

Implicature calculation and the pragmatics of experiments



Pranav Anand, Caroline Andrews & Matthew Wagers
Department of Linguistics • University of California, Santa Cruz



Introduction

- Experiments are underspecified social interactions
 - subjects fill in missing social context from experience-based expectations [5]
- Central Concern: How does experimental context influence participant responses?
- Investigative Approach
 - Task focus: Image verification
 - Empirical focus: Scalar implicature rate for *some* (note wide rate variation in previous work: 34% to 93% [e.g., 1, 2])

- 3 factors investigated
 - Linguistic context: specify the relevant bound at issue
 - upper bound [*not all*]
 - lower bound [*at least some*]
 - null context [target in isolation]
 - Evaluation apprehension [4]: anxiety of being judged
 - experimental protocol
 - annotation protocol
 - Task response categories [3]: structure of permitted responses
 - binary accuracy + Don't Know
 - ternary informativity + False

- Methodological conclusions
 - Implicature rate is affected by protocol type and response category
 - Response category interacts with linguistic context
 - but only rates for upper bound contexts are affected
 - Population effects may be large: university students and crowd-sourced workers differ
 - students sensitive to protocol
 - students and workers differently affected by linguistic context

- debriefing questionnaires reveal that task responses can underestimate implicature rates

Acknowledgements

Thanks to Jaye Padgett, and the UCSC Semantics Lab audience for feedback. Many thanks to Rosie-Wilson Briggs for stimulus construction.

References

- Geurts & Poussoulous 2009. *Embedded implicatures?!*, S&P.
- Katsos 2008. *The semantics/pragmatics interface from an experimental perspective: the case of scalar implicature*, *Synthese*.
- Orne 1962. *On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications*, *American Psychologist*.
- Rosenberg 1969. *Conditions and Consequences of Evaluation Apprehension*, *Artifacts in Behavioral Research*.
- Stohmertz 2008. *Research Artifacts and the Social Psychology of Psychological Experiments*, *Social and Personality Psychology Compass*.

Experimental Design

Task Instructions

Vary the social situation subjects believe they are participating in.

Experimental Protocol: Typical experiment instructions

Annotation Protocol: Checking the work of unaffiliated, possibly error-prone annotators

Context Story

All-Relevant - Upper bound is relevant; implicature is supported

All - Irrelevant - Upper bound is not relevant; literal interpretation is supported

None

Image Verification Task

Question: Would you be accurate if you said ...

... *'some of the spices have red lids'* ?

