Production Expectations Modulate Contrastive Inference

Elisa Kreiss (ekreiss@stanford.edu)
Department of Linguistics, 460 Jane Stanford Way
Stanford, CA 94305 USA

Judith Degen (jdegen@stanford.edu)
Department of Linguistics, 460 Jane Stanford Way
Stanford, CA 94305 USA

Abstract

Contrastive inferences, whereby a listener pragmatically infers a speaker's referential intention of a partial referring expression like *the yellow* by reasoning about other objects in the context, are notoriously unstable. We report a production-centric model of interpretation couched within the Rational Speech Act framework. Adjective production probabilities a listener expects for objects in a context drive the size of contrastive inferences: the greater the asymmetry in expectation for a speaker to use a pre-nominal adjective for the target rather than for competitors, the greater the listener's resulting target preference. Modifier production probabilities were collected (Exp. 1) and used to make predictions about comprehension in an incremental decision task (Exp. 2). The model's interpretation predictions are supported by the data. This account has the potential to explain the fluctuating appearance of contrastive inferences and shifts the explanatory focus away from contrastive inference towards online interpretation of referring expressions more broadly.

Keywords: contrastive inference; RSA; typicality; incremental processing

Introduction

One of the most interesting features of language is its flexibility. In referring to an object, speakers choose from a wealth of possible referring expressions. *The banana*, *the yellow banana*, and *the curvy fruit* are all expressions that can refer to the same object. Moreover, the same utterance – e.g., *the banana* – can be used to refer to different kinds of objects (yellow bananas, brown bananas, etc.). This flexibility poses a challenge for listeners, who have been shown to rapidly draw pragmatic inferences about speakers' referential intentions in online processing. Consequently, understanding how listeners process referring expressions – in particular, to what extent contextual information enters into this process – has been a central topic of psycholinguistic research.

Language is processed incrementally. Eye-tracking experiments have shown that upon hearing an incomplete utterance like *the yellow* in a display like Fig. 1a, listeners start to fixate the yellow objects more than other objects even before they hear the disambiguating noun *banana* (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). Additionally listeners go beyond the information contained in the signal itself in processing language; they also take into account contextual information – including the nature of other possible referents – to draw rapid pragmatic inferences about a speaker's intended referent. One such inference that has received much attention in recent years is the so-called *contrastive inference* (Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Aparicio, Kennedy, & Xiang, 2018; Grodner & Sedivy, 2011; Rubio-Fernandez, Terrasa, Shukla, & Jara-Ettinger, 2019; Ryskin, Kurumada, & Brown-Schmidt, 2019). Consider the context in Fig. 1b that shows a yellow and an orange banana, a yellow corncob and some other distractor item. When a listener is asked to *Click on the yellow...*, there are two eligible objects to choose from: the yellow banana and the yellow corncob. Rather than considering both yellow objects equally likely target referents, listeners often exhibit a preference, evidenced by increased looks, for the yellow object that has a contrasting member of the same type and different color in the display (i.e., the banana, Sedivy et al., 1999; Sedivy, 2003). When the contrast is absent, as in Fig. 1a, listeners have no such preference. This preference for the target over the competitor elicited by the presence of a contrast (i.e., the orange banana) is considered the result of drawing a contrastive inference.

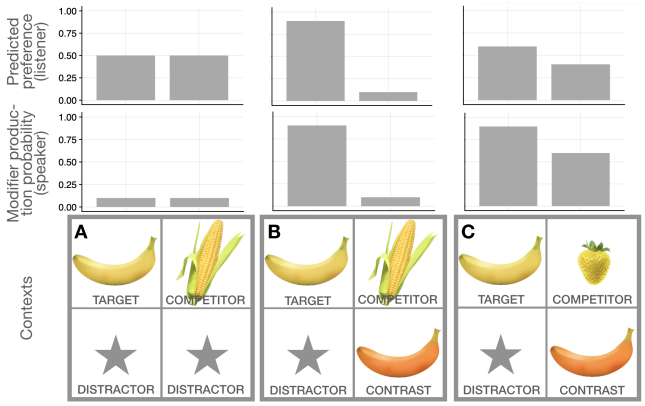

Figure 1: Three contexts, each with a yellow banana as the target and another yellow object as its competitor. The competitor can be typical (A, B) or atypical (C), and a contrast can be absent (A) or present (B, C). Gray stars represent other distractors that do not share color or type with any other object.

