Combining Multiple Perspectives in Language Production: A Probabilistic Model

Mindaugas Mozuraitis
Dept. of Computational Linguistics & Phonetics, Saarland University
mindauga@coli.uni-saarland.de

Suzanne Stevenson
Dept. of Computer Science
University of Toronto
suzanne@cs.toronto.edu

Daphna Heller
Dept. of Linguistics
University of Toronto
daphna.heller@utoronto.ca

Abstract

While speakers tailor referring expressions to the knowledge of their addressees, they do so imperfectly. Our goal here is to provide an explanation for this type of pattern by extending a probabilistic model introduced to explain perspective-taking behavior in comprehension. Using novel production data from a type of knowledge mismatch not previously investigated in production, we show that production patterns can also be explained as arising from the probabilistic combination of the speaker's and the addressee's perspectives. These results show the applicability of the multiple-perspectives approach to language production, and to different types of knowledge mismatch between conversational partners.

Keywords: language production; computational modeling; reference; audience design; common ground; perspective-taking; pragmatics; probabilistic models.

Introduction

The process whereby speakers tailor their utterances to the knowledge state of their addressee is known as "audience design". Much research on audience design has focused on reference – i.e., the labeling of objects. Reference is an ideal test bed for audience design, because of the clear action associated with referring: the speaker needs to produce a linguistic form that enables her addressee to identify the intended object. For example, a speaker should only use the name *Aloysius* if she can assume that her addressee will be able to map this name onto the intended person. More generally, in order to be understood, a speaker should rely on shared information when choosing a referring expression (Clark & Marshall, 1981).

Indeed, psycholinguistic research has shown that speakers rely on shared information when choosing a referring expression (e.g., Nadig & Sedivy, 2002). However, their referring expressions are not always based on shared information alone. For example, when producing language under time pressure, speakers do not distinguish between shared information and their own privileged knowledge (Horton & Keysar, 1996). But even when speakers do distinguish the two, their utterances are sometimes formed using privileged information. For example, when there is one shared triangle and a larger triangle known only to the speaker, speakers sometimes say *the small triangle*, even though for the addressee it would be sufficient to use the unmodified expression *the triangle* (Wardlow Lane & Ferreira, 2008; Yoon et al., 2012). Furthermore, when their addressee does not know a name, speakers sometimes include it nonetheless, along with the descriptive information that would allow the addressee to identify the referent (Isaacs & Clark, 1987; Heller et al., 2012). In sum, speakers do not fully adapt to addressees' knowledge, and production often reflects some egocentric tendencies.

The goal of this paper is to propose an explanation for these "mixed" patterns. Specifically, we extend to cover language production a computational model previously introduced to explain reference comprehension through the probabilistic combination of speakers' and addressees' perspectives (Heller et al., 2016). We first present a new production experiment, and then turn to modeling its results. While much research on reference production has considered knowledge mismatch due to differences in visual perspectives (e.g., Horton & Keysar, 1996; Nadig & Sedivy, 2002; Wardlow Lane & Ferreira, 2008; Yoon et al., 2012), our experiment involves a case where knowledge differs with respect to the *function* of objects. The predictions of our model, which combines the contribution of the speaker's perspective and the addressee's perspective, are a good fit to the human data.

Production Experiment

Our experiment investigates the production of referring expressions in cases where the speaker and the addressee have differing knowledge about the *function* of an object. To create such knowledge mismatch, we use Visually-Misleading Objects [VMOs] whose appearance does not match their function, such as a crayon that is shaped to look like a Lego block (similar objects were used in Mozuraitis et al., 2015, a comprehension study). One novel aspect of our design is that it examines the effects of knowledge mismatch about the VMOs' function *indirectly*: As shown in Figure 1(a), these objects were not described by participants (i.e., they were not the target). Instead, they were paired with a target (a typical object) that either matched their appearance (e.g., an actual Lego block) or that shared their function (e.g., a regular crayon).

What happens when both interlocutors have been demonstrated the function of the VMO (**the shared condition**)? If the VMO is categorized based on its appearance (e.g., as a Lego), then there are two Legos in the ***appearance*** condition (Fig. 1(a-1)), and a referring expression should include a modifier to distinguish the appearance-target from the contrast (e.g., *the Lego on your left*). If the VMO is categorized based on its function (e.g., as a crayon), then there are two crayons in the ***function*** condition (Fig. 1(a-3)), thus a modified referring expression should be used to distinguish the function-target from the contrast (e.g., *the longish crayon*). Importantly, these are not mutually exclusive: it is possible that both appearance and

function play into the categorization of objects, and speakers will use a modified expression in both cases.