Available responses:
Yes, No, Don't Know

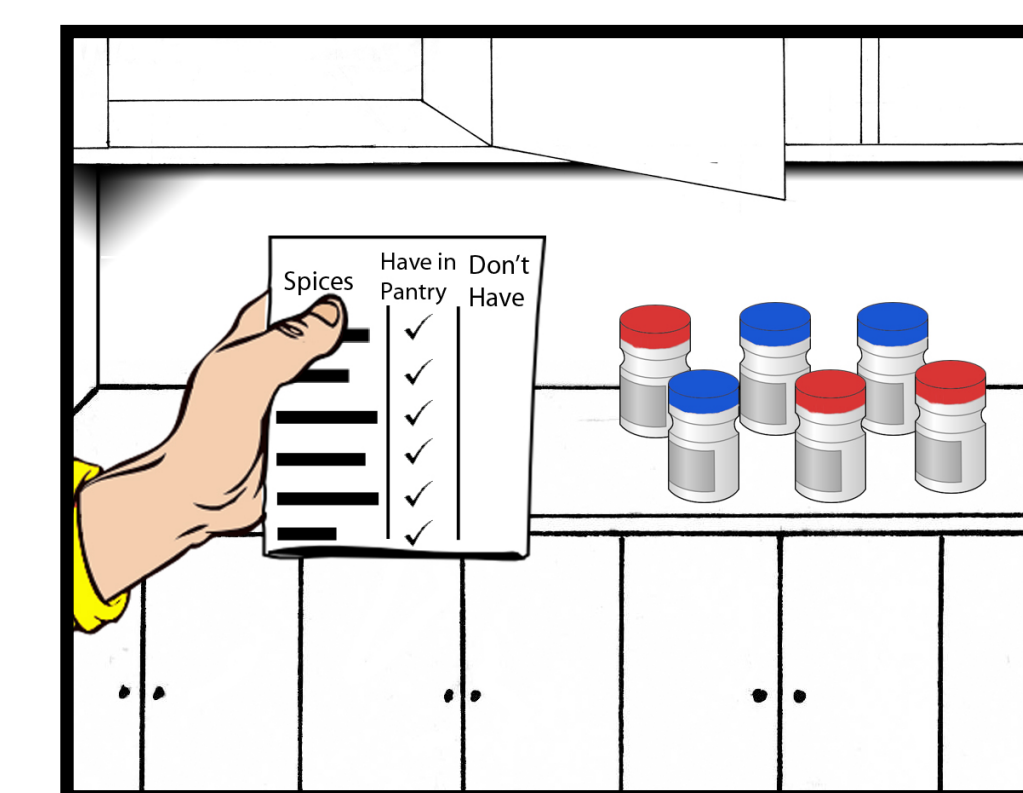


Image 1: Upper and lower bounds are satisfied

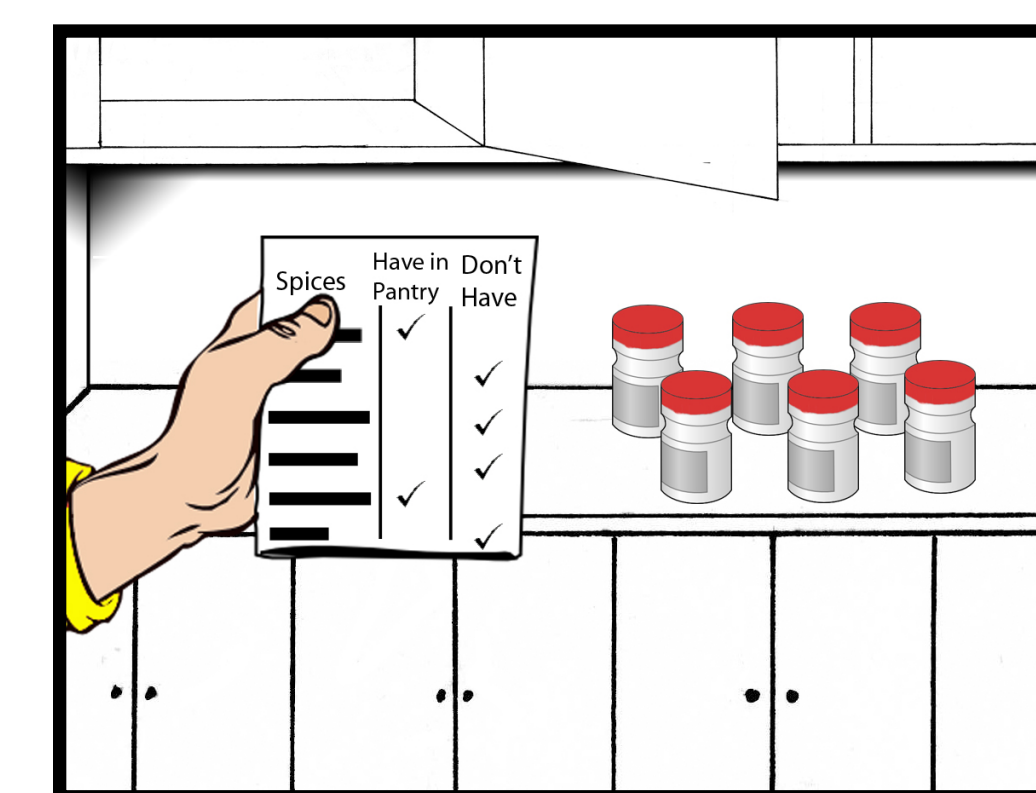


Image 2: Upper bound is violated

Experiment 1-2: Common design details	
Participants	Items
▶ Between-groups design.	▶ Within-groups design, counterbalanced sentence-picture pairs
▶ 20 participants per conditions	▶ 10 <i>all/some</i> sentence-picture pairs <ul style="list-style-type: none"> ▶ 4 implicature targets: <i>some</i> sentence, <i>all</i>-true picture ▶ 2 <i>some</i> sentence, <i>all</i>-false picture ▶ 4 <i>all</i> sentences: 2 <i>all</i>-true pictures, 2 <i>all</i>-false pictures
	▶ 20 non-quantificational fillers (3 ambiguous)

All-Relevant Story	All-Irrelevant Story
A major company that sells kitchen spices just issued a warning saying that their spices had been tainted. Any of the spices with the company's signature red lid need to be thrown out.	A major company that sells kitchen spices just issued a warning saying that their spices had been tainted. If you have eaten anything with one of their spices in the last two months, you'll need to go to the doctor. You pulled out all the spices that you've used in the last two months. The company always uses red lids to mark their spices.