Contrastive inferences arise as the result of listeners expecting a cooperative speaker to not be more informative than required by the context (Grice, 1975). The presence of a contrast object makes it contextually necessary to include the adjective. In contrast, the adjective is not necessary to refer to the competitor object. Upon observing the adjective, listeners can reverse-engineer that the intended referent must be

the color-congruent object with a contrast member, i.e., the yellow banana in Fig. 1b (Aparicio et al., 2018; Grodner & Sedivy, 2011; Ryskin et al., 2019; Sedivy et al., 1999).

This simple Gricean account that only takes into consideration the contrastive function of the adjective predicts that contrastive inferences should arise whenever the target object occurs in the presence of a contrast object. It is surprising, then, that contrastive inferences are not consistently observed across experiments. While the contrastive inference effect has been replicated reliably in the size adjective domain (Aparicio et al., 2018; Grodner & Sedivy, 2011; Heller, Grodner, & Tanenhaus, 2008; Ryskin et al., 2019; Sedivy et al., 1999), the effect is less stable with color adjectives (Sedivy, 2003). Sedivy (2003) reports that the contrastive inference arises in contexts where the target object has a predictable color (such as the yellow banana in Fig. 1) but not when it is replaced by an object with an unpredictable color like a cup, which comes in many colors. She shows that these objects differ in how likely a speaker is to produce the color modifier for the object in isolation: in the absence of a contrast, a yellow banana is usually called *the banana* while a yellow cup is often called *the yellow cup*, which Sedivy (2003) calls these objects' *default descriptions*. Only in cases where the modifier is not part of the default description, she argues, is its observation surprising and can be interpreted as a signal that will elicit the contrastive inference.

In addition to expectations of informativity as described above, contrastive inferences have been proposed to depend on the semantics of the adjective involved, such that it reliably arises with relative adjectives (e.g., size adjectives) and maximum standard absolute adjectives (e.g., *full*), but not with minimum standard absolute adjectives (e.g., *empty*); while the evidence from color adjectives is conflicting (Rubio-Fernandez et al., 2019; Sedivy, 2003). Furthermore the effect disappears when the listener considers the speaker unreliable (Grodner & Sedivy, 2011; Ryskin et al., 2019).

In this paper, we investigate an account of contrastive inference that has the potential to unify the above properties by reducing them to listeners' expectations about the speaker's contextual probability of producing the pre-nominal adjective. In so doing, we follow recent research highlighting the importance of listeners' generative model of the speaker in generating pragmatic inferences (Hawkins, Gweon, & Goodman, 2018; Kao & Goodman, 2015; Kleinschmidt & Jaeger, 2011; MacDonald, Pearlmutter, & Seidenberg, 1994; Mitchell, Cuetos, Corley, & Brysbaert, 1995; Rubio-Fernández & Jara-Ettinger, 2018). We formalize the relevant listener-side reasoning within the Rational Speech Act (RSA) framework (Frank & Goodman, 2012; Goodman & Frank, 2016; Cohn-Gordon, Goodman, & Potts, 2019), a state-of-the-art computational framework that models pragmatic inference as the result of listeners performing Bayesian inference on their speaker model and their prior beliefs about likely meanings, thereby giving the speaker model a central role in the inference. It provides a way to quantitatively assess the probabilities that a listener assigns to possible referents after observing partial sentences of the form *Click on the yellow...* given their prior beliefs and expectations about the speaker. This account shifts the explanatory focus away from specific cognitive and linguistic factors that influence contrastive inference and towards listener's production expectations (and their prior beliefs, which we don't treat in depth in this paper).

For this investigation it is important to distinguish between two notions: the behavioral pattern that manifests as a *target preference*, i.e., a preference for the target over the competitor; and the theoretical construct of a *contrastive inference*, i.e., the increase in target preference when a contrast is present vs. when it is absent.

We proceed by first showing that our production-centric account makes the same qualitative predictions about the basic contrastive inference effect as for instance the default description account proposed by Sedivy (2003). We then derive new predictions about the size of target preferences across different contrast-present and contrast-absent contexts. We report a free production study (Exp. 1) we conducted to elicit modifier probability estimates, which we used to determine quantitative model predictions. We evaluate the model by comparing those predictions to empirical comprehension data which we elicited using an incremental decision task (Exp. 2).

