Probabilistic weighting of perspectives in dyadic communication

Rachel Ryskin (ryskin@mit.edu)
Department of Brain & Cognitive Sciences
Massachusetts Institute of Technology

Suzanne Stevenson (suzanne@cs.toronto.edu)
Department of Computer Science
University of Toronto

Daphna Heller (daphna.heller@utoronto.ca)
Department of Linguistics
University of Toronto

Abstract

In successful communication, speakers tailor their language to the context and listeners make inferences about the speaker's knowledge. Several current accounts propose that both speakers and listeners accomplish this by rational analysis of the statistics in the environment, including their partner. Here we examine perspective-taking behaviour in a dyadic conversation task, where the same individuals act in the role of both speaker and listener. We model perspective-taking in both production and comprehension, taking into account the dyadic situation. Our findings suggest that conversational partners weight their own perspective more than the partner's when speaking, and the partner's perspective more than their own when listening. We also find that in both production and comprehension, conversational partners change the weighting of perspectives over time, moving towards relying more on the partner's perspective at the expense of their own perspective. Surprisingly, we find little evidence that listeners or speakers adapt to the idiosyncratic statistics of their partner's linguistic behaviour.

Keywords: perspective-taking; pragmatic inference; dyadic communication; common ground; reference.

Introduction

The goal of a typical conversation is to exchange information. This is enabled by the fact that different individuals have different knowledge and beliefs. Yet this asymmetry also presents a challenge, as it requires interlocutors to tailor the message to their partner's perspective in order for it to be understood. Largely focused on reference, much research has demonstrated that speakers tailor referential forms to the knowledge of the listener (e.g., Isaacs & Clark, 1987; Wilkes-Gibbs & Clark, 1992), and that listeners are sensitive to the knowledge of their speaker (Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Hanna, Tanenhaus, & Trueswell, 2003; Heller, Grodner, & Tanenhaus, 2008; Nadig & Sedivy, 2002). At the same time, much work has shown that interlocutors also consider their own perspective, both in production (e.g., Lane, Groisman, & Ferreira, 2006) and in comprehension (e.g., Keysar, Lin, & Barr, 2003).

Recent work has proposed that the *combination* of influences from *both* perspectives is, in fact, what underlies perspective-taking behaviour (Heller, Parisien, & Stevenson, 2016; Mozuraitis, Stevenson, & Heller, 2018). This model is in line with a larger trend in pragmatics to explain aspects of communication as probabilistic inference over the statistics of the context (Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013). How speakers and listeners select the relative weighting of the two perspectives remains an open question.

Here we reanalyze data from a dyadic communication task collected in Ryskin, Benjamin, Tullis, and Brown-Schmidt (2015) using the multiple-perspective model (Heller et al., 2016; Mozuraitis et al., 2018). This dataset has several unique properties that allow us to ask novel questions. First, it comprises data from 152 participants which is large compared to most interactive studies (usually 20-60 participants), thus providing a better test of the probabilistic framework.

Second, two naive participants *took turns* as speaker and listener. The fact that all exchanges were unscripted, and thus included some errors, provides a more ecologically-valid representation of perspective-taking behaviour. Because individuals acted as both speaker and listener, this data set allows us to compare – for the same individual – perspective-taking behaviour in production and comprehension. If weighting is tied to individuals, the same individual will show the same weighing across production and comprehension. However, it is also possible that production and comprehension do not correlate, because speakers and listeners are subject to different pressures.

Finally, the dyadic structure of the task allows us to ask whether interlocutors adapt to the specific linguistic behaviour of their partner. Indeed, previous research has demonstrated that speakers and listeners adapt to similar statistical properties. For example, listeners who receive instruction from an unreliable speaker do not draw the same inferences as those who interact with a reliable speaker (Grodner & Sedivy, 2011; Ryskin, Kurumada, & Brown-Schmidt, 2019). However, these findings have been restricted to experiments where the speaker and listener have access to the same information. When a speaker and listener have mismatching information, any infelicitous language may instead be attributed to differences in perspective, leading to less adaptation.

The dataset: Ryskin et al. (2015)

The dataset comprises 152 participants, paired in 76 dyads, and engaged in a referential communication task. Conversational partners, seated in front of separate computers

and communicating via speakers, took turns instructing each other about which object to click on out of an array (see Fig. 1) over the course of 288 trials[1]. Each partner saw 7 objects in a "cubby hole" display : 6 objects had a white background, meaning that they were also visible to the other partner, one object had a grey background, meaning that it was visible to them alone, and one was "hidden" behind a black square, indicating that the partner had an object in that location.