EXPERIMENT 2 TASK RESPONSE CATEGORIES

Do specific prompt types affect implicature rates?

Yes - Response categories affect not only the baseline implicature rate but also **interact with linguistic context**.

New response prompt: **informativity**

How informative is this sentence?

Scale:

<False, Not Informative Enough, Informative Enough, Too Much Information>

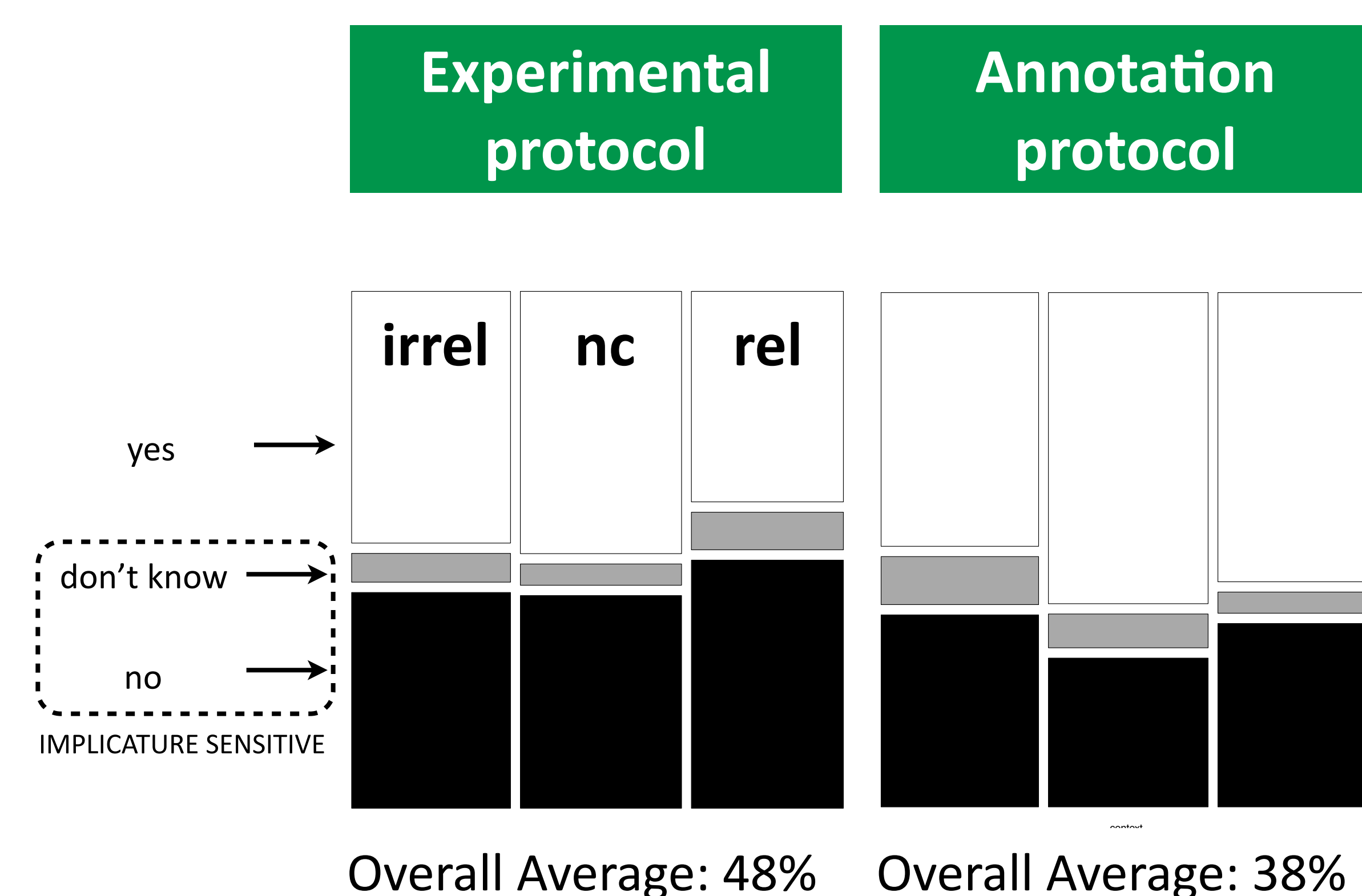
New Population: Amazon Mechanical Turk

EXPERIMENT 1 SOCIAL CONTEXT

Does social context of experiments change the rate of implicature responses?

Yes. Social context has an effect even when immediate linguistic context does not ($p < 0.05$).

