Production Expectations Modulate Contrastive Inference

Anonymous CogSci submission

Abstract

Listeners rapidly draw pragmatic inferences about speakers' intentions during online sentence processing. One such type of inference is the *contrastive inference*, which arises when a speaker uses an adjective as in *the yellow banana* and a listener infers, even before observing the disambiguating noun *banana*, that the speaker intends to refer to a yellow object that contrasts with another object of the same type (e.g., an orange banana), rather than to a yellow competitor object that does not participate in a contrast. This inference has been shown to be notoriously unstable – it is often observable with size adjectives, but much less frequently so with color adjectives. A unifying account of this inferential instability is lacking. Here, we argue for a highly production-centric model of comprehension couched within the Rational Speech Act framework, whereby the adjective production probabilities a listener expects for the objects in a given context drives the presence and size of contrastive inferences. The greater the asymmetry in expectation for a speaker to use a pre-nominal adjective for the target rather than the competitor, the greater the listener's resulting bias towards a target interpretation, irrespective of whether or not a contrasting object is present. Modifier production probabilities were collected in a free production study (Exp. 1) and used to make predictions about comprehension in an incremental decision task (Exp. 2). The model's fine-grained comprehension predictions are supported by the data. This account has the potential to explain the now-you-see-it-now-you-don't nature of contrastive inferences and shifts the explanatory focus away from contrastive inference narrowly and 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 must choose one of an indefinite number of possible referring expressions. *The banana*, *the yellow banana*, *the yellow, curvy fruit-thingy* for example are all utterances that can refer to the same object. At the same time, the same utterance – e.g., *banana* – can be used to refer to different kinds of objects. This flexibility poses a challenge for a listener, who needs to pragmatically infer a speaker's intended referent. Consequently, trying to understand how a listener processes these utterances and what affects their interpretation has become a central question in psycholinguistic research. [jd: last sentence too vague but maybe ok for first paragraph]

A basic fact about language is that listeners process utterances incrementally, i.e., new information is incorporated into the interpretation of the utterance as soon as it becomes available (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). For instance, eye-tracking experiments have shown that if a listener hears the incomplete utterance *the yellow* in a display like Figure 1a, they fixate on the yellow objects in the display even before they hear the disambiguating noun *banana* and [ek: arrive at] the final referent.

But 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 [ek: cite]. One such inference that has received much attention in recent years is the so-called *contrastive inference* (Sedivy, Tanenhaus, Chambers, & Carlson, 1999)[jd: include other references for studies on contrastive inferences]. Consider the context in Figure 1b that shows a yellow and orange banana, a yellow corncob and some other distractor item. When a listener is asked to *pick out 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 in increased looks, for the yellow object that has a contrast member of the same type but of a different color in the display (i.e., the banana) (Sedivy et al., 1999; Sedivy, 2003). When the contrast is absent, as in Figure 1a, listeners have no such preference. The effect of seeing an increased target preference over the competitor that is elicited by the presence of a contrast (i.e., the orange banana) is called the contrastive inference.

Broadly speaking, a contrastive inference arises because a listener expects a cooperative speaker to not be more informative than required by the context (?, ?). The presence of a contrast object makes it contextually necessary to include the adjective for establishing a unique referent. Therefore a listener should be able to infer the presence of a contrast when hearing the adjective.

This simple Gricean account that only takes into consideration the contrastive function of the adjective predicts that contrastive inferences should arise wherever and whenever the target object occurs in the presence of a contrast object. It is surprising, then, that contrastive inferences have been found to vary substantially: While it has been replicated reliably in the size adjective domain (Aparicio, Xiang, & Kennedy, 2016; Aparicio, Kennedy, & Xiang, 2018; Grodner & Sedivy, 2011; ?, ?; Sedivy et al., 1999), the effect seems to be less

