Evidence for Adaptation in Scalar Implicature Calculation

Background. Scalar implicature (SI), i.e., the enrichment of a sentence like *Mary ate some of the cookies* to *Mary ate some, but not all, of the cookies*, is widely acknowledged to be a variable, context-dependent inference. Yet it was observed as early as Bott & Noveck (2004) that in an experimental setting, some participants exhibit consistent behavior by either always calculating SI or never calculating it. Only a subset of people vary in their response to SI-probing questions. At the same time, it is typical of experimental studies on SI to report findings in terms of "SI rate", that is, the average rate of SI calculation abstracting over participant differences. This includes studies reporting e.g., that SI rates increase with a supportive Question Under Discussion (Ronai & Xiang, 2021, a.o.) or following a salient alternative prime (Rees & Bott, 2018, a.o.) or with certain lexical scales (van Tiel et al., 2016, a.o.). There do exist some studies that have explored by-participant variation in SI calculation (Spychalska et al., 2016; Hunt et al., 2013; Degen & Tanenhaus, 2015; Marty et al., 2024, a.o.). But such studies have varied greatly in how they define *consistent responders* vs. *inconsistent responders* (e.g., does a participant have to be 100% consistent in calculating SIs, or does 70% suffice?) and in what distinct participant groups they identify: pragmatic (=SI calculator) vs. literal (=SI non-calculator), or consistent vs. inconsistent, or literal vs. pragmatic vs. inconsistent. This latter, three-way distinction is supported by Heyman & Schaeken's (2015) latent class analysis (see also Ramotowska, et al., 2024 for a Bayesian modelling of by-participant variation). Crucially, however, no existing study - to our knowledge - has investigated how (and indeed whether) participant response types develop over time, in the course of an experiment.

Research Question and Hypotheses. The present study explores the dynamics of participant behavior in SI calculation. We focus on inconsistent participants, which we define as anyone not 100% consistent in providing a pragmatic or literal response on every trial of an experiment. (100% consistent participants trivially show no change over time.) We contrast two hypotheses:

H1: Propensity to (not) calculate SI is part of one's mental grammar and the rate of SI calculation is thus fixed for a speaker. E.g., someone might calculate SIs 30% of the time, and someone else 70%, etc., in a stable manner. This hypothesis is tacitly assumed by prior experimental studies on SI that explain some other variable (such as processing cost) by differences among participant groups but overlook the potential change in participants' responses over time (Degen & Tanenhaus, 2015; Ramotowska, 2025, a.o.).

H2: People who enter an experiment with some amount of uncertainty regarding SI calculation adapt to the experimental context and develop strategies. This adaptation may happen in both directions: people may become more consistently literal or more consistently pragmatic throughout an experiment. Whether someone converges on a literal or a pragmatic response strategy may depend, for instance, on what Question Under Discussion they assume. Such context-adaptation behavior has been observed before in SI priming (Marty et al., 2024, a.o.).

Methods. We analyzed publicly shared data from two published papers on SI calculation, Kursat & Degen (2020, henceforth "K&D") and Ronderos and Noveck (2023, "R&N"). They used very different methodologies: K&D used the gumball paradigm, in which participants hear a sentence such as *You got some (of the) gumballs* and indicate whether they agree with this statement after seeing a situation where all of the gumballs in the machine dropped. "Disagree" indicates SI calculation (=pragmatic response). R&N's participants were shown sentences like *Some cats are mammals* which they had to judge to be true or false. "False" indicates SI calculation. Each paper includes multiple experiments using the same paradigm. K&D manipulated partitivity and a background story, resulting in 4 data sets for our analysis. We analyze R&N's Exps.2-3, which manipulate the presence of a specific speaker (yielding 2 data sets). To test our hypotheses, we investigate: 1) the change over time of participant uncertainty about SI calculation; 2) the development over time of the eventual distribution of responses. We operationalized 1) as follows: for each data set, we created time windows with the widths of 3 trials (the n^{th} window includes the n^{th} to the $(n+2)^{th}$ trials), and calculated the Shannon

information entropy for each participant's responses within each window ($H=-p\cdot\log_2(p)-(1-p)\cdot\log_2(1-p)$, where p is the percentage of pragmatic responses). This provides a measure of the participant's response consistency at different positions of the experiments. H2 predicts an entropy decrease over time, as participants' responses become more consistent. Under H1, we should see no such systematic decrease in entropy over the course of the experiment.

