Modelling reference production using the simultaneity approach: A new look at referential success

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

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

Abstract

When a speaker produces a referring expression, their overarching goal is to get the addressee to identify a particular object in the context. This goal leads to the expectation that speakers will use a referring expression tailored to the perspective of the addressee. While research in psycholinguistics has indeed found that speakers tailor their referring expressions to the addressee's perspective, they also find egocentric tendencies; namely, a sensitivity to the speaker's own perspective. Mozuraitis, Stevenson and Heller (2018) make the novel proposal that "mixing" perspectives is a design feature of the production system, modelling data from an experiment where knowledge mismatch concerned object function. Here we further test this model on the more common knowledge mismatch of visual perspective, modelling data from Vanlangendonck, Willems, Menenti and Hagoort (2016). The modelling results shed new light on concept of "referential success" that has been assumed to guide reference production.

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

Introduction

Audience design refers to the phenomenon where speakers design their linguistic utterances to fit their audience, based on their assessment of their addressee. It seems intuitively necessary for speakers to engage in audience design if the goal of communication is for the addressee to recover the message they encode in their utterance. But psycholinguistic research on audience design, which focuses mainly on the forms of referring expressions, has produced mixed results. While much research indeed demonstrates that speakers adapt to their addressee in choosing the form of their utterances (Nadig & Sedivy, 2002; Heller, Gorman & Tanenhaus, 2012; Yoon, Koh, & Brown-Schmidt, 2012; Gorman, Gegg-Harrison, Marsh & Tanenhaus, 2013), other work argues that speakers are egocentric, tailoring linguistic forms to their own perspective (Brown & Dell, 1987; Horton & Keysar, 1996; Wardlow Lane & Ferreira, 2008).

A closer look at the referring expressions produced across the different studies suggests that speakers' behavior might be better characterized as a "mixture" of two perspectives, namely some adaptation to the addressee, along with some egocentric tendencies. Indeed, Mozuraitis, Stevenson and Heller (2018) were the first to propose that "mixing" is a design feature of the system. Specifically, they propose that in the tailoring of referring expressions, speakers simultaneously consider their own (egocentric) perspective and their addressee's perspective (see Heller, Parisien & Stevenson, 2016, for a similar proposal about the comprehension of referring expressions).

The simultaneity approach has an interesting property where it does not encode a global consideration of referential success. Because in this approach referring expressions are evaluated relative to each of the perspectives *separately*, and a referring expression is selected based on "mixing" perspectives, the referring expression selected is not directly evaluated as to whether it would allow the addressee to identify the intended object. This aspect of the simultaneity approach seems non-intuitive given that the goal of referring is to get the addressee to identify a certain object.

Referential success has been seen as a central goal for referring expressions at least since philosopher Keith Donnellan (1966) who coined the term *referential* for those uses of descriptions where the goal is for the addressee to choose an object intended by the speaker. Indeed, considerations of referential success have led Clark and Marshall (1981) to propose that referring expressions are tailored relative to shared knowledge. Even approaches that argue that referring expressions are tailored to the egocentric perspective alone (e.g., Horton & Keysar, 1996) include a second step of "monitoring-and-adjustment" that checks whether the resulting referring expression would allow the addressee to identify the intended object. Thus, the simultaneity approach contrasts with other approaches to the production of referring expression.

The goal of the current paper is to test this aspect of the simultaneity approach by modelling production data from Vanlangendonck, Willems, Menenti and Hagoort (2016). We chose to model this study because it contains an explicit manipulation that tests the role of referential success, namely, a case of audience design where egocentricity could possibly lead to referential failure, and a second case of audience design where egocentricity is unlikely to be harmful to referential success. Modelling these conditions allows us to test this aspect of the simultaneity model directly, as well as test its generality beyond the original set of data for which it was developed.

The Vanlangendonck et al. (2016) study

Vanlangendonck et al. (2016) (henceforth VWMH) examine the production of referring expression in a dialogue situation,

where one participant acts as the speaker and a second participant acts as the addressee. The speaker and the addressee each saw an array of objects on their screens, as if they were sitting on the two sides of a vertical shelving unit – see Figure 1. The most important aspect of this setup is that it allows creating knowledge mismatch. Specifically, some objects were visible to both participants (the objects with the white background), while other objects were visible only to the speaker and hidden from the view of the addressee (the objects with the dark background) – see Figure 2. Thus, as in many other studies on audience design (e.g., Horton & Keysar, 1996; Nadig & Sedivy, 2002; Wardlow Lane et al., 2006; Wardlow Lane & Ferreria, 2008; Yoon et al., 2012) knowledge mismatch was established by visual co-presence.