(a) Target x Contrast conditions (b) Example critical display

	Target	Contrast
1)	Appearance	VMO Lego-crayon
2)	Appearance	Control
3)	Function	VMO Lego-crayon
4)	Function	Control

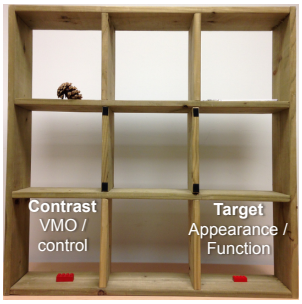

Figure 1: Sample materials

How will this pattern change (if at all) when only the speaker has been shown the function of the VMOs (the privileged condition)? As demonstrated in Mozuraitis et al., (2015), when people are *not* demonstrated the function of these objects, they categorize them based on appearance (e.g., as a Lego). Thus, if speakers (participants) tailor the referring expression to the perspective of the addressee (who can be assumed to think that the VMO is a Lego), they should use a modified expression when referring to the appearance-target (e.g., *the Lego on your left*), but not with the function-target (which they will simply call *the crayon*).

Method

Participants We report data from 80 native English speakers from the University of Toronto community, paid $10 each. An additional 7 participants were excluded because they failed to follow instructions (n = 3) or they reported a suspicion that their partner was not naïve (n = 4).

Materials and design All displays contained four objects (e.g., in Figure 1b the distractor objects are a pinecone and a wand). Three factors were manipulated in a 2 x 2 x 2 design: Knowledge state (shared vs. privileged; between-participants), type of target (appearance vs. function; within-participants), type of contrast (visually-misleading vs. control; within-participants).

Knowledge state. In the shared conditions (n=40), at the beginning of each trial, the function of the VMO (and other objects) was demonstrated non-verbally: e.g., the experimenter drew on a piece of paper with the crayon that looks like a Lego, while both the speaker (participant) and the addressee (confederate) were facing the objects. In the privileged conditions (n=40), the function of objects was demonstrated only to the speaker, while the addressee (confederate) faced away from the display (and wore headphones). In the latter condition, the goal was to give clear and consistent cues to the speaker that the addressee is unaware of the true (unexpected) function of VMOs.

Type of target. In the appearance conditions, the target shared the appearance of the VMO (e.g., a regular Lego with the Lego-crayon; see Fig. 1(a-1)). In the function conditions, the target shared the function of the VMO (e.g., a regular crayon with the Lego-crayon; see Fig. 1(a-3)).

Type of contrast. In the visually-misleading (critical) conditions, the contrast was one of 8 VMOs (appearance-function): Lego-crayon, paintbrush-eraser, baseball-yoyo, cigarette-pencil, book-box, lighter-pencil sharpener, pen-screwdriver, lightbulb-candle. In the control conditions, the VMOs were replaced by objects from an unrelated category that were similar to the VMOs in size and color (Fig. 1(a-2) and 1(a-4)). The control conditions were intended to provide baseline modification rates for the target objects when the display does not contain any contrasting objects.

A list design cycled the pairing of target and contrast objects such that participants saw a given array only once. Across participants, each array occurred in all of the experimental conditions. Across the experiment, target and contrast objects appeared in each of the four display positions equally often. To counteract any contingencies created by the critical items, we added 24 filler displays. To ensure that VMOs are not always relevant for the instruction, two fillers had a VMO paired with an object matching its appearance or function, but neither was the target. To have speakers refer to strange objects, eight fillers had a difficult-to-name object (not VMOs), with six of those as the target. To divert attention from appearance and function contrasts, ten fillers had a pair of objects contrasting in materials, with four as the target. The final four fillers had four unrelated objects. Trials were presented in random order with no two consecutive critical trials.

Procedure Participants (always the speaker) were led to believe the confederate was a naïve participant, and that they were assigned to their respective roles arbitrarily. Following Kuhlen & Brennan (2013), we used confederates blind to the purpose of the experiment. Two confederates participated in the shared condition. A third confederate participated in the privileged condition, to ensure she had not previously seen the function of the VMOs.