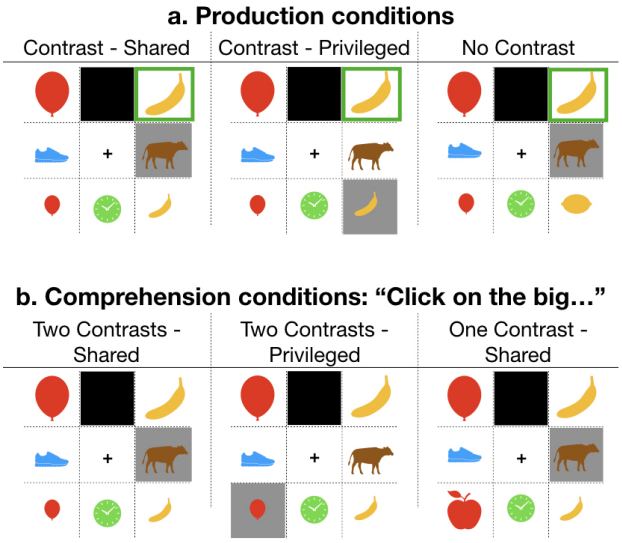

Figure 1: Stimuli from Ryskin et al. (2015) production (a) and comprehension (b) conditions. In each display, the target (e.g., banana) and competitor (e.g., balloon) were phonological onset competitors; this was done in order to increase the ambiguous portion of the speech stream.

During production, speakers provided an instruction to their partner to click on the object marked with a green frame (Fig. 1a). There were three types of critical trials (N=64 per participant): (i) No Contrast (n=16): The target object was the only one of its nominal category, and thus a bare noun would suffice (e.g., *"the banana"*); (ii) Contrast-Shared (n=32)[2]: the display contained two objects contrasting in size, which meant that size information was required (e.g., '*"the big banana"*); (iii) Contrast-Privileged (n=16): the second object was only visible to the speaker — this was the critical test case for perspective-taking.

[1]96 production trials [64 critical + 32 other] + 96 comprehension trials [64 critical + 32 other] + 48 production fillers + 48 comprehension fillers = 288 total trials. Production trials for one dyad member serve as comprehension trials for the other member and vice versa. 64 trials were critical for both members of the dyad: labeled as Contrast-Shared in Production and Two Contrasts - Shared in Comprehension (see Figure 1). "Other" trials are ones which are only useful insofar as they provided the stimulus for the partner.

[2]Unlike other trial types, Contrast-Shared/Two Contrasts-Shared trials were critical both for the producer and the comprehender. Their number was 2x the other trials to allow for certain analyses that are relevant to the current paper.

On average, speakers produced size adjectives on 8% of No-Contrast trials and 98% of Contrast-Shared trials: these provide a baseline of modification behaviour as a function of the absence or presence of contrast. The Contrast-Privileged trials are the critical test of the combination of the speaker's and listener's perspectives: including size information reflects influences from the speaker's own perspective, whereas producing a bare noun reflects influence of the listener's perspective. On average, participants modified on 66% of these cases—an intermediate behaviour between the two extremes.

During comprehension, listeners had to click on an object given a referring expression from their partner. In all three types of critical trials (N=64 per participant) (Fig. 1b), the intended referent was an object in a contrast set, e.g., the big banana, with the appropriate instruction, *"Click on the big banana"*: (i) In the One Contrast - Shared condition (n=16), upon hearing *"the big"*, listeners should expect the speaker to refer to the big banana, because it is the only object with a size contrast (see Sedivy, Tanenhaus, Chambers, & Carlson, 1999); (ii) In the Two Contrasts - Shared condition (n=32), the fragment *"the big"* should be ambiguous between the big banana and the big balloon, because both have a size contrast (cf. Heller et al., 2008); (iii) The Two Contrasts - Privileged condition (n=16) was the critical test of perspective-taking: here there was a second small object (e.g., small balloon) which was only visible to the listener. The influence of the listener's own perspective will lead to a pattern of ambiguity, parallel to Two Contrasts - Shared, whereas the influence of the speaker's perspective will lead to earlier bias towards the target, parallel to One Contrast - Shared.

Interpretation was measured using eye-movements, which were recorded for both partners. Trials with an incorrect instruction or an incorrect response were excluded ($\sim 2\%$ of trials). The timing of the adjective and the noun in the speech stream were marked, and eye movements were aligned to those onsets. We used the same analysis window as Ryskin et al. (2015): from 200ms after adjective onset to 800ms after noun onset (200ms for oculomotor delay + 600ms for the average noun).