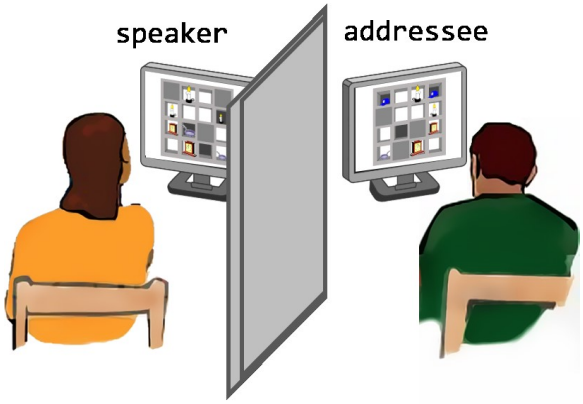

Figure 1: the experimental setup in VWMH

The critical conditions of VWMH required audience design: these are situations where the most appropriate referring expression is different when it is tailored relative to the speaker's perspective versus when it is tailored relative to the addressee's perspective. VWMH tested two such case: in the ADVISABLE condition, even if the referring expression is tailored relative the speaker's perspective, the addressee is likely to identify the correct object despite the fact that this is not the ideal referring expression from their own perspective. In the OBLIGATORY condition, in contrast, if the referring expression is tailored to the speaker's perspective, the addressee may not be able to identify the intended referent, leading to referential failure. Let us consider these in order.

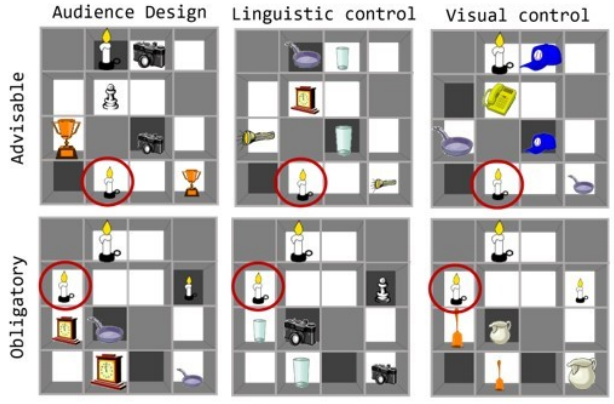

Figure 2: VWMH conditions

In the ADVISABLE condition (top-left display in Figure 2), the target objet (marked in red) is a candle, but, crucially, the speaker also sees a second, bigger candle that is not visible to the addressee. If the speaker tailors the referring expression to the perspective of the addressee, they should use an unmodified expression (e.g., *the candle*). If, however, they tailor the referring expression based on their own perspective, they will produce a modified referring expression (e.g., *the small candle*). VWMH label this condition "advisable" because adaptation can be seen as advisable rather than necessary, as the use of a modified expression would nevertheless allow the addressee to choose the intended object and thus lead to referential success. This critical condition of audience design was accompanied by two control conditions: linguistic control with one candle (top-middle display in Figure 2), and visual control with two candles (top-right display in Figure 2).

Because the research question concerns adaptation to the addressee's perspective, the results in this condition focus on the proportion of trials where speakers produced a bare noun (e.g., *the candle*), the expression expected from the addressee's perspective; the results are summarized in Figure 3. First, speakers behaved as expected in the control conditions. In the linguistic control condition, which is parallel to the addressee's perspective in the audience design condition, speakers mostly produced bare nouns (87.6%), whereas in the visual control condition, which is parallel to the speaker's perspective in the audience design condition, they produced a bare noun very rarely (1%). In the critical case of audience design, speakers mostly produced bare nouns (79.8%), exhibiting adaptation to the addressee. Crucially, however, the adaptation is not complete, because this proportion is significantly lower than the one in the linguistic control condition. In this case, the lack of complete adaptation might be due to the fact that not adapting would not have a harmful effect on referential success.