Partners were seated on opposite sides of a table with a 3x3 cubbyhole display (Fig. 1(b)); the middle cubbyhole was covered to avoid eye contact that might reveal referential intent. Each trial began by the experimenter demonstrating the function of objects non-verbally, and placing them in the display. Then, the speaker saw a picture on a computer monitor (not visible to the addressee), indicating the object to be moved and its target location. Speakers were given no further instructions, except to refrain from pointing or other gestures. Confederates did not act until speakers completed instructions (they were instructed to this effect during their training).

Results

The dependent variable was whether speakers' referring expressions for the target included a modifier (e.g., *the Lego on your left* or *the smaller Lego*, coded as 1) or not (e.g., *the Lego*, coded as 0). The data were analyzed using a mixed-effects logistic regression model with participants and items as crossed, independent, random effects. We used models

with the maximal random-effect structure that led to convergence. The model that converged included random intercept for participants and random intercept for items.

The 2x2x2 model (summarized in Table 1) revealed a main effect of contrast type: as expected, speakers were overall more likely to use modifiers in referring to the target when it occurred with a VMO than with a control object (.82 vs. .16). There were 3 significant interactions: Target type x Contrast type, Target type x Knowledge state, and the 3-way interaction. Here we focus on unpacking the 3-way interaction, both for theoretical reasons, and because the 2-way interactions are subsumed by it.

Table 1: The 2x2x2 model. Significant effects are bolded.

Effect	*β*	*SE*	*z*	*p*
Knowledge	0.29	0.32	0.90	0.367
Target	-0.32	0.26	-1.24	0.215
Contrast	**4.06**	**0.33**	**12.46**	**<0.001**
Knowledge x Target	**2.20**	**0.53**	**4.18**	**<0.001**
Knowledge x Contrast	-0.93	0.53	-1.74	0.081
Target x Contrast	**1.40**	**0.52**	**2.69**	**0.007**
Knowledge x Target x Contrast	**3.50**	**1.05**	**3.33**	**<0.001**

Separate follow-up analyses were conducted for the two types of contrasts – see Figure 2. When the contrast was a **control** object, there was a main effect of target type (β=-1.07, *SE*=0.37, *z*=2.88, *p*=0.004), indicating that participants were overall more likely to use a modified expression to refer to the function-target rather than the appearance-target (.22 vs. .11); recall these are simply different objects and thus the pattern is not meaningful beyond providing a baseline. The main effect of knowledge state was marginal (*p*=0.074): there was a trend for speakers modifying more in the privileged conditions. However, the Target type X Knowledge state interaction was not significant (*p*=0.581), indicating that, in the absence of a contrasting object, the knowledge state manipulation did not change the modification pattern differentially for the function and appearance targets.

When the contrast was a **VMO**, the Target type X Knowledge state *was* significant (β=4.08, *SE*=0.83, z=4.90, *p*<0.001; main effects were not, *p*s>0.250). Pairwise comparisons revealed that **when the target matched the VMO in function**, participants were *less* likely to use modified expressions to describe it in the privileged condition (.65 vs. .92; β= -1.95, *SE*=0.49, *z*=-3.95, *p*<0.001). This pattern indicates that speakers adapted to the addressee's perspective: they were less likely to use a modifier to differentiate the target (e.g., a typical crayon) from the VMO (e.g., the Lego-crayon) when they had no reason to assume that the addressee knew these shared their function. When **the target matched the VMO in appearance**, the pattern was reversed: participants were *more* likely to use modifiers in the privileged condition (.94 vs. .75; β=2.33, *SE*=0.92, *z*=2.53, *p*=0.012). This pattern further demonstrates that speakers adapted to the addressee's perspective: they were more likely to use a modifier to differentiate the target (e.g., a regular Lego) from the VMO (e.g., the Lego-crayon), when they knew the addressee did not know the real function of the VMO, and thus would expect the VMO to have a function consistent with its appearance.