Results. Figure 1 shows the mean entropy value for each time window for K&D's Dataset 1 and R&N's Exp.2 (the other datasets show the same qualitative pattern). We fitted regression models predicting entropy values by window position; entropy values decrease as window position increases for all data sets we analyzed (K&D 1: β=-0.45, p<0.05; K&D 2: β=-0.84, p<0.001; K&D 3: β=-1.37, p<0.001; K&D 4: β=-0.73, p<0.001; R&N Exp.2: β=-0.20, p<0.05; R&N Exp.3: β=-0.16, p<0.01). This suggests that participants in these studies become gradually more consistent with their responses throughout the experiments, in line with H2. To better visualize this process, Figure 2 shows the cumulative distribution of participant responses throughout R&N Exp.2: most participants start out without a strong bias for either pragmatic or literal responses; as the experiment progresses, a bimodal pattern emerges, with participants developing biases toward one of the two response types. Other data sets show similar trends.

Figure 1. Moving-window response entropy plot

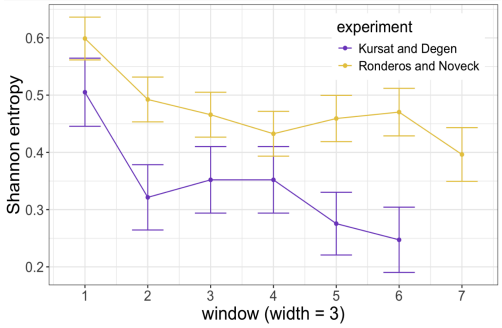
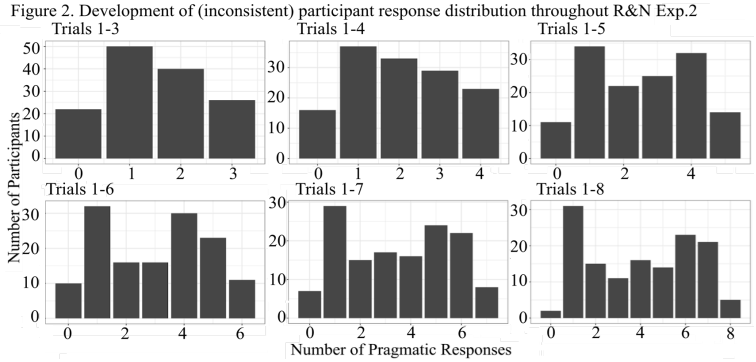


Figure 2. Development of (inconsistent) participant response distribution throughout R&N Exp.2



Discussion and Conclusion. By-participant variation remains a lesser studied topic within SI, with many papers instead reporting averaged "SI rates". Existing work that has explored this question is quite heterogeneous in what participant types it posited and how it defined them. Crucially for us, no existing study has probed whether the theorized participant types are stable throughout an experiment. Here, we analyzed data from published work on SI and found that participants' response consistency increases (operationalized as a decrease in entropy) over time in all analyzed data sets. We take this as evidence of context adaptation, that is, change in participant response behavior induced by the experimental context. This suggests that "inconsistent" participants (=those not providing 100% consistent responses) don't have a fixed likelihood of SI calculation as part of their mental grammar. Rather, they gradually resolve their uncertainty regarding SI calculation by developing toward (and possibly eventually into) a fully consistent literal or pragmatic participant. Thus, it can only be maintained that there are two participant types in SI calculation (literal and pragmatic), while the inconsistent third type is not a stable one (cf. Heyman & Schaeken, 2015, a.o.). It is also possible, though our findings don't directly speak to this, that literal and pragmatic participants became fully consistent by undergoing adaptation at some point prior to the experiment. Our study has important consequences for the SI literature, since individual variability and different participant types have previously been invoked as impacting various experimental findings, such as processing cost (Degen & Tanenhaus, 2015; K&D, a.o.) or the direction of priming (Marty, et al., 2024). Given our finding that response consistency evolves throughout an experiment, we should be cautious about deriving other properties from participant types.

Selected References. K&D, 2020, Probability and processing speed of scalar inferences is context-dependent, Proceedings of Cogsci. | R&N, 2023, Slowdowns in scalar implicature processing: Isolating the intention-reading costs in the Bott & Noveck task, Cognition.