Fig. 2 plots proportion of fixations over time. For analysis purposes, fixation durations were binarized (any duration > 0 was coded as 1), in order to provide a proportion measure analogous to the production measure. Logistic mixed-effects regression (with random slope for condition, by participant and by item) indicates, first, that listeners were more likely to fixate the target object in the One Contrast condition compared to Two Contrasts Shared ($\beta = 0.54, SE = 0.13, p < 0.001$): this establishes the difference between no ambiguity and temporary ambiguity. Importantly, listeners were also more likely to fixate the target object in the Two Contrasts Privileged condition compared to Two Contrasts Shared ($\beta = 0.29, SE = 0.13, p < 0.05$): this is the effect of perspective-taking, reflecting the influence of the partner's perspective. There was no evidence that these fixation patterns changed systematically over the course of the experi-

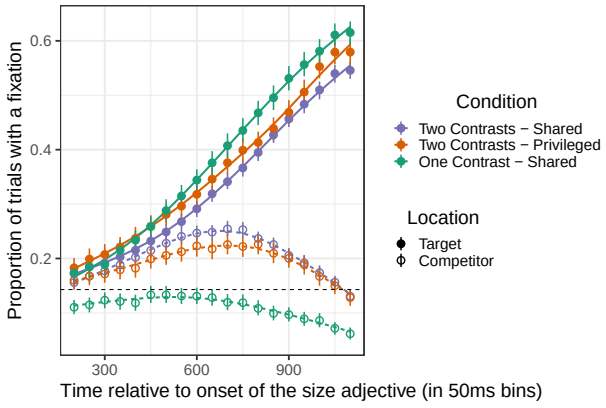

Figure 2: Timecourse of eye gaze to target and competitor objects over the course of the critical time window (onset+200ms to ∼ onset+1300ms), by condition.

ment (interactions with trial order $ps > 0.1$).

In what follows, we model the empirically observed behaviors of speakers and listeners during the perspective-taking task described above within the perspective-weighting framework (Heller et al., 2016; Mozuraitis et al., 2018). We first use each speaker's production data to estimate the weighting of their own and their partner's perspectives. We then estimate the weighting parameter for comprehension on the basis of individual listeners' eye-tracking data. In the third section, we ask whether partners in a dyad influence each other's perspective-weighting.

Perspective-weighting in production

We model reference production following Mozuraitis et al. (2018) as the probability of a particular referring expression, RE, being used to refer to an object, obj:

$$P(RE \mid obj) = \sum_{d \in D} P(RE \mid obj, d) P(d) \qquad (1)$$

where $D = \{s, \ell\}$ is the set of all available referential domains (i.e., perspectives on the objects in the situational context that are relevant to the selection of referring expressions); s is the speaker's perspective, and ℓ is the listener's perspective. Because s and ℓ exhaust the space of D, we can let $\alpha_s = P(d = s) = 1 - P(d = \ell)$ and rewrite Eqn. 1 as:

$$\begin{aligned} P(RE \mid obj) = & \alpha_s P(RE \mid obj, d = s) \\ & + (1 - \alpha_s) P(RE \mid obj, d = \ell) \end{aligned} \qquad (2)$$

The weight α_s represents how much the speaker's *own* perspective is weighed – the rest of the weight is given to the listener's perspective. Thus, when α_s is closer to 0, this reflects a speaker who weighs their own perspective less (they consider the listener's perspective more), whereas when α_s is closer to 1, this reflects a speaker who is more egocentric (they rely on the listener's perspective less). Talking about the weighting in terms of α_s is an arbitrary choice we make for expository reasons: it is equally possible to talk instead about the weight assigned to the listener's perspective (which we will do when modelling comprehension).

We first compute an aggregate α_s for the critical condition Contrast - Privileged, computing the weight that results in $P(RE = \text{“}big...\text{”} \mid target) = 0.66$ (the average modification rate in this condition). To this end, we use the production behaviour from the two control conditions: Contrast - Shared representing the speaker's own perspective – $P(RE = \text{“}big...\text{”} \mid target, d = s) = 0.98$, and No Contrast representing the listener's perspective – $P(RE = \text{“}big...\text{”} \mid target, d = \ell) = 0.08$. Solving for α_s in Eqn. 2 gives us $\alpha_s = 0.64$, suggesting that, in aggregate, speakers considered their own perspective somewhat more than the listener's perspective.