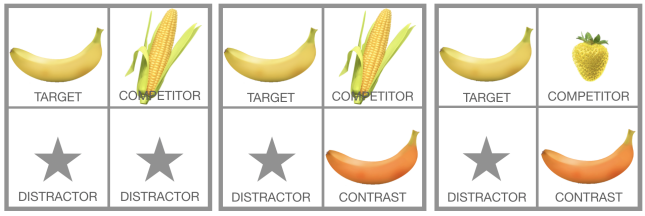

Figure 1: Three contexts, each with a yellow banana as the intended target and another yellow object as its color competitor. The competitor can be typical (in A and B) or atypical (in C), and a contrast can be absent (in A) or present (in B and C).[ek: label with A and B (and C) – possibly remove the atypical comp context]

stable with color adjectives (Sedivy, 2003). Sedivy (2003) report that the contrastive inference arises in contexts where the target object has a predictable color (such as the yellow banana in Figure 1) but not when it is replaced by an object with an unpredictable color like a cup, which comes in all sorts of colors. They suggest that these objects differ in how likely a speaker is to produce the color modifier for the object in isolation, i.e., in the absence of a contrast, a yellow banana is usually just called *a banana* while a yellow cup is often called *a yellow cup*. Sedivy (2003) call these references the *default descriptions* of the objects. They suggest that the contrastive inference only arises with objects that do not have the adjective in their default description, since this makes the observation of the adjective surprising and a listener then interprets the adjective contrastively.

In the interim, contrastive inference has been shown to be modulated by multiple factors. First of all, its occurrence seems to be dependent on the expectations of informativity, such as the presence of a contrast that makes mentioning the adjective contextually necessary (Aparicio et al., 2016, 2018; Sedivy, 2003). However, adjective semantics also seem to matter, such that a contrastive inference can arise with color adjectives, relative adjectives and maximum standard absolute adjectives, but not with minimum standard absolute adjectives (Aparicio et al., 2018). Furthermore the effect only arises when the listener considers the speaker as a reliable speaker who does not violate communicative norms (Grodner & Sedivy, 2011; ?, ?).

In this paper, we provide a novel 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 (Degen, Hawkins, Graf, Kreiss, & Goodman, 2019; ?, ?)[ek: jd: which Hawkins paper do you mean?]. We propose to formalize the relevant listener-side reasoning within the Rational-Speech Act (RSA) framework (Frank & Goodman, 2012; Goodman & Frank, 2016), 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 probability that a listener with prior beliefs and expectations about the speaker assigns to possible referents after observing partial sentences of the form *Click on the yellow*. This shifts the focus away from specific cognitive and linguistic factors that influence contrastive inference onto listener's production expectations (and their prior beliefs).

For this investigation it is important to distinguish between two concepts: *contrastive inference* and *target preference*. The target preference quantifies to what extent a listener prefers the target over the competitor interpretation in any given context. For example, if they consider both to be equally likely referents, the probability for target preference would be 50%. Even though a contrastive inference can lead to a preference of the target over the competitor, this does not need to be the case. A contrastive inference occurs when the presence of the contrast increases the target preference probability over a context where the contrast is absent. That means there can be a contrastive inference even if there is no target preference in the contrast-present condition, simply because there was a target dispreference in the contrast-absent condition.

In this paper we will first show how the RSA account can make the same predictions about the basic contrast effect as for instance the default description account (Sedivy, 2003). Furthermore we derive new predictions about the contexts that *should* affect contrastive inference and target preference. We report a free production study we conducted to elicit modifier probability estimates, which will be used to determine quantitative model predictions. For the evaluation of those predictions, we compare them to empirical comprehension data which we elicited using an incremental decision task.

Accounts of contrastive inference [ek: Model section]

[ek: "Likelihood" and "probability" are all scrambled up here. Don't forget to fix that!]

In the literature, different factors have been considered to give rise to contrastive inference which indirectly put more or less relevance onto the speaker.