Data analyzed with a **mixed-effects, cumulative link model** (ordered logistic regression), treating items as random effects. The only fixed effect justified for these models was the **protocol** factor. The control stimuli showed no effect of protocol.



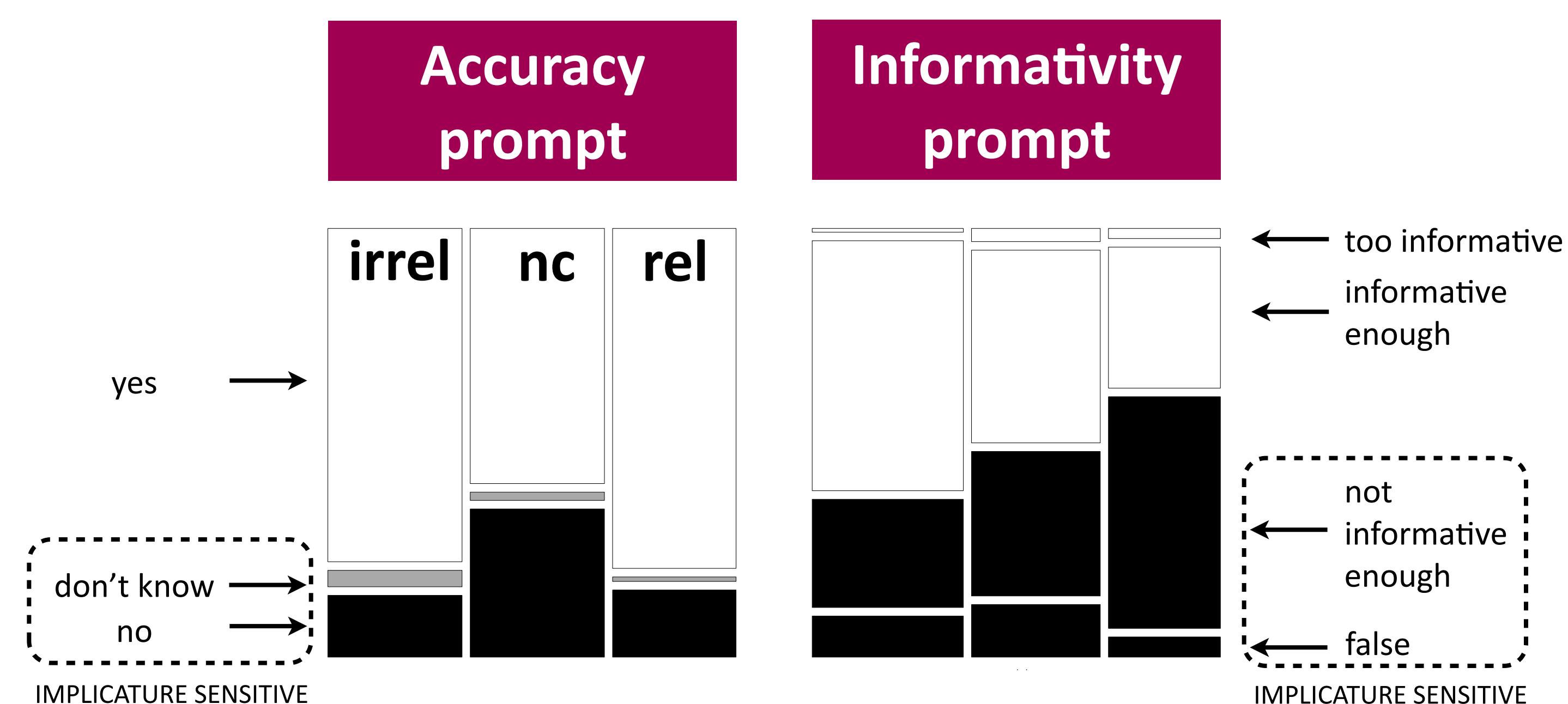
- Low implicature rates in the Annotation Protocol may be due to an abundance of *False* controls appropriate to an Experiment but not a quality-control-style task.
- Subjects contextualize a sense of *false*-ness specific to the task, and find the implicature cases do not fit their subsequent notion of *false* in context.