More interestingly, we consider α_s *separately* for each individual speaker, solving Eqn. 2 for each individual using their production behaviour. We first estimated α_s for each participant by considering their behaviour over the whole experimental session. Fig. 3 plots the distribution of α_s for 134 speakers: this reveals that there is much variability across different speakers (values from 18 participants were excluded because α_s exceeded 1 or was negative, likely due to noise). In addition, we compared α_s for the first and second halves of the session, finding that it decreased over time (first half $\alpha_s = 0.73$ vs. second half $\alpha_s = 0.57$; paired $t = 6.53, p < 0.001$). This indicates that speakers grew to consider their partners' perspective *more* over time, perhaps due to their own experience in the listener role.

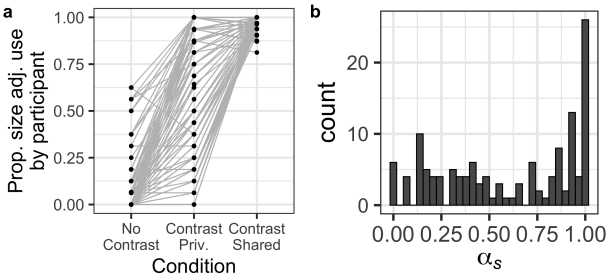

Figure 3: Proportion of size adjective use by condition for individual participants (a) and distribution of individual speakers' α_s values (b).

Perspective-weighting in comprehension

Reference comprehension asks a different question: what is the probability that an object, obj, is the intended referent, given a referring expression RE. Unlike in production, the probability here cannot be directly estimated from a listener's exposure to language, and so Bayes' rule is used to rewrite it as in Eqn. 3 (Heller et al., 2016; Mozuraitis et al., 2018).

$$P(obj \mid RE) \propto \sum_{d \in D} P(RE \mid obj, d) P(obj \mid d) P(d) \qquad (3)$$

As in Heller et al. (2016), the prior $P(obj \mid d)$ is estimated based on theoretical considerations: we assume occluded objects are considered by the listener less likely to be a referent and set $P(occluded\ obj \mid d) = 0.05$). The remaining probability mass is distributed equally among non-occluded objects. Because both domains – or perspectives – d have 7 such objects, we set $P(visible\ obj \mid d) = 0.136$.

The likelihood $P(RE \mid obj, d)$ is estimated from the production data, based on the assumption that these probabilities arise—in the mind of a listener—from experience with speakers' productions. The Contrast - Shared condition serves as a proxy for how likely speakers are to use a size adjective to describe a target object when it's in a pair (0.98). We estimate how likely speakers are to use a size adjective when the target object is a singleton using *both* the No Contrast and Contrast - Privileged conditions: this is because both are singletons from the listener's perspective. Given that listeners experience 80 trials where the referent is a singleton, including 48 fillers, 16 No Contrast trials, and 16 Contrast - Privileged trials, we set this value to 0.196 ($(48*0.08+16*0.08+16*0.66)/80 = 0.196$). Fig. 4 summarizes these distributions for each domain.[3]

Figure 4: Probabilities of referring expression with a size adjective for each object and domain.

Parallel to production, $D = \{s, \ell\}$ is the complete set of domains (or perspectives), so Eqn. (3) can be written as:

$$P(obj \mid RE) \propto (1-\alpha_\ell) P(RE \mid obj, d=s) P(obj \mid d=s) + \alpha_\ell P(RE \mid obj, d=\ell) P(obj \mid d=\ell) \quad (4)$$

To focus on the partner who is processing the language, here we set $\alpha_\ell = P(d=\ell)$ which encodes how much the *listener's* perspective is weighed (in production, we talked about $\alpha_s = P(d=s)$). This change in focus allows for an easy comparison of α_s, whereby a higher weight of α always indicates greater emphasis on one's own perspective, across both production (α_s) and comprehension (α_ℓ) (α is how much one weights one's own perspective).

The estimates of $P(obj \mid d)$ and $P(RE = \text{"}big...\text{"} \mid obj, d)$ are used to predict $P(obj \mid RE = \text{"}big...\text{"})$, as in Eqn. (4), for both targets and competitors across the range of possible α_ℓ values, as shown in Fig. 5. Recall that an α_ℓ closer to 0 means less egocentricity (because there is more weight on the speaker's perspective), whereas an α_ℓ closer to 1 means more egocentricity (more weight on the listener's perspective).

