Predicting the discourse dynamics of imprecision from adjectival scale structure

Introduction. Speakers often make imprecise utterances that, strictly speaking, are false (1a, Fig. 2). The *standard of precision* (SoP) [1: Lewis, 1979; 2: Lasersohn, 1999; 3: Lauer, 2012; 4: Leffel et al., 2016; 5: Klecha, 2018; 6: Beltrama & Schwarz, 2021] governing a discourse can be negotiated through metalinguistic disagreements (MDs) such as (1), Fig. 2, where a rejection (1b) increases the SoP. **The puzzle.** Lewis [1] claims that the SoP can be raised, but not lowered, by MD; Klecha [5] argues that lowering the SoP requires an explicit metalinguistic negotiation; and Lauer [3] attributes the purported asymmetry to the perceived risk of a later requirement for retraction. This standard view has difficulty capturing the fact that MDs can update the SoP bidirectionally in an experimental setting [7: Wu & Aparicio, 2025]—a pattern typically associated with relative adjectives, whose standard is context-dependent [8: Barker, 2002; 9: Kennedy, 2007]. However, [7] observed an asymmetry: while imprecise speakers readily abandon the lower SoP after a MD, precise speakers are less willing to discard their own SoP, consistent with [1, 3, 5]. The coexistence of bidirectional update and the associated asymmetry remains unexplained. **Contribution.** We argue that both generalizations follow from (i) the scale structure of absolute adjectives, together with (ii) probabilistic reasoning about likely SoPs. We obtained data from [7] and developed two models of discourse update: one encoding the upper-bounded semantics of absolute adjectives, and another encoding the open-scale semantics of relative adjectives. By comparing these models in terms of expected log predictive density (ELPD), we find that the Absolute model yields a substantially better fit, offering preliminary evidence that experimental participants' ability to reason about scale structure drives the discourse dynamics of imprecision. In short: the scale structure of absolute adjectives biases the probability mass toward the endpoint of the scale, yielding a qualitative asymmetry, while permitting enough flexibility to yield constrained bidirectionality.

Data. We model the experimental task of [7] as a discourse, following [10: Grove & White, 2025a; 11: Grove & White 2025b] . [7] designed 24 five-point scales featuring maximum standard adjectives (e.g., *empty bottle*, Fig. 3). [7] used S4 as the critical scale point, since it is compatible with an imprecise interpretation; S5 as a control in which the predicate applies; and S1 as a filler. Participants were instructed to engage in a conversation with a remote human participant who saw the same items; in reality, they interacted with a chatbot (*bot*) with predetermined SoP preferences: an *imprecise bot* (Exp1) accepting the predication at S4, and a *precise bot* (Exp2) that rejected it, with both bots accepting S5 and rejecting S1 (Fig. 4). The discourse proceeded in three phases, which we map to particular discourse moves. (1) *Initial Utterance:* in Part 1 (P1; [7]'s Block 2), participants initiated the dialogue by selecting one of the choices in Fig. 5a. We model this move as sampling a value from the adjective's distribution over possible standard thresholds; if the object's intensity exceeds the sampled standard, the predication is accepted. Subsequently, the bot agreed or disagreed, according to its own preferences. Thus at S4, MDs arose for strict participants who rejected S4 in Exp1 and imprecise participants who accepted S4 in Exp2 (Fig. 5a). (2) *Dialogue Perception:* in Part 2 (P2), participants judged whether their own or their interlocutor's assertion was correct by selecting one of the three choices in Fig. 5b. (3) *Post-MD:* in Part 3 (P3; [7]'s Block 3), participants re-evaluated the predication, following the MD (Fig. 5c). We model this move as *resampling* the adjectival threshold.

Modeling. We assume that to apply a predicate to an object with intensity (e.g., *emptiness*) c, a speaker samples θ from a probability distribution over thresholds determined by the adjective's scale structure, and that the probability of acceptance p_i is the probability that $c \geq \theta$. We test two competing hypotheses about the shape of this threshold distribution. ***Absolute model (Abs):*** this model encodes an upper-closed scale [9], such that standard thresholds below the endpoint may only arise due to pragmatic slack [2]. We encode this slack (i.e., the *pragmatic halo* [2]) in terms of a normal distribution truncated at the upper endpoint (see Fig. 1), and we model the probability of accepting the predication p_i as the cumulative distribution function (CDF) of this distribution up to the object's value c. As shown in the left panel of Fig. 1, deviations away from the endpoint (i.e., *lowering the SoP*) extend into the distribution's tail, while movements toward the endpoint (i.e., *raising the SoP*) face less resistance. ***Relative Model (Rel):*** this model encodes an open scale without any endpoints [9], so that the standard threshold takes a normal distribution. As shown in the right panel of Fig. 1, the resulting distribution's mass is concentrated around its mean; thus, shifting the standard away from the mean gives rise to equal resistance in both directions. We link the resulting theoretical probabilities to responses in a Bayesian model with ordered logit likelihoods: selections accepting the predication (i.e., '*This bottle is empty*' in P1/P3, '*Only the other participant is right*' in P2, Exp1 and '*Only I am right*' in P2, Exp2) are coded as 3; selections tolerating either acceptance or rejection (i.e., '*Both descriptions work*' in P1, '*Either A or B*' in P3, and '*Both of us can be right*' in P2) are coded as 2; and selections rejecting the predication (i.e., '*This bottle is not empty*' in P1/P3; '*Only I am right*' in P2, Exp1 and '*Only the other participant is right*' P2, in Exp2) are coded as 1. We predict the responses in P1 and P2 in terms of the CDF of the distribution over standard thresholds (θ), up to the relevant item's intensity (c), translated to a log-odds scale. We model the P3 (the post-MD trial) responses as arising from participants' decision to *resample* the standard threshold from the distribution associated with the adjective: participants abandon their previous threshold in response to disagreement, using the same distribution.