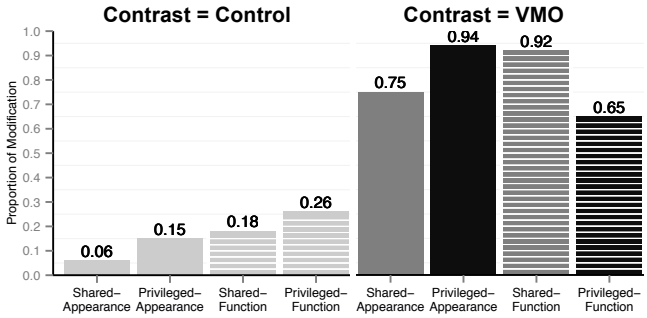

Figure 2: Modification rates produced.

Taken together, the crossover interaction indicates that speakers were sensitive to the knowledge state of their addressee in tailoring referring expressions. But note that if speakers in the privileged condition completely adapted to the addressee's perspective, they should not modify at all in the VMO-privileged-function condition (rightmost column in Figure 2): this is because the function-target (a crayon) would not need to be distinguished from the VMO (the Lego-crayon that the addressee did not know was a crayon). However, speakers used more modifiers in this case than in the corresponding control condition (.65 vs. .26; β=1.84, *SE*=0.41, *z*=4.52, *p*<0.001). This difference reveals that speakers did not completely adapt to the addressee's perspective. Instead, this pattern reflects the consideration of the addressee's perspective along with their own.

The Probabilistic Model

The probabilistic model of Heller et al. (2016) was developed to account for apparently-inconsistent results in the literature on perspective-taking in comprehension. The approach models reference resolution (i.e., the comprehension of a referring expression) as the probability P(*obj*|*RE*) that a certain object *obj* is the referent intended by the speaker, given a referring expression *RE*. Following standard practice, the probability on the left is rewritten using Bayes rule, as a product of two probabilities; importantly, the component probabilities are then conditioned on the *domain of reference* (*d*):

$$P(obj|RE) =^{def} \alpha P(RE|obj,d=e)P(obj|d=e) + (1-\alpha)P(RE|obj,d=c)P(obj|d=c) \quad (1)$$

An important aspect of this model is the conditioning on the domain *d*. This conditioning captures the fact that reference depends not just on what the intended referent is, but also on what other objects need to be distinguished from the referent. The likelihood P(*RE*|*obj*,*d*) captures the preference for using the referring expression *RE* to describe the various objects *obj* in domain *d*; the prior P(*obj*|*d*) captures the probability of *obj* being referred to in the context of domain *d*.

A second important aspect of the model is that reference resolution is guided by *both* the common ground perspective *c* (those entities which both the speaker and the addressee

can see; which was identical to the speaker's perspective), and the egocentric perspective *e* (all the objects the addressee can see). This contrasts with other Bayesian models of reference, which have not considered the combined influence of the speaker's and addressee's perspectives (e.g., Frank & Goodman, 2012; Goodman and Stuhlmüller, 2013; Kehler & Rohde, 2013). These two domains are weighed in formula (1) by α and (1-α), respectively, to ensure the combination of the component probabilities forms a probability distribution.

In order to model **language production**, we extend Heller et al.'s (2016) model in two ways. First, since we are modeling the speaker's choice of referring expression rather than the addressee's search for a referent, the probability of interest is P(*RE*|*obj*) rather than P(*obj*|*RE*) – that is, we directly model the preference for various referring expressions *RE* assuming that the object *obj* to be referred to is a given. In this case, there is no need to apply Bayes rule. Second, we observe that the use of the domain-weighting constant α in the Heller et al. formulation can be avoided: we can rewrite P(*RE*|*obj*) by marginalizing over all possible values of the domain variable *d* (where *D* is the set of possible referential domains for production):

$$P(RE|obj) = \sum_{d \in D} P(RE|obj,d)P(d) \quad (2)$$

Marginalizing over *d* provides a well-motivated way within probability theory to condition the probability of the *RE* on the domain *d* when its value is not a given. By capturing the degree of influence of each domain *d* in the probability P(*d*), the formula in (2) more naturally encodes the kind of domain weighting that Heller et al. achieved with the ad hoc constant α in formula (1). This approach also provides a more general formulation that can be readily extended to the case of more than two domains. It is important to note that formula (1) of Heller et al. can be recast within this same marginalization approach as:

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

where $D=\{e, c\}$, and P(*d*=*e*) replaces α and P(*d*=*c*) replaces (1-α). Thus, together the two probability formulas we propose, (2) and (3), provide a unified way of modeling both production and comprehension of referring expressions under the influence of multiple perspectives.