Sedivy et al. (1999) for instance only considers a very limited speaker model. As mentioned above, a contrastive inference only arises when the modifier is not a component of the object's *default description*. It therefore only assumes speaker considerations as to the creation of the default descriptions for objects and is completely independent of the context the target is presented in. Other research has extended this account and found that the size of the contrastive inference can vary (Aparicio et al., 2018)[ek: Rubio-Fernandez]. [ek: Figure out how much they really consider the speaker here... actually not much]

A Rational-Speech Act account gives the speaker a central

role in the predictions, since the listener P_{L_1} determines the most likely intended referent r by directly reasoning about the speaker's possible utterances u for each object in the display C.

$$P_{L_1}(r|u,C) = \frac{P_{S_1}(u|r,C) * P(r|C)}{\Sigma P_{S_1}(u|r_i,C) * P(r_i|C)} \quad (1)$$

To simplify the following example, we will assume that listeners have a uniform prior $P(r|C)$ over all objects in the display. Then the RSA model predicts a direct relationship between the production probabilities P_{S_0} and the listener's distribution over possible referents P_{L_1}.

$$P_{L_1}(r|u,C) \propto P_{S_1}(u|r,C) \quad (2)$$

However so far in the literature, RSA has been mainly applied to the analysis of full utterances [ek: Reuben]. To receive RSA predictions for an incomplete reference such as *the yellow...*, we take P_{S_0} to simply be the underlying probability of color mention for each referent in the display. Now we can turn to the qualitative predictions this account makes.

Consider the example contexts in Figure 2. Upon hearing the modifier *yellow*, the pragmatic listener P_{L_1} will consider 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 will 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. The hypothetical modifier production probabilities are written under each item in Figure 2.[1]

If the contrasting banana is absent (Figure 2)a, speakers are equally unlikely to include the color modifier when referring to the banana (10%) and its competitor, the corncob (10%). For the pragmatic listener predictions, these probabilities are simply normalized, resulting in the target preference probability of 50% (i.e., chance). It therefore predicts that a listener will not be able to prefer one potential referent over the other.

The presence of the orange banana (Figure 2b) will not affect the listener's expectations about modifier production for the competitor (corncob). However, the presence of another banana (i.e., a contrast) makes the inclusion of the modifier now contextually necessary for a unique identification (90%). After renormalizing the production probabilities, the model predicts a target preference probability of 90%. This is a clear increase in target preference compared to the 50% in the contrast-absent context in Figure 2a. This predicts the generally reported case of contrastive inference.

In contrast to previous accounts of contrastive inference, the modifier production probabilities are expected to directly drive the contrastive inference and target preference. Since

[1] All of the following modifier production probabilities for the target and competitor are hypothetical and the empirically elicited probabilities.

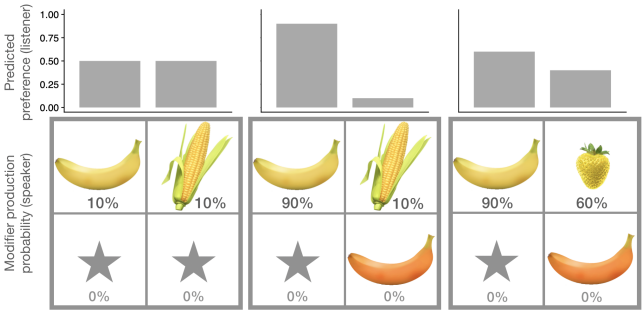

Figure 2: The three contexts from Figure 1 with expected modifier production probabilities for each of the object (the number labels in the cells) and the corresponding model predictions for the intended target after hearing *Click on the yellow...* (bar plots). [ek: still not great... any ideas?][jd: how about making it a three-tier figure: at the bottom, you plot modifier probabilities for target and competitor (ie, the numbers that are currently inside the figure, ignoring the zero-probabilities). in the middle, the predicted listener preferences, like you already have. and at the top you have the three contexts (with label "Contexts" instead of "Modifier production probability", and "target", "competitor" etc labels reintroduced). this also allows you to remove the figure that has only the contexts (currently fig 1). ie essentially you're combining figs 1 and 2. also: increase the size of the axis labels!]