A Bayesian account of contrastive inference

The Rational Speech Act framework (Frank & Goodman, 2012; Goodman & Frank, 2016) is a probabilistic (and thus non-deterministic) Bayesian account of natural language which ascribes a central role to the speaker in pragmatic interpretation. The core idea of the model is that a listener and a speaker recursively reason about each other: A pragmatic listener L_1 wants to infer the meaning of an utterance u, as formulated by the pragmatic speaker S_1. Possible referents r are assigned a probability proportional to the probability that S_1 will produce u to convey r multiplied by the listener's prior belief in r $P(r)$, as defined by Bayes' Rule.[1]

$$P_{L_1}(r|u) \propto P_{S_1}(u|r) * P(r) \tag{1}$$

To simplify the following example, we will assume that listeners have a uniform prior $P(r)$ over all objects in the display[2]. Then the RSA model predicts a direct relationship between the production probabilities P_{S_1} and the listener's distribution over possible referents P_{L_1}.

While RSA has typically been applied to the analysis of full utterances, it can straightforwardly be extended to generate predictions at the sub-sentential level[3]. To generate RSA

[1] The pragmatic speaker model and further recursive steps are spelled out in detail elsewhere (e.g., Goodman & Frank, 2016). Since we will elicit speaker probabilities empirically, we need not be concerned with the details of the speaker model.

[2] The simplifying assumption is justified by the results of Exp. 2.

[3] One exception is the Incremental Iterated Response Model of Pragmatics, which is also shown to qualitatively predict contrastive inference in general (Cohn-Gordon et al., 2019).

predictions for an incomplete referring expression such as *Click on the yellow...*, we take P_{S_1} to correspond to the contextual probability of color mention for each referent in the display. This corresponds to marginalizing over the probabilities of all continuations of the utterance (i.e., *Click on the yellow banana/corncob/lettuce/...!*). Let's investigate this account's qualitative predictions:

Consider the example contexts in Fig. 1a and 1b. Upon hearing the modifier *yellow*, the pragmatic listener P_{L_1} considers how likely a speaker is to include this modifier in their referring expression for each object in the display. Since only the target (yellow banana) and the competitor (corncob) are yellow, we assume that the production probabilities of *yellow* for the other objects in the display are 0. This only leaves the target and the competitor as potential referents.

Hypothetical modifier production probabilities for target and competitor are shown in the middle row of Fig. 1. Assume that in the absence of a contrast object (Fig. 1a), speakers are equally unlikely to include the color modifier when referring to the target banana (probability 0.1) and its color competitor, the corncob (0.1). Pragmatic listener predictions are obtained by renormalizing these probabilities, resulting in a target preference of 0.5, i.e., the pragmatic listener does not prefer one potential referent over the other.

Does RSA predict the target preference and therefore contrastive inference in context Fig. 1b? Assuming that the presence of the contrasting orange banana does not affect the speaker's modifier production probability for the competitor corncob but does increase modifier production probability for the target banana to 0.9, renormalizing the production probabilities results in a target preference of 0.9 – thus reproducing the classic contrastive inference.

Unlike previous accounts of contrastive inference, modifier production probabilities are expected to directly drive the contrastive inference and associated target preference. Since the contrastive inference is the difference in target preference between contrast conditions and the target preference depends on the modifier production probabilities of the target and the competitor, the competitor takes on a central role in these predictions. This suggests that increasing the modifier production probabilities for the competitor should lead to a decrease in target preference. It has been established that speakers are more likely to include color modifiers in referring expressions for objects in isolation when they appear in an atypical rather than in a typical color (Rubio-Fernández, 2016; Westerbeek, Koolen, & Maes, 2015; Degen, Hawkins, Graf, Kreiss, & Goodman, 2020). Thus the atypical yellow strawberry in Fig. 1c is more likely to elicit a color modifier than the typical corncob in Fig. 1b. Assuming a modifier production probability of 0.6, this contrast-present context yields a much smaller increase in target preference compared to the contrast-absent context. In other words, the size of the target preference is predicted to be dependent on the choice of competitor in the contrast-present vs. contrast-absent conditions, keeping target typicality constant. This predicts that the size of the contrastive inference can vary depending not only on features of the target (as previously shown by Sedivy, 2003; Rubio-Fernandez et al., 2019), but also crucially depending on features of the competitor (and more generally, any other objects in the display that may plausibly elicit the relevant modifier).

