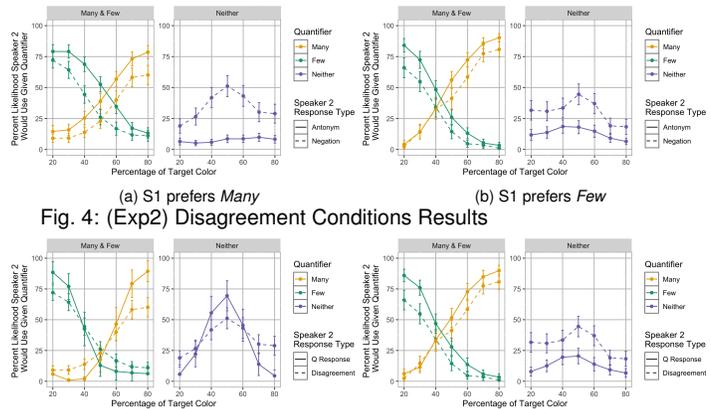


Fig. 4: (Exp2) Disagreement Conditions Results

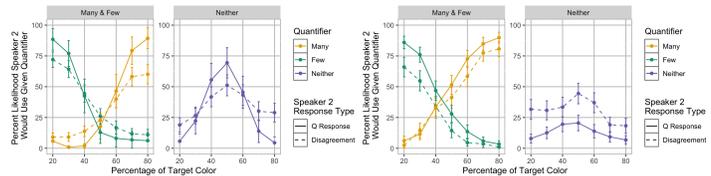


Fig. 5: (Exp2) Neg. Disagreement and Question Response Results

[1] Barker (2013). *Inquiry*. [2] Wu & Aparicio (2025). *ELM*. [3] Pecsok & Aparicio (2024). *CogSci*. [4] Yildirim et al. (2016). *JML*. [5] Schuster & Degen (2020). *Cognition*. [6] Horn (1989). *A Natural History of Negation*. [7] Ruytenbeek et al. (2017). *Glossa*.