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 raises the question of what happens if we change the modifier production probabilities for the competitor. It has been established that speakers are more likely to include the color modifier for objects in isolation when they appear in an atypical than in a typical color (?, ?, ?). Therefore the yellow strawberry in Figure 2c receives a higher modifier production probability (60%) than the corncob competitor in Figure 2a (10%). With the atypical competitor, this contrast-present context shows a much smaller increase in target preference probability to the contrast-absent context. In other words, there is a smaller contrastive inference than in the context with the typical competitor (corncob). This predicts that the size of the contrastive inference can change, dependent not only on the target in the display, but crucially also the color competitor.

[ek: mention somewhere that it differs from default description account because it considers production probabilities which are not reduced to a binary elicitation]

To investigate when a listener with a generative speaker model **should** draw a contrastive inference, we need to manipulate and elicit how likely a listener can expect a modified over an unmodified referring expression for each object in the display. To evaluate the performance of the model and to gain information about the prior of the objects, we also need com-

prehension data that informs us which object is considered the most likely target referent.

Experiment 1: Modifier Production in an Interactive Reference Game

Experiment 1 was aimed at obtaining modifier production probabilities for all the displays ultimately used in the contrastive inference experiment (Experiment 2). We elicited these probabilities in a free production interactive reference game. We expect that the typicality of a color for an object will affect these modifier production probabilities, i.e., we expect speakers to call a yellow banana simply *the banana*, but an orange banana *the orange banana*. We take the results as the modifier production probabilities a listener can expect for each object.

Participants

We recruited 282 participants over Amazon's Mechanical Turk, who were randomly matched to form listener-speaker chat pairs (i.e., 141 pairs in total). Each participant was paid \$2.30 (approximately \$11-\$14/hr)[2]. We restricted participation to workers with IP addresses in the US and an approval rate of previous work above 97%. We excluded 29 participant pairs because of multiple participation and for primarily using unnatural descriptions such as *should be yellow, must have teeth to eat* for *red corn*.

Material

Each context included four items, as displayed in Figure 3. The pool of items consisted of 10 types (banana, broccoli, carrot, corn, egg, lettuce, pumpkin, strawberry, swan, tomato), each of which could occur in a typical and atypical color. For example, the broccoli could occur in its typical color green or in the atypical color red. The resulting pool contained 20 items, 10 of which were atypically colored. The number of colors were counterbalanced such as each color occurred twice as a typical and twice as an atypical instance. All items were carefully normed for color-diagnosticity (?, ?), typicality and nameability.

Design

The contexts varied in the typicality of the target, the typicality of the competitor and the presence of a contrast, resulting in eight conditions. We needed to elicit the modifier production probabilities for the target and the competitor. In contexts where the contrast was absent, this distinction is irrelevant. For example, when target and competitor are both (a)typical, either could be underlyingly coded as the target. Similarly, the modifier production probability for a typical target with an atypical competitor is the same as the probability for a typical competitor in a context with an atypical target. In contrast-present contexts, the target-competitor distinction matters, which is why speakers had to communicate the competitor half of the trials and the target in the other half.

The fillers were eight randomly created contexts where the contrast or the distractor had to be communicated. Overall, each participant saw 60 different contexts (32 critical trials) in a completely randomized order.

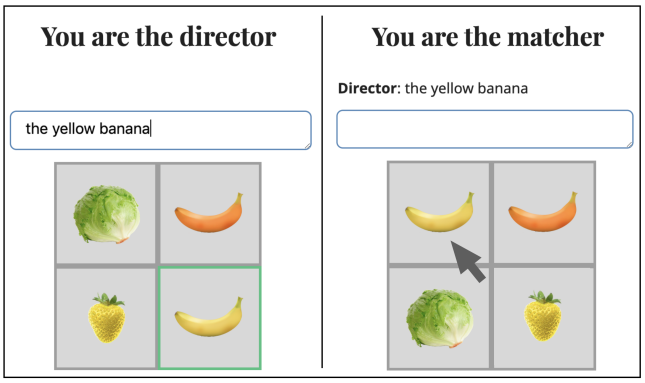

Figure 3: Example display for the interactive reference game (Experiment 1). Both, the speaker (here *Director*) and listener (*Matcher*) see the same four objects but in a scrambled order. Additionally, the speaker sees a green border around one of the objects, marking the intended target which the listener needs to select.