To investigate this novel prediction, we first elicited modifier production probabilities (i.e., an estimate of $P_{S_1}(u|r)$) in a free production interactive reference game (Exp. 1) in contexts that varied in the presence of a contrast, the typicality of the target, and the typicality of the competitor. This allowed us to generate pragmatic listener probabilities for each display. We then evaluated model performance by comparing these predictions to empirically elicited interpretations (Exp. 2).

Experiment 1: Modifier Production in an Interactive Reference Game

The goal of Exp. 1 was to obtain color modifier production probabilities for the items in the displays ultimately used in the contrastive inference experiment (Exp. 2). In particular, we elicited production probabilities for those items that functioned as targets and competitors in Exp. 2.[4] Probabilities were elicited in a free production interactive reference game. We expected modifier production probability to be higher for atypical objects and in the presence of a contrast. For instance, we expected speakers to call a yellow banana simply *the banana*, but an orange banana *the orange banana*. We employed the elicited modifier production probabilities as the pragmatic speaker probabilities in the subsequent model evaluation.

Method

We recruited 282 participants over Amazon's Mechanical Turk, who were randomly matched to form director-matcher dyads (i.e., 141 pairs in total).

Each context included four objects, as displayed in Fig. 1. The pool of objects consisted of 10 items (e.g., broccoli), each of which could occur in a typical (green broccoli) and atypical color (red broccoli). All objects were carefully normed for color-diagnosticity (Tanaka & Presnell, 1999), typicality, and nameability. Both director and matcher saw the same four objects, but in scrambled positions. The director also saw a green border around one object which was to be described to the matcher through a chat window. The matcher's task was to click on this object.

On critical trials, participants saw critical displays from Exp. 2. The object to be communicated could be either the object that functioned as the target or the object that functioned as the competitor in that display in Exp. 2, as exemplified in Fig. 1. We continue to refer to 'target' and 'competitor' in the reporting of this experiment, terms which refer to the

[4] We assumed that the production probability of the relevant color modifier was close to 0 for the remaining distractor objects in the display and did not elicit these explicitly.

function of the object to be communicated in Exp. 2. Contexts varied in the typicality of the target and the competitor and the presence of a contrast, resulting in eight conditions. Participants saw each context exactly once. Throughout the experiment, half of the critical trials required the speaker to communicate the 'target' and in the other half the 'competitor'.

In contexts where the contrast was absent, the distinction between target and competitor was meaningless and thus one of the color competitor objects was arbitrarily coded as the target and the other as the competitor. Fillers were randomly created contexts where the 'contrast' or the 'distractor' from Fig. 1 was the object to be communicated. Overall, each dyad saw 60 contexts (32 critical trials) in randomized order.

Results

We excluded two dyads because of multiple participation and 27 dyads for primarily using playful descriptions, e.g., *should be yellow, must have teeth to eat* for the *red corn* object, which left 112 dyads for the analysis.

Fig. 2 shows the proportion of color modifier mentions for the target and competitor in each condition. We conducted a Bayesian mixed effects logistic regression predicting color mention for each item from centered fixed effects of contrast presence, target typicality, and competitor typicality, as well as random by-participant intercepts (the most complex random effects structure that allowed the model to converge).

There was strong evidence of contrast presence ($E = 5.25$, $CI = [4.82, 5.69]$), such that when a contrast to the object was present (e.g., another banana, see target proportions in the upper row in Fig. 2), participants were more likely to mention the color modifier than in the absence of a contrast (see target proportions in the lower row in Fig. 2 and competitor proportions overall). This was especially true when the object was atypical[5]. There was also strong evidence for the object's typicality ($E = 2.82$, $CI = [2.52, 3.12]$), such that participants were more likely to include a color modifier when referring to an atypical object than a typical one.

