Uniqueness vs. familiarity in interpreting definite descriptions

Sadhwi Srinivas & Kyle Rawlins (Johns Hopkins University)
sadhwi@jhu.edu

[Introduction.] Uniqueness theories of definite descriptions claim that a description is felicitous IFF a unique referent satisfies its literal meaning (Russell, 1905; Evans, 1977; Lobner, 1985). Orthogonally, familiarity theories claim that reference succeeds IFF the referent has been made salient in the preceding discourse context (Kamp, 1981; Heim, 1982). Here, we experimentally investigate the process by which hearers interpret descriptions in English to moderate between these 2 theories of definite descriptions. We then implement a Rational Speech Acts model (RSA; Frank & Goodman, 2012), and discuss in what ways hearers adhere to and diverge from rationality.
[Experiment.] 40 M-Turk participants read descriptions of 10 stories imagining themselves as one of the characters (the *hearer*). After each story, they interpreted a definite description uttered by the character they were interacting with (the *speaker*). We used a 2 (speaker uniqueness) x 2 (hearer uniqueness) x 3 (linguistic salience) within-subjects design, where ***speaker uniqueness*** (***hearer uniqueness***) was true if the speaker (hearer) knew there to be a unique referent satisfying the description. ***Linguistic salience*** was our proxy for familiarity which held of an object if it had been mentioned prior to the description to be interpreted. In every story, two objects featured prominently. Their status varied independently on speaker and hearer uniqueness. As for salience, it could hold of neither or of 1 of the 2 objects. Fig (1) shows a trial in which speaker uniqueness was true (the auditor knew only the labeled jar to be *the jar of camphor*) but hearer uniqueness wasn't (to the lab worker, both the labeled and the salient unlabeled jar were jars of camphor). At the end, hearers chose the intended referent (the labeled jar, the unlabeled jar, or *"Don't know"*).
[Results.] The main findings from our experiment are as below:

1. Reference succeeded as long as speaker uniqueness held, regardless of hearer uniqueness.
2. There were ranked effects of uniqueness & familiarity. Uniqueness helped hearers pick a referent regardless of familiarity (right panel in fig 2). But in the absence of uniqueness, familiarity helped only about 60% the time (left panel in fig 2).
3. When the non-speaker-unique object (*distractor*) was salient (the unlabeled jar), hearers tended to choose it more often ($\sim$20%) than they did a non-salient distractor ($\sim$8%).

[Computational Model.] The RSA explicitly models how an agent accounts for their interlocutor's knowledge via iterative Bayesian reasoning within a communicative task. Our implementation of the RSA model builds in 2 main ways on the traditional implementations discussed in the literature:

1. Typically, the RSA is used in situations where the literal semantics are shared by both speaker and hearer. But this assumption wasn't appropriate in our scenarios which were explicitly designed to provide hearers with privileged knowledge. To handle the mismatch, we assumed that hearers began with a speaker-centered semantics (per experimental results); however, noise (a trained parameter) could be introduced by the hearer's own knowledge.
2. Traditionally in RSA, there isn't a way to avoid choosing any option at all; but this was allowed in our experiment. We used the idea that hearers choose *"Don't know"* proportionally to the entropy of the rational listener's belief distribution. The entropy was translated to a probability value by a logistic function and chances of picking each referent were then recomputed.

Salience was operationalized by incorporating it into the prior probability. Our model correctly predicts that salience doesn't add much on top of speaker-uniqueness when the latter holds (green in table 1). However, it underestimates hearers' susceptibility to salient distractors (in red), and overestimates how much salience helps in the absence of speaker-uniqueness (in orange). **Overall**, the contribution of our study is two-fold. First, we evaluated the relative strengths of uniqueness and familiarity cues towards definite description interpretation. We found ranked effects of both, with uniqueness being stronger. Then, we implemented an RSA model to find that participants

behaved “rationally” in many ways, but were also less capable of using salience as a cue as well as more susceptible to salient distractors than what fully rational reasoning would predict.

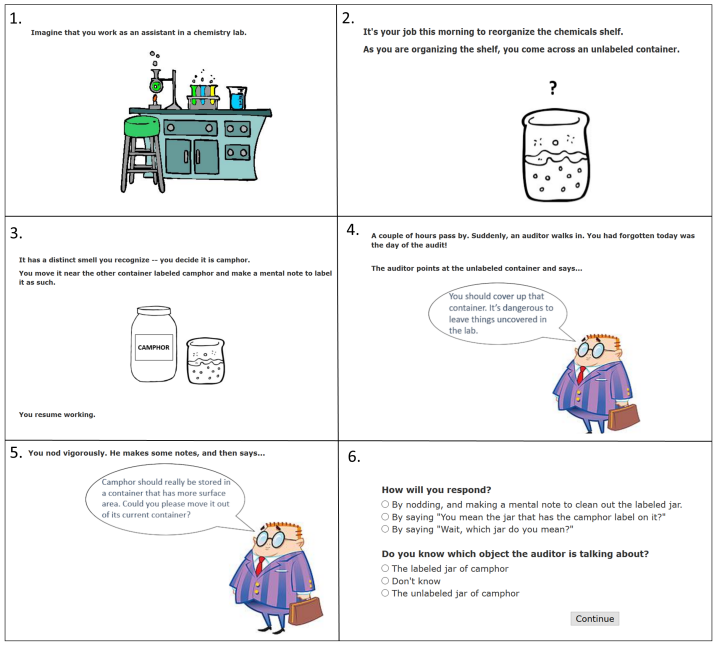

Figure 1: Example of an experimental trial

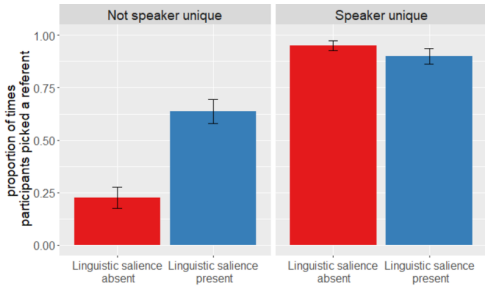

Figure 2: Uniqueness vs familiarity

Situation	RSA model			Actual data		
	p(A)	p(B)	p(failure)	p(A)	p(B)	p(failure)
A satisfies the literal content of the description; B does not; neither A nor B is salient	0.83	0	0.17	0.8625	0.0875	0.05
A satisfies the literal content of the description; B does not; A is salient	0.83	0	0.17	0.8625	0.0375	0.1
A and B both satisfy the literal content of the description; neither A nor B is salient	0.175	0.175	0.65	0.1	0.125	0.775
A and B both satisfy the literal content of the description; A is salient	0.71	0.03	0.26	0.55	0.0875	0.3625
A satisfies the literal content of the description; B does not; B is salient	0.63	0.05	0.32	0.6	0.23	0.175

Table 1: Comparing model predictions to actual experimental data

Selected References.

Russell, B. (1905). On denoting. *Mind*, 14(56), 479-493.

Evans,G.(1977). Pronouns, quantifiers, and rel clauses (I). *Canadian journal of phil*, 7(3), 467-536.

Lobner, S. (1985). Definites. *Journal of semantics*, 4(4), 279-326.

Heim, I. (1982). The semantics of definite and indefinite noun phrases, *UMass PhD*.

Frank, M. C.& Goodman, N. D. (2012). Predicting pragmatic reasoning in language games.*Science*, 336(6084), 998-998.