Pre-registration report

Study information

Title: Ignorance in context: The interaction of modified numerals and QUDs - A replication study

Authors: Mae Grenz, Merlin Marinova, Esma Sakalli, Johanna Venzke

Description:

Hypotheses:

1. Ignorance inferences are context-sensitive for both modifiers.
2. When the context is underspecified, there will be a difference between the modifiers *at least* and *more than*.

Design Plan

Study type: Browser-based experiment

Blinding: The participants and experimenters will not have direct contact, due to the study being conducted browser-based.

Study design: Within-subject design

Randomization: All participants see all experimental items and fillers, the items are randomized with the Latin square design and the fillers are randomized for each participant with a version of the fisher-yates shuffle. All participants are rotated randomly through the 6 lists. In total, there are 108 stimuli (36 items + 72 fillers) in each experiment with randomized order for each participant.

Sampling Plan

Existing data: Data from a pilot study (N=3) was available and guided the specification of statistical models. This data, however, will not be included in the final analysis.

Data collection procedures

Sample size: In the original experiment they used 35 participants, we want to achieve at least as many. Due to that we are aiming to recruit 40 participants (via studall mail and friends and family), therefore, we have room to exclude those from the study who did not properly execute the experimental task.

Sample size rationale: Since we have a deadline to hold and cannot reach many participants solely via our own contacts, and since we also do not have any

incentives to offer, we apologize in advance for the possibility of not fulfilling the desired number of participants.

Stopping rule: We will stop data collection when we reach the predefined sample size.
Or when our deadline is reached, which is at midnight, the 22nd of August.

Variables

Manipulated variables: We manipulate the type of question posed by the judge (QUD) and the type of scalar modifier in the witness's answer (MOD). The QUD has three levels: it ranges over {POLAR, WHAT, HOWMANY} (reference level: POLAR). MOD has two levels and ranges over {COMP, SUP} (reference level: COMP).

Measured variables: We record the reading times per word of the witness's answer and measure how justified the participants found the judge's conclusion. Concretely, variable RT is a metric variable capturing reaction times; variable VALIDITY_JUDGEMENT is discrete based on a 5-point Likert scale (1: not justified at all, 5: strongly justified). However, due to time restrictions, we will probably not be able to follow up on the RT measurements and analyze them.

Indices: We do not consider any indices.

Analysis Plan

Statistical models: We will run frequentist models with variate "VALIDITY_JUDGEMENT" (whether the participant thinks the judge's conclusion was justifiable on a 5-point Likert scale). The validity-judgement data is presented with mixed-effects ordinal probit regression models. The participants and the items are considered as random effects while the question and answer conditions are treated as fixed effects. An interaction model with all two-way interactions between the question conditions (POLAR, WHAT, HOWMANY) and answer conditions (SUP, COMP) and a main-effect only model will be run. All statistical modeling reported will be programmed in R. The models are estimated using the *ordinal* R package. More details are provided in the analysis script which you can find in the associated github repository.

Transformations: No transformations will be applied.

Data exclusion We control and exclude participants who do not judge the filler trials appropriately. Specifically, this is the case when a participant does not react truthfully to the obviously valid, plausible, implausible or invalid inferences (all of which was judged by the experimenters). This concerns only the filler trials. Obviously invalid inferences must be answered with a rating of 1, otherwise the reaction will be considered as not truthful. However, in order to verify the collected data, we will only exclude data from a participant who has reacted untruthfully on more than eleven obviously invalid sentences. Reaction to obviously valid, plausible and implausible sentences will not affect data exclusion.

Missing data: Should a data set not be recorded completely, we will use all data available from that participant.

Exploratory analysis: We intend to investigate also the following questions of interest if time allows it:

1. RTs will be higher as the ignorance inferences increase.