The results of this production experiment show that the probability of a speaker's modifier use is modulated by an object's color typicality, replicating previous results (Westerbeek et al., 2015). The results also confirm the assumption made by many contrastive inference studies that speakers are more likely to produce the color modifier in the presence of a contrast (Aparicio et al., 2018; Grodner & Sedivy, 2011; Sedivy et al., 1999), though this probability is modulated by the typicality of the object.

[5]A full interaction model did not converge because color was *always* mentioned in the contrast-present condition with atypical targets, which did not allow the model to generate estimates for interactions involving these conditions. We did not find evidence for any other interactions.

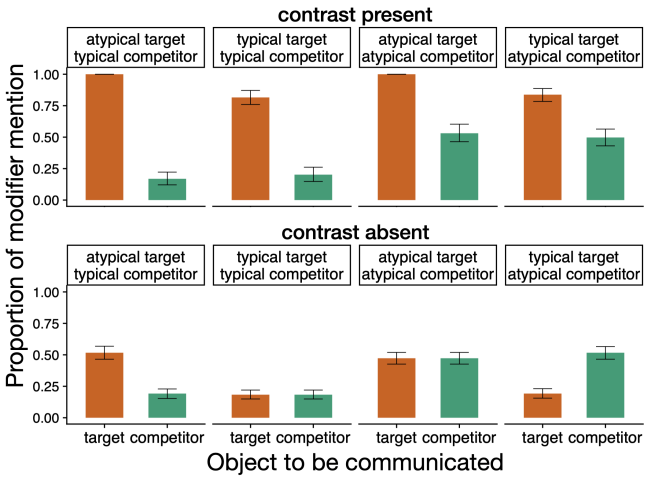

Figure 2: Proportion of modifier mentions in each condition for objects that functioned as target and competitor in Exp. 2. Error bars indicate 95% bootstrapped confidence intervals.

Experiment 2: Referential Interpretation in an Incremental Decision Task

To investigate which object listeners consider to be the most likely referent after observing the color adjective, we conducted an incremental decision task (Qing, Lassiter, & Degen, 2018). This is an offline task that allows for eliciting participants' belief distributions at multiple points in the unfolding referring expression.

Method

We recruited 239 participants over Amazon's Mechanical Turk. This experiment was a one-player comprehension-only adaptation of the production study described above and was implemented as an incremental decision task (Qing et al., 2018): Participants read sentences of the form "Click on the yellow banana", which contained a referring expression, and their task was to select the target in the display. Crucially, the sentence was only gradually revealed. Participants made a selection at each of three time points: (1) before receiving any information about the referent (i.e, after observing "Click on the", *prior window*), (2) after observing the adjective ("Click on the yellow"), *adjective window*, and (3) after observing the full referring expression with the disambiguating noun ("Click on the yellow banana"), *noun window*.

The critical displays were identical to the critical displays in Exp. 1. Target typicality and contrast presence were within-participant manipulations, competitor typicality was a between-participants manipulation[6]. All critical trials used

[6]The complexity of the 2x2x2 design and considerations of power required that either the number of trials per participant be high or one manipulation be between-participants. We decided for a smaller number of trials to minimize the probability of strategic responses or response fatigue developing over the course of the experiment. Contrast presence and target typicality could not be manipulated between-participants since these regularities are easily de-

color modified referring expressions. Filler trials were included that primarily used unmodified utterances and referred to one of the other three items in the display to avoid learning effects. Participants completed 55 trials (20 critical) in random order. To minimize the risk that the speaker was perceived as pragmatically uncooperative (Grodner & Sedivy, 2011; Pogue, Kurumada, & Tanenhaus, 2016; Ryskin et al., 2019), trials with modified utterances that referred to a typical object with no contrast only appeared after the 15th trial. To familiarize participants with the task, they first completed four practice trials in the director role.

Results

We excluded participants who participated multiple times (1), who indicated that they did the experiment incorrectly or were confused (13), whose self-reported native language was not English (6), and who gave more than 20% erroneous responses[7] (7). 211 participants remained; 108 saw atypical competitors and 103 saw typical competitors on critical trials.