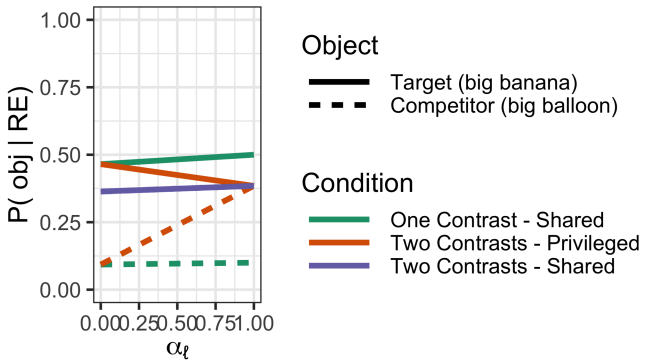

Figure 5: Model predictions for comprehension across a range of α_ℓ.

For all values of α_ℓ, participants are predicted to consider the target much more than the competitor in the One Contrast - Shared condition (green) and to consider them equivalently in the Two Contrasts - Shared condition (purple: the two lines in Fig. 5 completely overlap). In both of these conditions, the difference between the speaker's and listener's perspective is not relevant to the probability of considering the target. The critical case is the Two Contrasts - Privileged condition. Here, participants are predicted to consider the target more than the competitor for *all* values of $\alpha_\ell < 1$. When $\alpha_\ell = 1$, listeners consider only their own perspective, which means considering the target and competitor equally, and when $\alpha_\ell = 0$, listeners consider only their partner's perspective, which means considering the target almost exclusively.

How do these predictions map onto the eye-tracking data? Overall, the pattern of gaze from Fig. 2 are consistent with the model predictions for all values of $0 < \alpha_\ell < 1$. This indicates that, in aggregate, listeners considered both their own and the speaker's perspectives during comprehension. To estimate this relative weighting, we examine how models with different values of α_ℓ map onto the eye-tracking data.

Before we proceed, it is important to note that there is no agreement in the literature on how visual attention links to reference resolution, namely $P(obj \mid RE)$. For example, when an object is selected as the referent, it is not fixated at 100%, and when objects are inconsistent with the refer-

[3]The probability for the unknown object in the speaker's domain (question mark in Fig. 4) is estimated from the production data based on the assumption that it could be either a big pair member, a small pair member, or a singleton ($0.25*0.98+0.25*0.00+0.5*0.196 = 0.343$).

ring expression, they may nevertheless receive some visual attention. Nevertheless, we chose to use an *identity* linking function between gaze and model predictions, as this choice avoids making any further assumptions. We return to this issue in the discussion.

For the analyses that follow, dyads in which one of the members had an improper α_s were excluded, leaving data from 59 dyads (118 participants). Simulations across a range of possible α_ℓ values indicate that a model with an $\alpha_\ell = 0.27$ minimizes mean squared error for predicting the observed eye-fixation proportions of individual listeners (in the time window described above), averaged by condition and object. The model with $\alpha_\ell = 0.27$ (Fig. 6a; $R^2 = 0.55$) provides a better fit to eye gaze than a model which uses the weighting we estimated from the production data, namely $\alpha_\ell = 0.64$ (Fig. 6b; $R^2 = 0.50$; $BF_{10} = 5.5e15$ by BIC approximation[4]). This result suggests that interlocutors may weight their own perspective more during speaking than during listening, although this conclusion requires further investigation due to the different ways in which these weights are estimated.

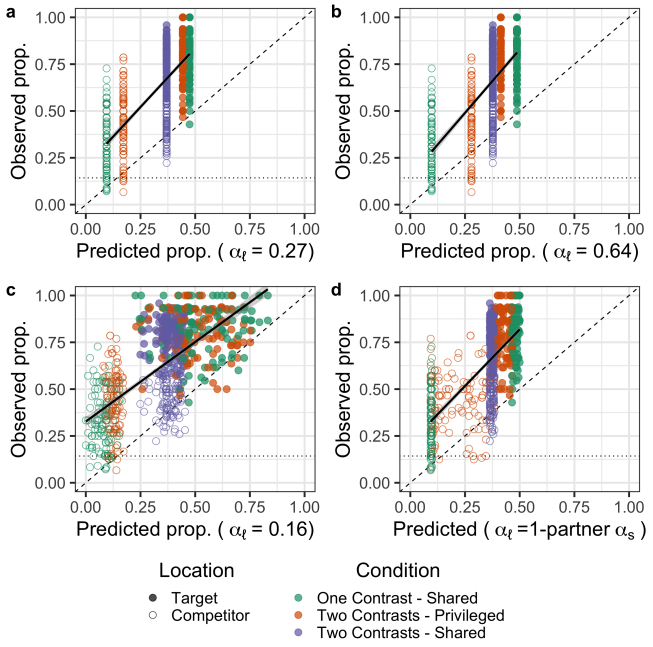

Figure 6: Observed proportion of eye fixations to target and competitor objects by condition over predicted proportions based on (a) $\alpha_\ell = 0.27$, (b) $\alpha_\ell = 0.64$, (c) $\alpha_\ell = 0.16$ with partner-specific $P(RE)$, and (d) $\alpha_\ell = 1-$ partner's α_ℓ .