Instantiating the Probability Model

The Variables in the Probability Formula

The Object *obj* The given object *obj* in our probability formula of Eqn. (2) will denote the *referent*: the object the speaker will label with the goal of the addressee choosing it.

The Referring Expression *RE* Because the dependent variable we used in the experimental results was modification rate, we consider *RE* for the purposes of modeling to represent modified referring expressions. That is, instead of determining the preference for a *specific* referring expression, we use the probability formula to calculate the probability of using modification, P(*RE*=*MOD*|*obj*=*target*), abbreviated as P(*MOD*|*target*).

The Domains *D* We consider the set *D* of multiple domains that influence the speaker's formulation of referring expressions to consist of two domains: the domain *s* is the speaker's (egocentric) perspective, and the domain *a* is the addressee's perspective. Thus, $D=\{s, a\}$. In general, *d*=*s* corresponds to the objects and their properties known by the speaker, and *d*=*a* corresponds to the objects and properties known by the addressee. Importantly, when the addressee did not know the true function of a VMO (privileged condition), we assume the domain *a* reflects the speaker's assumption about the addressee's false belief. (Note that Heller et al. (2016) used different labels for the perspectives of the interlocutors in modeling comprehension: domain *e* for the egocentric domain of the addressee and domain *c* for the common ground, which was identical to the speaker's perspective.)

The Resulting Probability Formula With the above variable values, we can instantiate formula (2) as:

$$P(MOD|target) = \sum_{d \in \{s, a\}} P(MOD|target, d)P(d) \quad (4)$$

Showing the sum over the two possible domains yields the following formula to model the experimental data:

$$P(MOD|target) = P(MOD|target, d=s)P(d=s) + P(MOD|target, d=a)P(d=a) \quad (5)$$

Estimating the Probabilities

P(*MOD*|*target*, *d*) Because our goal is to use the multiple domains approach to explain data patterns obtained under knowledge mismatch (i.e., those in the critical Privileged conditions) we derive probabilities from the Control and Shared conditions, and see whether their application in the Privileged conditions obtains the observed data pattern.

Specifically, we take each probability P(*MOD*|*target*,*d*) in one of the critical experimental conditions to reflect two influences. First, a "baseline" level of modification holds for different objects regardless of the presence of a VMO; this baseline is taken directly from the modification rates in the four Control conditions (we include these because of the variation). Second, when the target needs to be distinguished from a contrasting VMO, as in the Shared conditions, there is an additional level of modification. We thus take the Shared experimental conditions to reflect a "typical" modification rate for when the target shares only appearance or only function with the VMO (Shared-Appearance and Shared-Function, respectively).

Next, we use these values to predict the behavior observed in the two VMO-Privileged conditions. We set the value of the probability P(*MOD*|*target*,*d*) in each of the Privileged conditions to be the sum of:

(i) the baseline level of modification obtained in the corresponding Control condition, plus:

(ii) the amount of modification due to the target sharing either the Appearance or the Function of the contrast object, from the corresponding Shared condition.

These component estimates are shown in Fig. 3a. The resulting values for the Privileged conditions are shown below in Modeling the Production Data, where we consider the speaker's and the addressee's perspectives (other panels of Fig. 3, described below).

P(*d*) As in Heller et al. (2016), we consider that the language user's weighting of each domain is not directly observable. We determine the range of values of P(*d*) that yields a fit to the empirical data, and see if this range fits our hypothesis that speakers use both domains in formulating referring expressions. Because we assume that *d* can only take on the values *s* and *a*, those values exhaust the probability space, and so P(*d*=*s*)+P(*d*=*a*)=1, or P(*d*=*a*)=1–P(*d*=*b*). Given that there is only one parameter to consider here (the other value is the additive inverse), the model needs to account for *two* patterns of modification (i.e., in Privileged-Appearance and in Privileged-Function) with *one* parameter setting for weighing perspectives. Thus, our experimental design provides a critical test of the model.

Modeling the Production Data

To model the data, we must consider what the speaker's and addressee's perspectives are regarding the objects and their relevant properties, with attention to the relation of the VMO to the target. This relation dictates the component probabilities that must be added to the baseline preference for modification. Specifically, does the target share Appearance with the VMO, share Function with the VMO, or share neither? The answer depends on the perspective.