Fig. 3 shows the proportion of target and competitor selections in the adjective window (lighter colors) alongside the RSA model predictions derived from the Exp. 1 production probabilities (darker colors), grouped by condition.[8] We conducted a Bayesian mixed effects logistic regression on adjective window choices, predicting the log odds of target over competitor selections from centered fixed effects of contrast presence, target typicality, competitor typicality, and their interactions, prior window selection, as well as the maximal random effects structure that allowed the model to converge[9].

There was strong evidence for an effect of contrast presence ($E = 0.34$, $CI = [0.13, 0.53]$), such that when there was a contrast object (top panels), there was a general preference for target over competitor selections, replicating the standard contrastive inference effect. This preference was largest when the target was atypical and the competitor was typical and disappeared when the target was typical and the competitor was atypical, following the qualitative predictions discussed in the modeling section above and exemplified in Fig. 1. There was also strong evidence for an effect of competitor typicality ($E = -0.54$, $CI = [-0.90, -0.17]$), such that when the competitor was atypical, target selections decreased, which is again in line with our predictions.

Although object selections in the prior window were approximately at chance, there was strong evidence that it affected participants' specific selections of their adjective window choices ($E = 1.46$, $CI = [1.29, 1.63]$). These results suggest that when participants' prior selection is congruent with the newly revealed adjectival information, they stick with their previous choice.

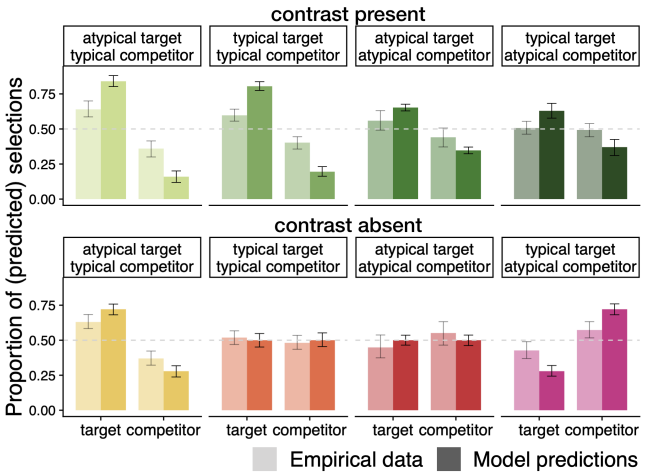

Figure 3: Empirical proportion (light bars) and model predicted probability (in darker colors) of object selections for each condition. Dashed line marks chance level of target versus competitor selections. Error bars indicate 95% bootstrapped confidence intervals.

Overall, these results suggest that the color typicality of not just the target, but of competitor objects in the display, too, affects the inferences listeners draw about the intended referent. An atypical competitor alone can promote the competitor over the target when the contrast is absent and can even make the target preference disappear when a contrast is present.

If one quantifies contrastive inference as an increased target preference in the adjective window in the contrast-present condition compared to its item-matched contrast-absent condition, the contrastive inferences is small or even non-existent when the target is atypical and the competitor typical (left column of Fig. 3). This may explain why contrastive inferences did not occur with target items of unpredictable colors (Sedivy, 2003). However, even though those items have been reported to have a higher modifier production probability in isolation (Sedivy, 2003), future work still needs to establish how those objects of unpredictable colors relate to (a)typically colored objects.

Model evaluation

Here we assess the extent to which RSA captures the comprehension data based on the empirically elicited modifier production probabilities. We assume a flat prior over all objects in the display, a choice justified by the uniform selection distribution over objects in the prior window of the comprehension experiment. The pragmatic listener probabilities assigned to the target over the competitor are then the normalized modifier production probabilities as shown in Equation (2), where r are contextually possible referents r_{target} and

tectable by a participant within an experiment. Between-participants manipulations are considered more conservative (Charness, Gneezy, & Kuhn, 2012) and random by-participant intercepts and slopes were included in the analyses to account for random by-participant variability.

[7] An incorrect response is defined as a selection of a non-target object after observing the fully disambiguating noun.

[8] Neither of the other two objects in the display was chosen after observing the adjective.