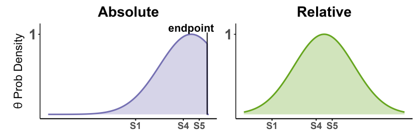


Fig. 1: Models.

Results. We compare the ELPD differences (s.e. in parentheses) of the **Abs** and **Rel** models fit to the full dataset and the P1 initial utterance subset in Table 1. **Abs** outperforms **Rel** on P1, a comparison strengthened on the full dataset, indicating that **Abs** better explains both initial judgments of the adjective's application and the (biased) bidirectional update observed by [7].

	Full dataset	P1
Rel vs **Abs**	-12.3 (4.5)	-8.9 (4.0)

Table 1: Model comparison results.

Discussion. By relating the discourse dynamics of imprecision to adjectival scale structure, we find preliminary evidence for a unified explanation of both the classical observation about the unidirectionality of precisification [1, 5] (i.e., MDs can successfully raise the SoP but cannot effectively lower it) and recent experimental evidence for bidirectional updates [7] (i.e., first-person MDs update the SoP both upward and downward). The **Abs** model concentrates probability mass near the scale endpoint, rendering precise standards difficult to abandon, while the distribution's tail permits sufficient probabilistic slack. In ongoing work, we aim to acquire parallel data on relative adjectives, in order to validate whether open scales yield the unconstrained bidirectional update patterns predicted by the **Rel** model.

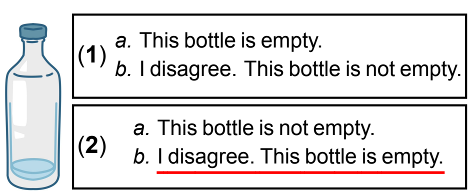


Fig. 2: Metalinguistic disagreements.

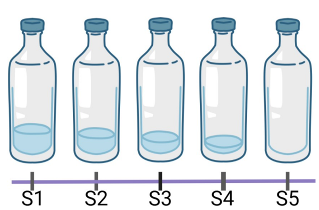


Fig. 3: Scalepoints.

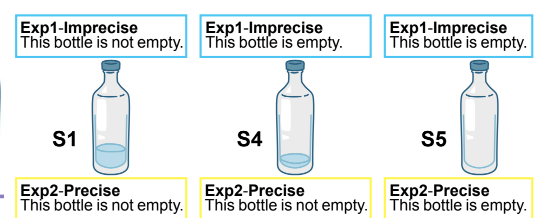
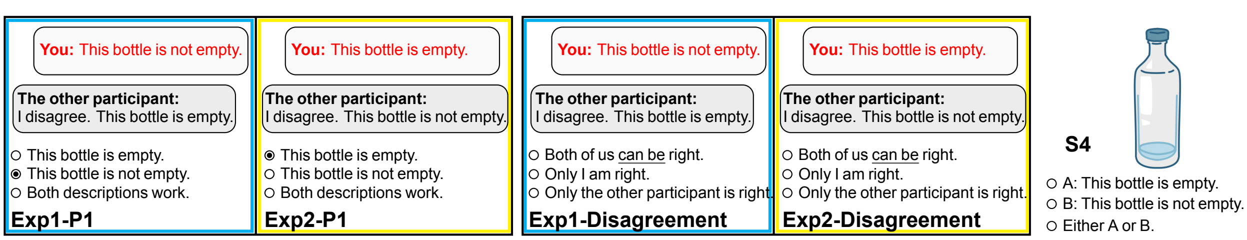


Fig. 4: Bot preferences.



(a) P1 *Initial Utterance* example. (b) P2 *Dialogue Perception* example. (c) P3 *Post-MD*.

Fig. 5: Experiment Block2-3 item example.