Speaker's Perspective (P(*MOD*|*target*,*d*=*s*): Fig. 3b) Because the speaker always knows the true function of the VMOs, the speaker's perspective is identical in the Shared and Privileged conditions. We calculate P(*MOD*|*target*, *d*=*s*) in the Appearance conditions by adding the shared-appearance probability (.69) to each corresponding baseline, and in the Function conditions by adding the shared-function probability (.74) to each corresponding baseline.

Addressee's Perspective (P(*MOD*|*target*,*d*=*a*): Fig. 3c) In contrast to speakers, the addressee's perspective changes across the knowledge state manipulation. In Privileged-Appearance, the similarity in appearance between the target and the VMO (e.g., a Lego and a Lego-crayon) should lead the addressee to assume that the objects *also share function*. Thus, we add the **shared-function** probability (.74) to the baseline to mimic the addressee's assumed perspective (despite the fact that in actuality the objects share only their appearance). In Privileged-Function (e.g., a crayon target and Lego-crayon contrast), the addressee's false belief should lead to an assumption that the VMO shares neither appearance nor function with the target, and thus a **value of 0** is added to the baseline.

Combining the Two Perspectives (P(*MOD*|*target*): Fig. 3d) Our proposal is that the speaker probabilistically combines both their own perspective with the addressee's perspective in considering the preference P(*MOD*|*target*) for a referring expression. Assuming that the probabilities in Fig. 3b and 3c are weighted equally in Eqn. 5 (P(*d*=*s*)=P(*d*=*a*)=0.5), we obtain the values for P(*MOD*|*target*) across the four experimental conditions illustrated in Fig. 3d. The levels of probability for the Privileged conditions (which are the values we aim to predict) have a very good fit to the behavioral data (Fig. 2; recall that the values for the Shared conditions are set according to the human data.) With 0.25 < P(*d*=*a*) < 0.65, the model yields values in the ranges consistent with the 95% confidence intervals for the Privileged conditions; the best fit is obtained with P(*d*=*a*)=0.47, which is very close to the equal combination in Fig. 3d. Importantly, because we are predicting two values (Privileged-Appearance and Privileged-Function) with one parameter (P(*d*=*a*)), this is not trivial (if it were one data point, there would always be *some* value for P(*d*=*a*) that would achieve the fit).

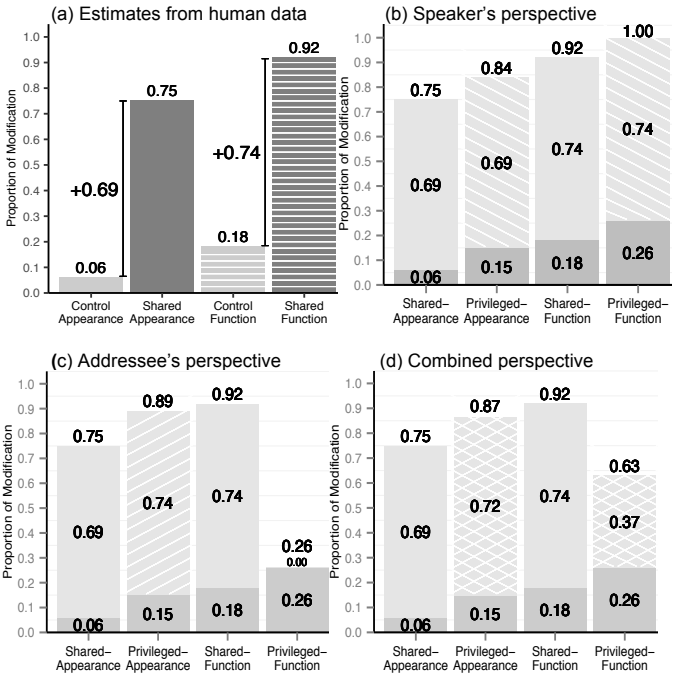

Figure 3: Modeling

Discussion

We present the first model of reference production that probabilistically combines the influence of multiple perspectives. Our model explains the modification patterns observed in our production experiment as a result of speakers considering both their own perspective and the addressee's perspective in choosing the form of a referring expression. The modeling results support the hypothesis that speakers use *both* perspectives, because weighing the two perspectives about equally gives a good match to the human data. A model where speakers are fully egocentric (equivalent to P(*d*=*a*) being close to 0) or fully adapt to the addressee's perspective (equivalent to P(*d*=*a*) being close to 1) cannot predict the pattern observed. Furthermore, this result is attractive because we are using the same approach that successfully modeled perspective-taking in comprehension (Heller et al., 2016).