[9] Random effects: $(1 + \text{contrast} * \text{target\_typicality}|\text{participant}) + (1 + \text{contrast} * \text{competitor\_typicality}|\text{target}) + (1 + \text{contrast} * \text{target\_typicality}|\text{competitor})$

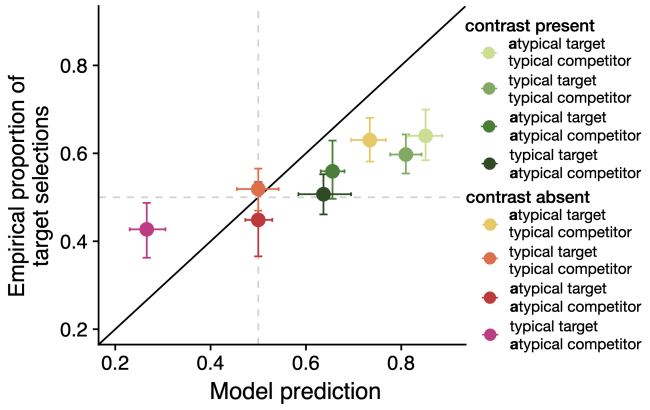

Figure 4: Empirical proportion of target selections against RSA model predictions. The dashed lines mark chance level for target over competitor selections. Error bars indicate 95% bootstrapped confidence intervals.

r_{comp}, and u the referring expression up to the contextually warranted color modifier, e.g., *the yellow*.

$$P_{L_1}(r|u) = \frac{P_{S_1}(u|r)}{P_{S_1}(u|r_{\text{target}}) + P_{S_1}(u|r_{\text{comp}})} \tag{2}$$

Fig. 3 shows the model predictions (dark bars) alongside the empirical results (light bars) for target and competitor selection in the adjective window. Using the modifier production probabilities obtained in Exp. 1, the model predicts the qualitative patterns for all the different context conditions.

Quantitatively we found strong evidence that the RSA model predicts the empirically elicited comprehension data ($E = 1.46, CI = [1.01, 1.92]$)[10] and its predictions are highly correlated with the empirical results ($r = 0.91$). However, it generally predicts more extreme probabilities than are borne out in the empirical data, as shown in Fig. 4. The model overpredicts target selections in high target preference conditions and underpredicts target selections in low target preference conditions.

One possible explanation for the mismatch between model predictions and observed target selections towards the extreme ends of the scale is that the empirically elicited contrastive inferences appear smaller due to participants' re-selection bias (as described in the results of Exp. 2). If a participant observes an adjective that could elicit a contrastive inference but the participant selected the competitor in the prior window, the re-selection bias counteracts the contrastive inference. This can explain why the range of empirical target selection proportions is compressed towards the center of the scale. We have since re-run this experiment without eliciting prior window selections, and the results support this explanation. In an eye-tracking version of the experiment, where participants do not make explicit looking decisions, the bias to continue looking at the same object may also be weaker.

Overall, these results suggest a strong connection between referring expression interpretation and production. Only using the probability of encountering the observed adjective, the RSA model can qualitatively and quantitatively predict the empirically elicited comprehension data.

General Discussion

In this paper, we tested a speaker-centric model of contrastive inference couched within the Rational Speech Act (RSA) framework. We used the model to make quantitative predictions about the behavior a pragmatic listener should exhibit in varying contexts. In contrast to previous accounts, it is not simply the modifier production probability for the target that modulates the inference (as suggested by, e.g., the default description account proposed by Sedivy, 2003), but more broadly the relative modifier production probabilities for *all* contextually relevant objects. This account shifts the focus away from specific cognitive and linguistic factors that have been discussed to affect contrastive inference and onto listener's production expectations, and away from contrastive inference narrowly to the interpretation of referring expressions more broadly.

We show that this speaker-centric model not only predicts the basic contrastive inference effect; it also provides possible explanations for why contrastive inferences are less stable with color adjectives. First, the nature of the competitor affects the behavioral patterns generally associated with contrastive inference, such that the target preference can disappear even when a contrast is present, as long as the expected modifier production probabilities are sufficiently similar for target and competitor. Second, higher modifier production probabilities for the target in contrast-absent contexts can decrease the difference in target preference compared to its otherwise matched contrast-present context.

These results also provide a challenge for accounts that would explain away variable contrastive inference behavior by pointing towards adjective semantics. The presented comprehension results show a high degree of variability in target preference within the color adjective domain, calling into question a generalizable contrastive inference pattern for color adjectives. While an adjective semantics-based account predicts greater variability between than within adjective types, a speaker-centric account predicts instead that differences in target preference and contrastive inference are mediated by a listener's expectations about how likely the modifier is to be produced.