Procedure

Participants were randomly paired up and each was randomly assigned either to the role of a speaker or listener. They could communicate freely through a real-time multi-player interface similar to [ek: Hawkins (2015)]. The speaker was instructed to communicate a target object out of a four-object context to the listener. The target could be identified by a green border surrounding it. The speaker and the listener saw the same set of objects but in a randomized order to avoid trivial position-based references such as "the left one". After the listener clicked on the presumed target, both the speaker and listener received feedback about whether the right object had been selected.

Results

Figure 5 shows the probability of color modifier mention for the target and competitor in each condition[3].

When a contrast to the target is present (e.g., another banana), a speaker needs to include the color modifier to fully disambiguate the two items (see the upper row in Figure 5). When the target was atypical, speakers always used the color modifier, while this was not always the case for a typical target.

When the contrast was absent (see the lower row in Figure 5), speakers were more likely to include a color modifier

[2]The experiment was preregistered on [ek: id]. Originally, we recruited 68 participants and then ran a follow-up with 214 more to get enough data for the evaluation of the RSA model. The results from the first 68 participants do not differ from the full data set, which is why we present them collapsed.

[3]Note that some data is duplicated in the conditions where the contrast is absent (as described in *Design*)

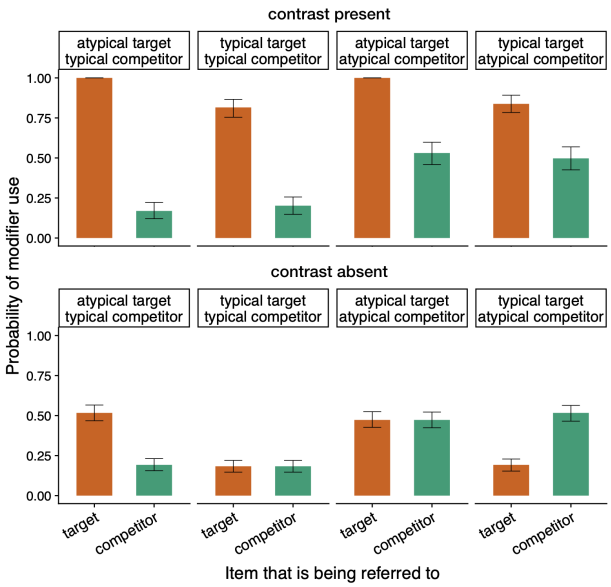

Figure 4: Results of the production study, showing the probability of modifier use for the target and competitor in each condition. Error bars are 95% bootstrapped confidence intervals.

when referring to an atypical target than a typical one.

Independent of contrast, speakers were more likely to include the color modifier for an atypical color competitor over a typical one [ek: stats?].

The results of this production experiment show that the probability of a speaker's modifier use is modulated by the color typicality of the item and the presence of a contrast. Our experiment therefore successfully manipulates the modifier production probabilities a listener can expect in different contexts.

Comprehension Experiment: An Incremental Decision Task

To investigate which objects listeners consider to be the most likely referent after observing the color adjective, we conducted an incremental decision task (?, ?). This is an offline task to investigate a listener's belief about the intended referent while gradually unfolding the referring expression.

Participants

We recruited 239 participants over Amazon's Mechanical Turk, 121 of which saw atypical color competitors and 118 saw typical color competitors in the critical trials[4]. Each of them were paid $1.80 for their participation (10-16/hr). We restricted participation to workers with IP addresses in the US

[4]The experiment was preregistered on [ek: id]. Originally, we recruited 80 participants and then ran a follow-up with 140 more to get enough data for the evaluation of the RSA model. The results from the first 80 participants do not differ from the full data set, which is why we present them collapsed.