Further, listeners appear to consider their partner's perspective more over the course of the experiment (first half best α_ℓ = 0.43 vs. second half α_ℓ = 0.20), suggesting that listeners may be adjusting to speakers who are focused on their own perspective by weighting their own perspective to the same extent as the partner (i.e., moving α_ℓ toward $1-\alpha_s$).

[4] $BF_{10} > 10$ are considered strong evidence for H1 over H0.

Perspective-weighting in dyads

The changes in both α_s and α_ℓ over the course of the task are suggestive of adaptation. One possibility is that interlocutors are adapting to the task: speakers may be learning what it is like to receive instructions that are ambiguous given the visual display, and listeners may become more aware of the cognitive burden on the speakers who must take into account what their partner does and doesn't know. However, interlocutors may also be adapting to the idiosyncrasies of their conversational partner; we investigate this possibility by taking account of the dyadic structure of the task.

Production of modified expressions

Participants varied substantially in terms of how much they weighted the two perspectives when designing referring expressions. Here we ask whether this variability was related to their partners' production behaviour. One might imagine that having a partner with a higher α_s, who frequently produces modifiers that are not necessary from the listener's perspective, might lead one to similarly overproduce unnecessary modifiers: This would lead to a positive correlation between α_s values within a dyad. Alternatively, the listener might try to set a good example for their egocentric partner by producing felicitous references. This would lead to a negative correlation between α_s values within a dyad. However, neither scenario was borne out: There was no correlation between the α_s values of members of the same dyad ($r = -0.04, CI = [-0.29, 0.23]$), nor did such a relationship emerge over the course of the task (first half $r = 0.12, CI = [-0.17, 0.37]$ vs. second half $r = -0.03, CI = [-0.30, 0.22]$).

Comprehension of modified expressions

Fig. 7 shows how model predictions for $P(obj \mid RE)$ vary as a function of individual production probabilities which are used as $P(RE \mid obj, d)$ (cf. average in Fig 5): while the same general pattern is predicted for all participants, there are nevertheless quantitative differences. For example, for listener 35 the model predicts target probabilities between 0.75 and 0.50 in the Two Contrasts - Privileged condition, whereas for listener 37, the predicted range is 0.50 to 0.40. Given the variability in how speakers produced modified referring expressions, a rational strategy for listeners may be to adapt their comprehension to the production statistics of the partner they have been paired with.

Dyad-specific probabilities of REs. We tested whether predictions generated from probability distributions computed on a per-dyad basis would provide a better fit to the observed gaze data. As in the overall comprehension modelling, we simulate model predictions across the range of potential α_ℓ values and find the one which minimizes mean squared error. In contrast to the previous analyses, the values of $P(RE \mid obj, d)$ are not the same across listeners but rather computed for each listener based on their partner's production probabilities. Predicted values from the best-fitting $\alpha_\ell = 0.16$ with $P(RE \mid obj, d)$ estimated from the partner's productions

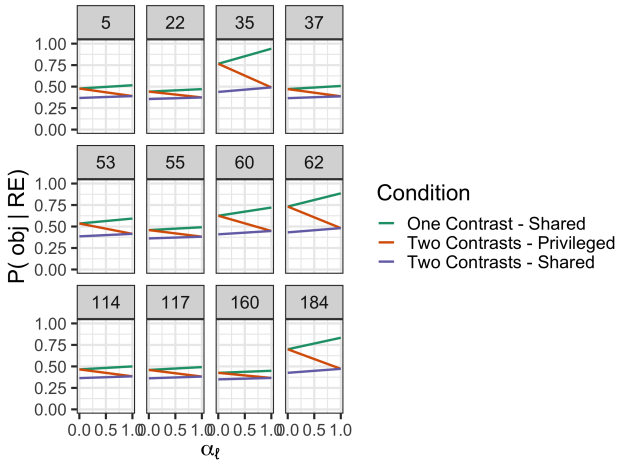

Figure 7: Model predictions of $P(target|RE)$ for a random sample of 12 participants

were robustly correlated with gaze data ($R^2 = 0.46$; Fig. 6c). However, these provided a worse fit than the earlier model, which used $P(RE \mid obj, d)$ estimated from the average of all participants' productions and $\alpha_\ell = 0.27$ ($BF_{10} = 6.3e-28$). The inferior model fit of the dyad-specific model suggests that listeners may not be rapidly adapting their pragmatic inferences to the production probabilities of their partners, though it is possible that some adaptation is masked by the additional variability in the predictions.

