The lying/misleading distinction from the viewpoint of truth evaluators

Background. The two dominant definitions of Lying face a challenge when attempting to distinguish between lying and misleading claims. According to the traditional view, lying involves false *explicit* content, whereas misleading claims involve false *implicated* content (1,7, but see 2). Recent empirical studies, however, indicate that speakers can be perceived as lying even when the believed-false content is implicated (e.g., 3). According to a more recent view, speakers are lying if they are perceived as committed to the false communicated content (5, 8). While this view effectively accounts for the empirical evidence, it seems to obscure the boundary between lies and misleading claims. That is to say, in a commitment-based framework for lying and given that commitments are measured on a continuum, it is unclear what constitutes the boundary between lies and misleading claims (as even a merely misleading claim involves some degree of commitment).

Objectives. This study aims to draw a clearer distinction between lies and misleading claims. To achieve this, I address a methodological issue that pertains to the inconsistency within the empirical literature regarding the means by which the difference between lying and misleading claims is gauged. Furthermore, I address the data through the oft-overlooked role of the truth evaluators. By doing so, I aim to highlight a conceptual difference between lying and misleading that pertains to the mindset of those tasked with their evaluation.

The current study. This study examines participants' disposition when asked to truth evaluate a literally true statement that indirectly conveys a believed-false claim. In this study, indirectness was manipulated using explicit, rather than implicated, expressions through the epistemic modals, *possible* and *possibly*, which express uncertainty. Participants were introduced to a context story about a party. Following the story, targets appeared in one of four conditions, see Table 1.

Table 1 – Experimental procedure

Both Aron and Dan like Jill, and they both know this about each other. One day, Aron invited Jill to a party, and she agreed. They danced all night and had a lot of fun. The day after the party, Aron met Dan and told him about the party. Dan immediately asked him if Jill came to the party too.	**Context story**, similar for all participants
Aron said: "It is possible that Jill came to the party."	**Target – Condition 1**: literally true *p* + possible
Aron said: "Jill possibly came to the party."	**Target – Condition 2**: literally true *p* + possibly
Aron said: "Jill did not come to the party."	**Target – Condition 3**: a literally false *p*
Aron said: "Jill came to the party."	**Target – Condition 4**: a literally true *p*
Did Aron lie?	**Task**, similar for all participants
|----------| No, he did not Yes, he did	**Measurement tool**, similar for all participants

Four screening/comprehension questions followed the task. Failure in any one of these questions led to removal from the analysis.

Method. A web-based paradigm with a between-subjects design was employed in Prolific (4). 320 participants (age range: 19–45, *M* = 31.85, *SD* = 7.15, 160 females) were recruited until a sample size of ~80 participants per condition, after exclusion, was met. All participants were native US English speakers and were paid regardless of their success in comprehension questions.

Results. A summary of the means' lie ratings per condition shows that the literally false claim received a high lie rating (Cond.3; *M* = 97.3, *SD* = 11.4) and that the literally truthful claim received a low lie rating (Cond.4; *M* = 6, *SD* = 12.2)—as expected. The hedged literally true claims received an intermediate lie rating when accompanied by the objective epistemic modal *possible* (Cond.1; *M* = 50.9, *SD* = 32.6) and by the subjective epistemic modal *possibly* (Cond.2; *M* = 56.6, *SD* = 12.2).

Data was analyzed using a Bayesian Zero-One-Inflated-Beta (ZOIB). The *emeans* package was used for later pairwise comparisons (6). Because this analysis uses a Bayesian framework, it is important to note that there are no clear thresholds to determine significance. Traditionally, if the coefficient intervals do not include 0, it can be deduced with adequate confidence that a significant effect was observed. The model revealed a significant effect of condition for the question, "Did [the protagonist] lie?" It specifically showed that the Highest Posterior Density (HPD) interval of Condition 1 with Condition 3 and of Condition 1 with Condition 4 did not include 0, indicating that an objective epistemic modal with a literally true claim is considered neither a full-fledged lie nor a truthful claim—and similarly for condition 2. The HPD interval of the comparison between Condition 1 and Condition 2 included 0, indicating that the truth evaluations in the two conditions were not significantly different. Lastly, the HPD interval of Condition 3 with Condition 4 did not include 0, indicating that the evaluation of full-fledged lies differs significantly from that of truthful claims.

***Discussion*.** These findings indicate that hedging a literally true claim using epistemic modals is a misleading act. It, thus, also indicates that misleading is not restricted to implicated content. These findings, however, here and in other studies, do not directly explain the lying/misleading distinction. To do this, it is essential to adopt the truth evaluators' perspective (rather than the content's explicitness/speaker's commitment).

A closer look at the truth evaluators' behavioral patterns suggests that two distinct mindsets underlie the evaluation of different forms of deception. In misleading claims, participants are conflicted, probably by the presence of two opposing truth values. They resolve this conflict by leaning towards one of the truth-values (as evident through the bimodal distribution and its wide range). In full-fledged lies and truthful claims, truth evaluators experience no such conflict (as evident in the skewed distribution with its narrow range).

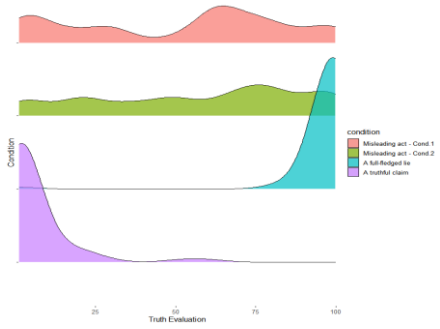

Figure 1 – The patterns of deception

***Future Directions*.** To the extent these patterns generalize, they provide insights into the mindsets of truth evaluators when evaluating different forms of deception. An ongoing experiment explores this using other stories and other modes of deception (e.g., politeness).

References: **(1)** Adler, J. E. (1997). Lying, Deceiving, or Falsely Implicating. The Journal of Philosophy, 94(9), 435–452.; **(2)** Meibauer, J. (2005). Lying and falsely implicating. Journal of Pragmatics, 37(9), 1373–1399.; **(3**) Orr, S., Ariel, M., & Peleg, O. (2017). The case of literally true propositions with false implicatures. In I. Chiluwa (Ed.), Deception and deceptive communication: Motivations, recognition techniques and behavioral control (pp. 67–107). Nova Science Publishers, Inc.; **(4)** Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. Journal of Behavioral and Experimental Finance, 17, 22–27; **(5)** Reins, L. M., & Wiegmann, A. (2021). Is Lying Bound to Commitment? Empirically Investigating Deceptive Presuppositions, Implicatures, and Actions. Cognitive Science, 45(2) e12936.; **(6)** Russell V., L. (2022). emmeans: Estimated Marginal Means, aka Least-Squares Means (R package version 1.8.3).; **(7)** Saul, J. M. (2012). Lying, Misleading, and What is Said: An Exploration in Philosophy of Language and in Ethics. Oxford University Press.; **(8)** Viebahn, E. (2021). The Lying-Misleading Distinction: A Commitment-Based Approach. The Journal of Philosophy, 118(6), 289–319.