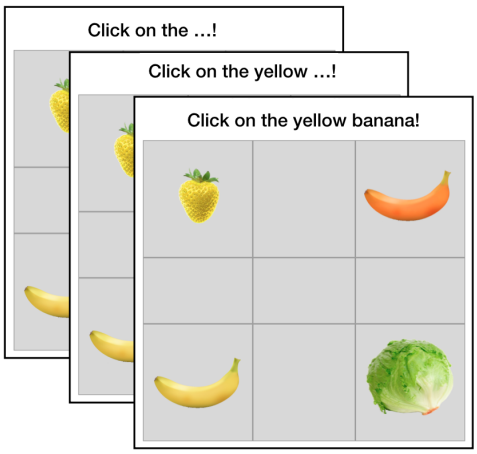

Figure 5: Design of the incremental decision task. The referring expression was placed above the grid and was gradually revealed to the participant. After each new word, the participants had to make a selection, indicating their best guess about the intended target.

and an approval rate of previous work above 97%. 27 participants were excluded because they indicated that they did the experiment incorrectly, English was not their native language, or they gave more than 20% erroneous responses[5]. 211 participants remain, 108 of which were in the atypical competitor and 103 were in the typical competitor condition.

Material

The item pool is the same as in the production study (Experiment 1).

Design

Participants completed 55 trials in total, 20 of which were critical trials and 35 were fillers. The contexts varied for each participant with respect to the presence of a contrast and the target's color typicality (within-subject manipulation). Participants were randomly assigned to see either typical or atypical competitors on critical trials (between-subject manipulation). All critical trials included color modified utterances. To avoid learning effects, we included filler trials with unmodified referring expressions and with referents other than the targets of the critical trials.

Procedure

This experiment is a one-player adaptation of the production study explained above and follows the design of an incremental decision task (?, ?).

All participant were assigned the role of the listener, which means that they needed to identify which object was the target given a referring expression placed above the context. Crucially, the referring expression was only gradually revealed

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

and participants had to choose an object each time before the trial continued. In each critical trial, three choices had to be made: (1) before receiving any information about the referent (i.e, after observing "Click on the"), (2) after receiving the adjective ("Click on the yellow") and (3) after receiving the full referring expression with the disambiguating noun ("Click on the yellow banana!").

To center the position of the mouse after each selection, a button appeared in the center of the grid which had to be clicked to reveal the next word or to advance to the next trial.

Trials were randomized with the only restriction that modified utterances that referred to a typical object with no contrast only appeared after the 15th trial to minimize the risk that the speaker was perceived as unreliable (Grodner & Sedivy, 2011).

Before participants proceeded to the main trials, they had to complete four practice trials constructed from the speaker perspective, which were introduced to familiarize the participants with the task.

Results

Figure 6 shows the proportion of object selections before reading the adjective (lighter colors) and after the adjective (darker colors), grouped by context condition. Before an adjective is observed, all items should appear equally likely to be the target, which is supported by the generally uniform distribution in all conditions. After the adjective is revealed (darker colors), only the target and competitor are legible options and we predicted that the presence of the contrast and the typicality of the objects will affect the listeners' object choices.

When the contrast is present (upper row in Figure 6), there is a general preference for target over competitor selections. This preference is biggest for the case when the target is atypical and the competitor is typical and disappears for when the target is typical and the competitor is atypical.

When the contrast is absent (lower row in Figure 6) and target and competitor differ in typicality, there is a preference for the item with the atypical color. When the two items share their typicality, the selection are approximately at the same rate.

Qualitative post-hoc analyses revealed that the selections after reading the adjective were also affected by the participant's previous selection. A participant who previously selected the competitor was more likely to select the competitor again than switch to the target (and vice versa). But since the object selections before the adjective occurred are uniformly distributed, any patterns that appear after the adjective cannot be an artifact of the reselection bias.