Dyad-specific probabilities of α_ℓ. It is possible that instead of changing their estimates of $P(RE \mid obj, d)$, listeners may be updating their α_ℓ to reflect their partner's perspective-weighting, α_s. We compare predictions from models with α_ℓ values corresponding to each listener's own α_s and $1-\alpha_s$ as well as their partner's α_s and $1-\alpha_s$. These four models also did not provide a better fit to gaze data than the model using the average $P(RE \mid obj, d)$ estimates and $\alpha_\ell = 0.27$ (own α_s: $R^2 = 0.46$, $BF_{10} = 6.7e-28$; partner's α_s: $R^2 = 0.45$, $BF_{10} = 1.3e-30$); 1-own α_s: $R^2 = 0.49$, $BF_{10} = 9.8e-20$; 1-partner's α_s: $R^2 = 0.50$, $BF_{10} = 7.8e-17$, Fig. 6d). Taken together, we find no evidence that listeners are tuned to their partner's perspective-weighting. However, as pointed out earlier, the additional variability in the predictions may make such effects more difficult to detect. It is interesting to note that, among these four models, the one that fits the data best is the one where α_ℓ is set to 1 - partner's α_s (own α_s: $BF_{10} = 8.6e-12$; partner's α_s: $BF_{10} = 1.6e-14$; 1-own α_s: $BF_{10} = 1.3e-3$). We speculate that this may reflect the listeners changing their perspective weights to match how much their partner weights the listener perspective during speaking.

Conclusions

In a dyadic referential communication task where interlocutors' perspectives (always) differed, we examined the perspective-taking behaviour of individuals taking turns as speakers and listeners. When they were in the role of speaker, there was much variability in how individuals weighted their own and their partner's perspectives, and yet overall they considered their partner's perspective more in the second half of the session. Interestingly, individuals in the current paradigm assigned relatively less weight to their own perspective compared to speakers in other (similar) experimental paradigms (Heller & Stevenson, 2018; Vanlangendonck, Willems, Menenti, & Hagoort, 2016). This might be because in the current task they also played the role of listener, putting them in a position to notice ambiguous and redundant instructions. This interpretation is further supported by the finding that speakers assigned less weight to their own perspective in the second half of the experiment.

When they were in the role of the listener, participants generally weighted their partner's perspective more than their own's, suggesting that listeners expected speakers to be using their own perspective. As in production, listeners also considered their partner's perspective more over time (i.e., α_ℓ decreased in the second half of the experiment relative to the first), perhaps as listeners come to better understand the cognitive pressures experienced by the speaker attempting to produce an unambiguous and felicitous referential expression.

Modelling comprehension required us to link the listener's probabilistic inferences and the eye-movements they execute; to our knowledge, this is the first attempt to fit model predictions of this kind to eye-tracking data. Because it is not clear what a reasonable linking function should be, as a first foray, we have used an identity link, which makes few assumptions but is likely sub-optimal. Finding a better linking function will be challenging, given that it may depend on the time window being examined, but is a crucial next step that may improve our ability to investigate perspective-taking in comprehension which is often measured with eye tracking.

Comparing the weightings used in production and comprehension suggests that speakers were more egocentric than listeners: that is, $\alpha_s > 0.5$ and $\alpha_\ell < 0.5$, indicating the speaker's perspective is weighted more than the listener's by both speakers and listeners. Note that we do not directly compare α_s and α_ℓ to determine the *relative level* at which they weight the speaker's perspective (i.e., how much above or below the "tipping point" of 0.5 they are). This is because we cannot guarantee that α_s and α_ℓ are directly comparable; they are estimated in different ways and thus may exist on distinct scales (despite both being bounded by 0 and 1). While further investigations are needed to determine whether these conclusions are robust to variation in the estimation procedure (e.g., time window for eye-tracking data), the basic asymmetry we observe between speakers and listeners is consistent with other findings in the literature (e.g., Hawkins & Goodman, 2016; Yoon & Brown-Schmidt, 2013).