Turning to the OBLIGATORY condition (bottom-left display in Figure 2), the target objet is a candle and there is a second, bigger candle visible to both conversational partners, but, crucially, the speaker can also see a third, *smaller* candle that is not visible to the addressee. Here. if the speaker tailors the referring expression to the addressee's perspective, they would say *the small candle*, whereas if they tailor the referring expression to their own perspective, they will say *the medium candle* (importantly, VWMH used objects in *four* different sizes, meaning that the expected size adjective could not be determined by the absolute size of the object). The OBLIGATORY condition is different from the ADVISABLE condition in that the two perspectives lead to *incompatible* referring expressions. Thus, if the speaker fails to adapt to the addressee in this case, the addressee might not be able to identify the correct referent, leading to referential failure. This condition was also accompanied by the two control conditions: Linguistic control (bottom-middle display in Figure 2) which is parallel to the addressee's perspective in the audience design condition (i.e., two candles), and visual control (bottom-right display in Figure 2), parallel to the

speaker's perspective in the audience design condition (i.e., three candles).

Here adaptation would lead to using the adjective *small* (or *large*), and hence the results are presented in term of the proportion of trials on which speakers produced these adjectives (e.g., *the small candle*) – see Figure 3. The control conditions showed the expected pattern: in the linguistic control (parallel to addressee's perspective), speakers mostly produced *the small candle* (97.3%), and in the visual control condition (parallel to the speaker's perspective) they rarely produced such modifier (1.4%). In the critical audience design conditions, speakers again showed adaptation to the addressee, mostly producing *the small candle* (89.9%). Here again, adaptation was not complete, as this value is significantly lower than in the linguistic control condition. But in this case the lack of complete adaption is surprising, because the lack of adaptation could potentially threaten referential success.

One of modelling this pattern using the simultaneity approach is to test whether the patterns observed here arise from the same "mixing" behavior.

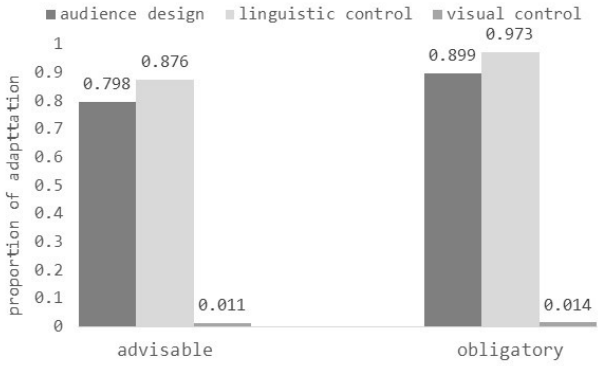

Figure 3: VWMH experimental results. The dependent variable plotted is the proportion of that behavior which is adaptive in the audience design condition: a bare noun in ADVISABLE and *the small N* in OBLIGATORY.

Modelling the production data

The Mozuraitis et al. (2018) simultaneity proposal is operationalized in a computational model as:

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

This formula encodes the observation that a referring expression depends not just on the referent object alone (i.e., *obj*) but also on the domain of reference (*d*), which is the set of contextually-relevant objects from which the referent needs to be distinguished in the current context. (Note that the right hand side of Eqn. 1 is not an application of Bayes rule.) This captures the fact that the same object may be called *the small vase* if it appears with a bigger vase, but will instead be called *the big vase* if it appears with a smaller vase.

The referring expression to be produced, *P(RE|obj),* requires *summing across the possible domains of reference in the context*: for each domain, it takes into account the probability of referring expressions for the object in that domain, *P(RE|obj,d)*, and also the probability of the domain itself, *P(d)*. In a situation of knowledge mismatch between the conversational partners, the different perspectives of the partners constitute two relevant domains of reference: *d=s* is the perspective of the speaker and *d=a* is the perspective of the addressee. In this situation, where *D*={*s*, *a*}, we can rewrite (1) as:

$$P(RE|obj) = P(RE|obj,d=s)P(d=s) + P(RE|obj,d=a)P(d=a) \qquad (2)$$

Thus, in the simultaneity approach a speaker doesn't choose *between* their own (egocentric) perspective and their partner's perspective, but instead uses both simultaneously, combining the contributions of the two perspectives. While it has been previously proposed that perspective information is probabilistic (e.g. Hanna, Tanenhaus & Trueswell, 2003), Mozuraitis et al., (2018) are the first to propose "mixing".