These results clearly show that the color typicality of the objects in the display affect the inference 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. It is therefore highly relevant to control for the quality of the competitor when assessing contrastive

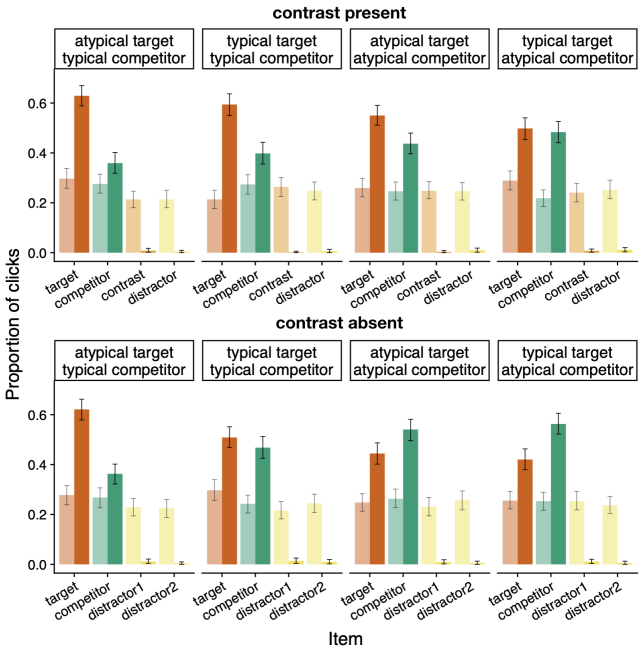

Figure 6: Results for the comprehension study, showing the proportion of selections for each item in the display and each condition. The bars in lighter colors indicate the selections before, the darker bars are the selections after the adjective was observed. Error bars are 95% bootstrapped confidence intervals. [ek: make bars and text bigger]

inferences. Choosing an atypical target makes the contrast-present and contrast-absent conditions more similar, suggesting a smaller contrastive inference. This replicates the finding that the contrastive inference did not appear with items of unpredictable colors (Sedivy, 2003).

Model evaluation

To assess the relationship between the modifier production probabilities and the comprehension data in the simplest way, we will assume a flat prior over all objects in the display. This choice is further justified by the uniform distribution over all objects before receiving any information about the target in the comprehension experiment. The probabilities to choose the target over the competitor are then the normalized modifier production probabilities as shown in Equation (3), where r is the possible referent, u the utterance and C the specific context.

$$P_{L_1}(r|u,C) = \frac{P_{S_0}(u|r,C)}{P_{S_0}(u|r_{target},C) + P_{S_0}(u|r_{comp},C)} \quad (3)$$

Figure 7 shows the model predictions (in blue) and the empirical results (in purple) for target and competitor selection after observing the adjective.

Using the modifier production probabilities obtained in Experiment 1, the model qualitatively predicts the patterns for the different context conditions.

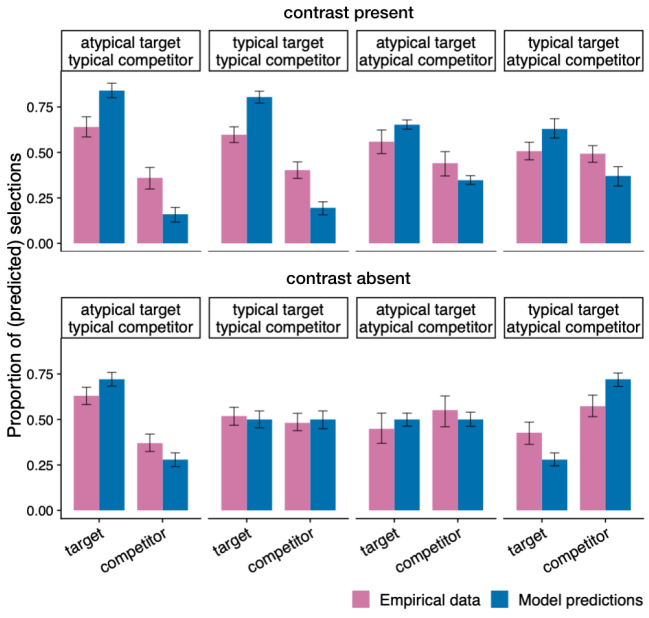

Figure 7: This is a figure.