Despite the substantial variability between dyads in terms of the statistics of the linguistic environment (see Fig. 7), using partner-specific estimates did not improve model fits. One possibility is that there is simply too much noise in the data

to use individual estimates. A more intriguing possibility is that speakers and listeners may not be tuned to the specifics of their partner because that idiosyncratic experience is outweighed by their prior expectations of speakers using adjectives felicitously in the real world. Note that this conclusion would contrast with prior results where listeners adapt their inferences to the speaker's pragmatic competence (Grodner & Sedivy, 2011; Ryskin et al., 2019), though these studies did not include differences in perspective.

The current paper is the first to analyze the production and comprehension for the same individuals using one model: the multiple-perspectives model (Heller et al., 2016; Mozuraitis et al., 2018). An important goal for future research is to compare the present model to alternative frameworks. For example, within the Rational Speech Act Framework (Frank & Goodman, 2012), Hawkins and Goodman (2016) come to complementary conclusions regarding the division of labor between speakers and listeners in conversation. Comparing both models against the same dataset may yield important insights about the computations that best capture perspective-taking in dyads. Alternatively, dynamical systems approaches provide a more sophisticated way to model perspective-taking effects as they unfold over time (Dale et al., 2018). Integrating probabilistic and dynamical approaches may prove to be a fruitful avenue for future work, particularly given the rich time-course information afforded by the use of eye-tracking.

References

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107*(3), 1122–1134.

Dale, R., Galati, A., Alviar, C., Contreras Kallens, P., Ramirez-Aristizabal, A. G., Tabatabaeian, M., & Vinson, D. W. (2018, September). Interacting Timescales in Perspective-Taking. *Frontiers in Psychology*, *9*.

Frank, M. C., & Goodman, N. D. (2012). Predicting Pragmatic Reasoning in Language Games. *Science*, *336*(6084), 998–998.

Goodman, N. D., & Stuhlmüller, A. (2013, January). Knowledge and Implicature: Modeling Language Understanding as Social Cognition. *Topics in Cognitive Science*, *5*(1), 173–184.

Grodner, D., & Sedivy, J. (2011). The Effect of Speaker-Specific Information on Pragmatic Inferences. In E. A. Gibson & N. J. Pearlmutter (Eds.), *The Processing and Acquisition of Reference* (pp. 239–272). The MIT Press. doi: 10.7551/mitpress/9780262015127.003.0010

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*(1), 43–61.

Hawkins, R. X. D., & Goodman, N. D. (2016). Conversational expectations account for apparent limits on theory of mind use. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society.* (pp. 1889–1894). Philadelphia, PA.

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

Heller, D., Parisien, C., & Stevenson, S. (2016). Perspective-taking behavior as the probabilistic weighing of multiple domains. *Cognition*, *149*, 104–120.

Heller, D., & Stevenson, S. (2018). Modelling reference production using the simultaneity approach: A new look at referential success. In *Proceedings of the 40th Annual Conference of the Cognitive Science Society.* (pp. 481–486). Madison, WI.

Isaacs, E., & Clark, H. H. (1987). References in Conversation Between Experts and Novices. *Journal of Experimental Psychology: General*, *116*, 26–37.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41.

Lane, L. W., Groisman, M., & Ferreira, V. S. (2006). Don't Talk About Pink Elephants!: Speakers' Control Over Leaking Private Information During Language Production. *Psychological Science*, *17*(4), 273–277.

Mozuraitis, M., Stevenson, S., & Heller, D. (2018). Modeling Reference Production as the Probabilistic Combination of Multiple Perspectives. *Cognitive Science*, *42*, 974–1008.

Nadig, A., & Sedivy, J. (2002). Evidence of Perspective-taking Constraints in Children's On-line Reference Resolution. *Psychological Science*, *13*(4), 8.

Ryskin, R., Benjamin, A. S., Tullis, J., & Brown-Schmidt, S. (2015). Perspective-taking in comprehension, production, and memory: An individual differences approach. *Journal of Experimental Psychology: General*, *144*(5), 898–915.

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

Vanlangendonck, F., Willems, R. M., Menenti, L., & Hagoort, P. (2016). An early influence of common ground during speech planning. *Language, Cognition and Neuroscience*, *31*(6), 741–750.

Wilkes-Gibbs, D., & Clark, H. H. (1992, April). Coordinating beliefs in conversation. *Journal of Memory and Language*, *31*(2), 183–194.

Yoon, S. O., & Brown-Schmidt, S. (2013, October). Lexical differentiation in language production and comprehension. *Journal of Memory and Language*, *69*(3), 397–416.