The strong correlation between the model predictions and target preference patterns in the comprehension experiment suggests a clear connection between production expectations and pragmatic interpretation. To derive listener predictions, the presented model used empirically elicited modifier production probabilities, essentially treating the speaker model as a black box. One avenue of future work is to apply exist-

[10] Results of a Bayesian mixed effects logistic regression model: target_selection ~ RSA_prediction + (1 + RSA_prediction|participant)

ing RSA models of modified referring expression production (Degen et al., 2020), which have been successful in modeling the empirically observed redundant use of color modifiers as a function of objects' typicality, to the production data reported here.

Finally, the RSA model predicts that a listener's prior beliefs about likely referents should affect listeners' inferences in a systematic way. In other words, a listener's target preference should be greater for objects they believe the speaker is more likely to refer to a priori. Explicit prior manipulations, and extensions of the model to other adjectives are promising new avenues to further probe the RSA account of the interpretation of referring expressions.

References

Aparicio, H., Kennedy, C., & Xiang, M. (2018). Perceived informativity and referential effects of contrast in adjectivally modified NPs. In *The semantics of gradability, vagueness, and scale structure* (pp. 199–220).

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, *81*(1), 1–8.

Cohn-Gordon, R., Goodman, N., & Potts, C. (2019). An incremental iterated response model of pragmatics. *Proceedings of the Society for Computation in Linguistics*, *2*(1), 81–90.

Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., & Goodman, N. D. (2020). When redundancy is useful: A bayesian approach to "overinformative" referring expressions. *Psychological Review*.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, *24*(6), 409–436.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *SCIENCE*, *336*, 1.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829.

Grice, H. P. (1975). Logic and conversation. *1975*, 41–58.

Grodner, D., & Sedivy, J. C. (2011). The Effect of Speaker-Specific Information on Pragmatic Inferences. In *The processing and acquisition of reference* (Vol. 2327, pp. 239–272). MIT Press.

Hawkins, R. X. D., Gweon, H., & Goodman, N. D. (2018). Speakers account for asymmetries in visual perspective so listeners don't have to. *CoRR*, *abs/1807.09000*.

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*(3), 831–836.

Kao, J. T., & Goodman, N. D. (2015). Let's talk (ironically) about the weather: Modeling verbal irony. In *CogSci.* (Proceedings of the Thirty-Seventh Annual Conference of the Cognitive Science Society)

Kleinschmidt, D., & Jaeger, T. F. (2011). A Bayesian belief updating model of phonetic recalibration and selective adaptation. In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics* (pp. 10–19).

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676.

Mitchell, D. C., Cuetos, F., Corley, M. M., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, *24*(6), 469–488.

Pogue, A., Kurumada, C., & Tanenhaus, M. K. (2016). Talker-specific generalization of pragmatic inferences based on under-and over-informative prenominal adjective use. *Frontiers in psychology*, *6*, 2035.

Qing, C., Lassiter, D., & Degen, J. (2018). What do eye movements in the visual world reflect? A case study from adjectives. In *CogSci.*

Rubio-Fernández, P. (2016). How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology*, *7*, 153.

Rubio-Fernández, P., & Jara-Ettinger, J. (2018). Joint inferences of speakers' beliefs and referents based on how they speak. In *CogSci.*

Rubio-Fernandez, P., Terrasa, H. A., Shukla, V., & Jara-Ettinger, J. (2019). *Contrastive inferences are sensitive to informativity expectations, adjective semantics and visual salience.* PsyArXiv.

Ryskin, R., Kurumada, C., & Brown-Schmidt, S. (2019). Information integration in modulation of pragmatic inferences during online language comprehension. *Cognitive Science*, *43*(8), e12769.

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, *32*(1), 3–23.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147.

Tanaka, J. W., & Presnell, L. M. (1999). Color diagnosticity in object recognition. *Perception & Psychophysics*, *61*(6), 1140–1153.

Westerbeek, H., Koolen, R., & Maes, A. (2015). Stored object knowledge and the production of referring expressions: the case of color typicality. *Frontiers in Psychology*, *6*.