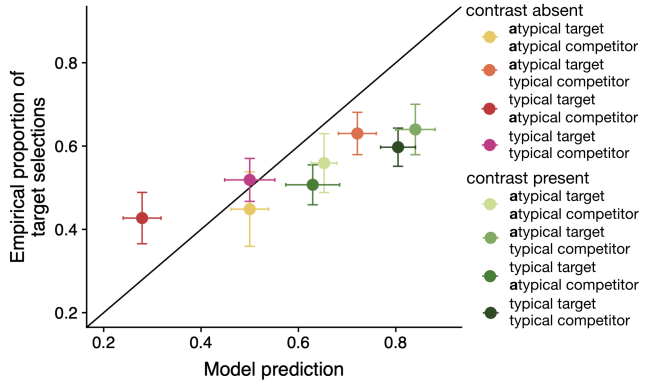

Figure 8: This is a figure.[ek: put contrast-present up]

Quantitatively the RSA model is a significant predictor for the empirically elicited comprehension data [ek: stats] and its predictions highly correlate with the empirical results ([ek: stats: r=]). However, it generally predicts stronger inferences than borne out in the empirical data, as shown in Figure 8. The model overpredicts target selection for high target preference conditions and underpredicts the target selection in low target preference conditions.

One possible explanation is that the inferences appear smaller in the comprehension study, especially because of the bias to reselect the previously selected object (as described in the results of Experiment 2). If a participant receives an adjective that is supposed to elicit a contrastive inference, but the participant has clicked the competitor before, the non-switching bias will counteract the contrastive inference choice. This can explain why the preference of the target over the competitor is flatter than expected and might be an artifact of the offline incremental decision task. Since an online eye-tracking experiment does not rely on incremental selections, we expect this bias to decrease and therefore the inferences to become stronger.

Overall, these results suggest a strong interconnection between referring expression interpretation and production. Only using the probability of encountering the utterance itself, the RSA model can qualitatively and quantitatively predict the empirically elicited comprehension data. Although we replicate that the contrastive inference can be elicited in an offline incremental decision task, the model results suggest that the selection biases in the paradigm might reduce the size of the inferences. We expect this bias to reduce in an eye-tracking paradigm, which is an immediate future direction for this work.

Discussion

In this paper, we provided a novel account of contrastive inference in which we argue for a speaker-centric model of comprehension. We use the Rational-Speech Act (RSA) framework to make quantitative predictions about the behavior a pragmatic listener *should* exhibit when provided with different contexts. This account shifts the focus away from specific cognitive and linguistic factors that have been discussed to affect contrastive inference in the literature onto listener's production expectations (and their prior beliefs). We show that this speaker-centric model cannot only account for the general case of contrastive inference, but makes further predictions onto why contrastive inference appears to be less stable with color adjectives.

In contrast to previous accounts, it is not simply the production probability of the target that matters (as suggested in a default description account (Sedivy, 2003)), but instead the relative modifier production probability of *all* objects in the display, which are then evaluated against each other. In this particular case this means that it assigns a central role to the color competitor in the display. The empirical results confirm that the choice of the competitor affects the interpretation of the utterance, providing evidence for this highly pragmatic account of comprehension.

We assessed the model predictions by collecting object selections in an incremental decision task (?, ?), replicating that it generally can elicit the contrastive inference. Crucially the results show high variation between context conditions dependent on the typicality of the target and competitor. In other words, by varying the modifier production probabilities, we can make the contrastive inference appear strong and almost make it disappear. This range provides a challenge for accounts that argue for a uniform quality about adjective semantics that affect contrastive inference. A speaker-centric account predicts instead that the general differences observed between different types of adjectives is in fact mediated by how likely a listener expects the modifier to be produced and an interesting new avenue for future investigations.

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). Springer.

Aparicio, H., Xiang, M., & Kennedy, C. (2016). Processing gradable adjectives in context: A visual world study. In *Semantics and linguistic theory* (Vol. 25, pp. 413–432).

Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., & Goodman, N. D. (2019). 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, May). 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.

Grodner, D., & Sedivy, J. C. (2011). 10 the effect of speaker-specific information on pragmatic inferences. In *The processing and acquisition of reference* (Vol. 2327, pp. 239–272). MIT Press.

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.