

“Some” tasks are not optimal: Concerns about truth value judgment tasks for assessing scalar implicatures

Les Sikos^{1,2}, Minjae Kim³, Jacqueline Lane⁴, and Dan Grodner⁵

¹Saarland University · ²Cluster of Excellence MMCI, Saarland University · ³Boston College · ⁴University of Pennsylvania · ⁵Swarthmore College

Introduction

Scalar Implicature

- Much of what we communicate in conversation is implicit
- If someone says “*Some students passed the test*,” listeners often infer that *not all* the students passed
- This **pragmatic inference** arises because communication is typically cooperative [1]
- Cooperative speakers are expected to deliver strongest (most informative) utterance [2]
- Scalar implicatures have become a central testing ground for investigating how implicit meanings are computed

Dominant View

Generating scalar implicatures is cognitively effortful

- Most evidence comes from **verification tasks** using **underinformative** sentences (UIs) [3]
“Some elephants are mammals” True/False
- UIs are literally true, but their implicated meaning is false
False → scalar implicature was computed
True → literal interpretation was computed
- Response patterns in this task are sensitive to cognitive load, which is expected if scalar implicatures require cognitive resources:
High load → increases acceptance rate [3-8]

Alternative Possibility

Binary decision is what makes task cognitively effortful

- UIs are neither patently true nor false, instead they are pragmatically odd
- Consequently, it may be difficult to outright accept or reject a UI because it is simply infelicitous

References

- [1] Grice (1975). *Logic and conversation*.
- [2] Levinson (2000). *Presumptive meanings*.
- [3] Bott & Noveck (2004). Some utterances are underinformative. *JML*.
- [4] Bott, Bailey & Grodner (2012). Distinguishing speed from accuracy in scalar inferences. *JML*.
- [5] De Neys & Schaeken (2007). When people are logical under cognitive load. *Exp Psych*.
- [6] Dieussaert, Verkerk, Gillard & Schaeken (2011). Some effort for some: further evidence that scalar inferences are effortful. *QJ of Exp Psych*.
- [7] Marty & Chemla (2013). Scalar implicatures: working memory and a comparison with only. *Frontiers in Psych*.
- [8] Feeney, Scranton, Duckworth & Handley. (2004). The story of some: Everyday pragmatic inference by children and adults. *Canadian J of Exp Psych*.
- [9] Foppolo, Gausti & Chierchia (2012). Scalar implicatures in child language: Give children a chance. *Lang Learning & Dev*.
- [10] Sikos, Kim & Grodner (2016). Speaker likeability influences utterance acceptability: Social context modulates tolerance for pragmatic violations in adults. *CUNY*.

Methods

Participants

- 65 native English speakers
- Individual differences assessed via Operation Span Task (OST) $M = 93.9\%$ (cf. 62.4% [6])

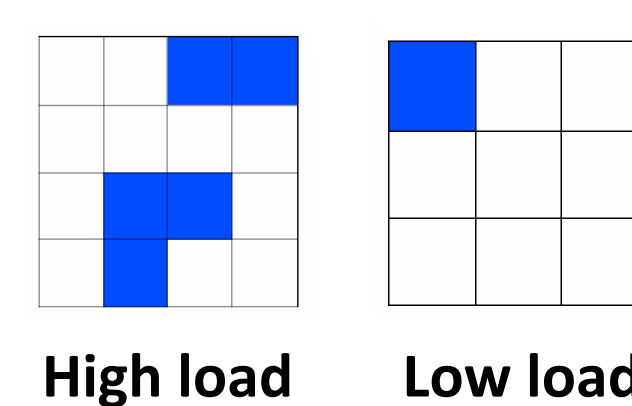
Dual Task

- Remember a **low** or **high load** visual pattern
- Sentence acceptability judgment using **binary response** (T/F) or **7-pt scale** (1: unnatural, 7: natural)

Stimuli

- 120 items (60 UI)
- Critical UI sentences (T6) were pseudorandomly intermixed with 5 types of fillers

Load Manipulation



Type	Example	Literal Veracity
T1	All dogs are spaniels.	False
T2	All spaniels are flowers.	False
T3	All spaniels are dogs.	True
T4	Some dogs are spaniels.	True
T5	Some spaniels are flowers.	False
T6	Some spaniels are dogs.	???