Experimental protocol

Reliable effects:
context ($p < .001$)
prompt ($p < .001$)
context:prompt ($p < .06$)

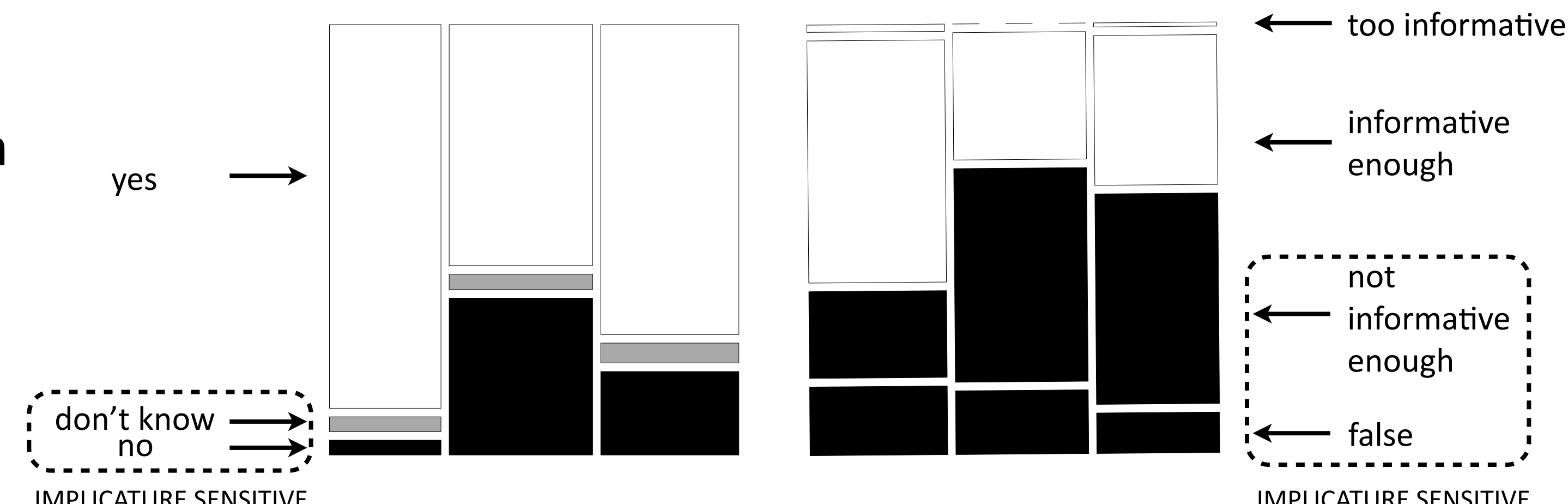
Effects reported for coefficients in a **mixed-effects, logistic regression model**, treating items as random effects.

- ▶ To compare prompt conditions, the response categories labeled as **IMPLICATURE SENSITIVE** were collapsed.
- ▶ We also separately analyzed each prompt condition with a cumulative link model, as in Experiment 1, and found the same pattern for context for both prompt types.



Annotation protocol

Not reliably different from experimental protocol



Post-experiment debriefing

- subjects asked to comment on target trial responses
- a majority show awareness of *all* alternative.
- many indicated a preference for *all* alternative:

"ALL ... would be a better response, but SOME ... is true."
"...all have nuts, not just some. I didn't feel the categories fit."

Implicature awareness	73%
Spatial ambiguity of stimuli	5%
Unclear stimuli	16%
Other	6%