Why would interlocutors use both their own perspective along with their partner's? One reason is that it may be cognitively taxing to completely suppress one's own perspective. It has already been shown in the literature that perspective-taking abilities are tied to inhibition control (Brown-Schmidt, 2009; Nilsen & Graham, 2009), and to the salience of one's own perspective in the situational context (Wardlow Lane & Ferreira, 2008).

There are also *linguistic* reasons, however, for why interlocutors might use both perspectives, namely performing felicitous moves in conversation. For example, producing a meaningful assertion requires choosing information from one's own perspective and assessing that one's partner does not know the information; formulating a question, in contrast, involves identifying a knowledge gap in one's perspective and assessing that the partner *does* have this information.

Finally, even when tailoring referring expressions, it may not be ideal for speakers to focus on the addressee's perspective alone: what we have been calling "the addressee's perspective" is just a *hypothesis* on the part of the speaker about the addressee's knowledge. While this hypothesis may have been developed based on strong cues (e.g., the fact that the addressee did not see the experimenter demonstrate the function of VMOs), other cues in the situation may suggest to the speaker that the addressee actually shares their knowledge (e.g., they may note small perceptual cues on the VMOs that are suggestive of their unexpected function). Given the uncertainty in assessing the addressee's knowledge state, the speaker may do well to consider their own perspective as relevant to expressing their intent.

Note that this consideration and integration of multiple communicative contexts is different from that explored in Goodman and Stuhlmüller (2013). There, probabilistic weighing is used to model *uncertainty about the other partner's state of knowledge*, with the goal of maximizing adaptation to the partner. This fundamentally differs from our approach of weighing and integrating the differing perspectives *of the two partners*. In future work, insights from the Goodman and Stuhlmüller model of uncertainty with respect to a single perspective can be integrated within our approach that combines multiple perspectives.

Thus, although on the surface it may seem ideal to choose a referring expression that is fully tailored to the addressee's perspective (cf. Clark & Marshall, 1981), our view is that balancing the different demands of conversation requires actively maintaining – and integrating – a representation of both partners' perspectives. In other words, perspective-taking behavior is achieved neither by focusing on shared information nor by trying to fully adapt to one's partner, but rather by probabilistically integrating the perspectives of all interlocutors. Future work will aim to disentangle the considerations that lead to the weighing of perspectives, as well as the effect of feedback on weighing and re-weighing.

Acknowledgments

We are grateful to Rafiya Asad for her help with data collection and coding, and we acknowledge support from NSERC and SSHRC of Canada.

References

Brown-Schmidt, S. (2009). The role of executive function in perspective-taking during on-line language comprehension. *Psycho. Bulletin & Review 16*, 893-900.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. Joshi, B. Webber, & I. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63).

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

Goodman, D. N., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science, 5*, 173–184.

Heller, D., Gorman, K. S. & Tanenhaus, M. K. (2012). "To name or to describe: shared knowledge affects referential form". *Topics in Cognitive Science*, *4*, 290-305.

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

Horton, W. S. & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*, 91-117.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, *116*, 26–37.

Kehler, A. & Rohde, H. (2013). A Probabilistic Reconciliation of Coherence-Driven and Centering-Driven Theories of Pronoun Interpretation, *Theoretical Linguistics*, *39*, 1-37.

Kuhlen, A. K. & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin & Review*, 20, 54-72.

Mozuraitis, M., Chambers, G. C., & Daneman, M. (2015). Privileged vs. shared knowledge about object identity in real-time referential processing. *Cognition*, *142*, 148-165.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science, 13,* 329–336.

Nilsen, E., & Graham, S. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology 58*, 220-249.

Wardlow Lane, L. & Ferreira, V. S. (2008). Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *6*, 1466-1481.

Yoon, S. O., Koh, S., & Brown-Schmidt, S. (2012). Influence of perspective and goals on reference production in conversation. *Psychonomic Bulletin & Review*, *19*, 699–707.