Block	List 1	List 2	List 3	List 4
1	TF_high	TF_low	R_high	R_low
2	TF_low	TF_high	R_low	R_high
3	R_high	R_low	TF_high	TF_low
4	R_low	R_high	TF_low	TF_high

Design

- 2 (cognitive load) x 2 (response type) block design
- Cognitive load and response type manipulated across blocks
- Participants first saw either true/false or ratings blocks, within which high and low load blocks were in same order
- Conditions were evenly distributed across blocks
- Block order was counterbalanced across lists

Predictions

- If generating scalar implicatures is effortful, then responses to UIs should be affected by load manipulation for both response types
- If difficulty is primarily driven by choosing between suboptimal options, then load should affect responses only in true/false condition
- Responses in graded condition should be unaffected because intermediate ratings should be preferred regardless

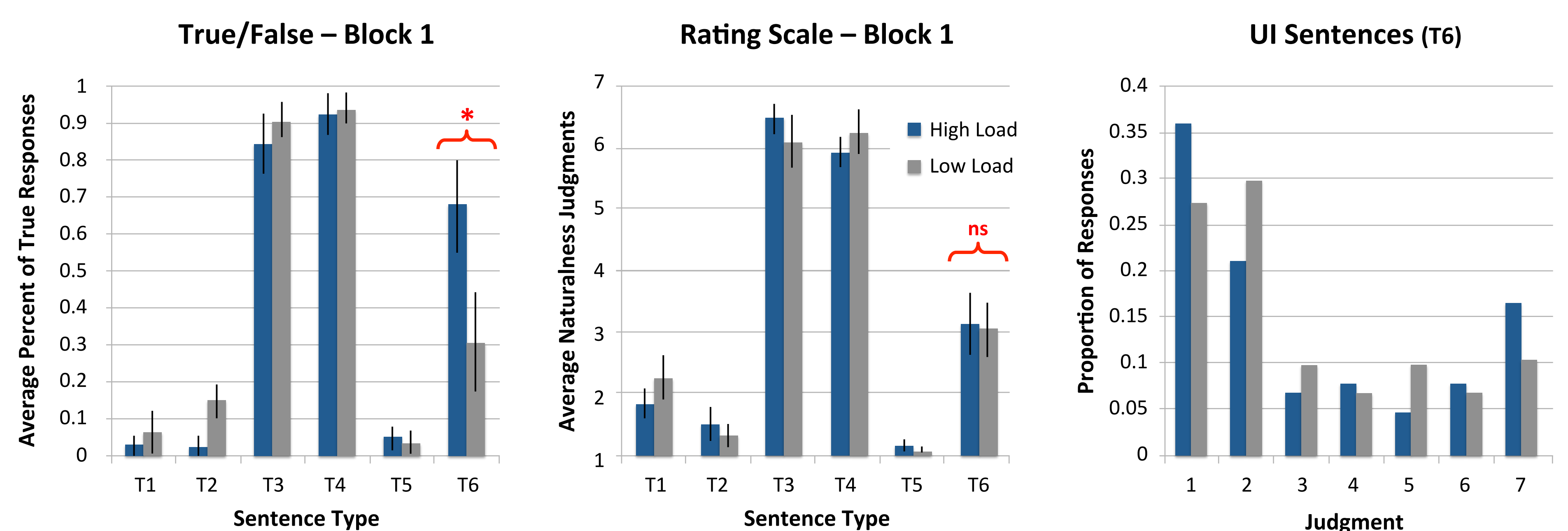
Results and Conclusions

Exclusion Criteria

- 3 participants excluded for accuracy < 80% on fillers
- 1 participant excluded for accuracy < 80% on OST
- 1 participant excluded for asking about UI sentences

Analysis of all blocks

- No reliable effect of load on T/F or Ratings responses
- Follow-up analyses suggested that participants were no longer affected by load after first block of that type



Results were consistent with alternative possibility

- True judgments increased under high load, but naturalness ratings were unaffected by load
- When given intermediate response options, UI sentences were judged to be more natural than patently false sentences, but less natural than patently true sentences
- No reliable effect of load on fillers
- This pattern of results is consistent with previous work [8]

Conclusions

- Findings indicate participants can judge UIs as acceptable even when they understand that UIs are pragmatically inappropriate
- This pattern of results argues against a strictly competence-based or resource-based view of pragmatic inference
- Results also raise concerns about the widespread use of binary choice tasks for investigating pragmatic processing
- Binary judgment tasks should be used with care