We use this approach to predict the behavior in VWMH where there is knowledge mismatch, namely in the audience design conditions (top- and bottom-left displays in Figure 2).

P(*RE*|*target*, *d*). The first step is to estimate the probabilities of referring expressions in the addressee's perspective, *P(RE|target,d=a)*, and in the speaker's perspective, *P(RE|target,d=s)*.

Recall that VWMH's Linguistic control conditions (top- and bottom-middle displays in Figure 2) are equivalent to the addressee's perspective in the Audience Design displays. Therefore, we use the production patterns in this condition to estimate P(*RE*|*target*,*d=a*); see also Figure 3.

Recall, further, that VWMH's Visual control conditions (top- and bottom-right displays in Figure 2) are equivalent to the speaker's perspective in the Audience Design displays. Therefore, we use the production patterns in this condition to estimate P(*RE*|*target*,*d=s*); see also the relevant columns in Figure 3.

We follow VWMH's analysis, and use as the dependent variable that form which would be the adaptive behavior in the audience design condition: for the advisable condition, it is *the N*, and for the obligatory condition, it is *the small N* – see again Figure 3.

P(*d*). Since the weighting of the two perspectives is not directly observable (cf. Mozuraitis et al., 2018), our approach is to determine the value, or range of values, for the weight that yields a fit to the behavioral data. The resulting P(*d*) indicates the degree to which speakers engage in audience design. Because we assume that *d* can only take on the values *speaker* and *addressee* (see Mozuraitis et al., 2018 for discussion), the two values exhaust the probability space, and so *P(d=s)*+*P(d=a)*=1, or *P(d=a)*=1–*P(d=s)*. In other words, there is only one parameter to consider here, P(*d=a*), as the other value can be derived from it; we therefore refer to the parameter as *P(a)*.

We evaluate our modelling results by looking at what the *P(a)* we obtain tells us about three issue: (1) how different types of knowledge mismatch affect the weight *P(a)*, (2) the consistency of the weight for individuals; and (3) the consistency of this weight across referring situations.

Question 1: Comparing across situations with different cues to the mismatched information

The first question we address in modelling VWMH is whether situations with different cues to shared versus mismatched information lead to different weighing of the two perspectives, as has been proposed in Heller et al. (2016) and Mozuraitis et al. (2018). Specifically, in production, the idea is that the more salient the addressee's perspective is, the more influence it will have (i.e., *P(a)* will be higher), and the less salient it is, the less influence it will have (i.e., *P(a)* will be higher).

Mozuraitis et al. (2018) created situations in which the knowledge mismatch between interlocutors concerned objects' function. To this end, they used visually-misleading objects: objects whose function is not consistent with their appearance, such as a crayon that is shaped like a Lego brick. The mismatched situation they modeled was such that the speaker knew the unexpected function of the object (this function was demonstrated to them by the experimenter), but the addressee did not (they turned their back to the experimenter during the demonstration). The modelling results showed, first, that the pattern of referring expression used is not consistent with the speaker using only their own perspective, or only the addressee's perspective, even when taking into account reasonable amount of noise in the data. Instead, this data was successfully accounted for by "mixing" the two perspectives. The best fit to the human data was achieved when the two perspectives were weighed about equally: $P(a) = 0.48$ and $P(s) = 0.52$. (Again, these sum to 1, so in what follows we only report $P(a)$). When considering the 95% confidence intervals of the means for the modelled condition, the range is $0.26 \leq P(a) \leq 0.64$.

Our goal here is to model the VWMH data, and compare the *P(a)* we obtain from that data to Mozuraitis et al.'s modelling results. What is the prediction with respect to how these should compare? We predict that in VWMH in VWMH, where the cues to mismatch information are visual, the addressee's perspective will be weighed *more* than in Mozuraitis et al. (2018), where the knowledge mismatch concerned object function. This is because, first, the visual mismatch in VWMH has a constant perceptual correlate: objects that are not visible to the addressee have a background with a different color, whereas in Mozuraitis et al. (2018) speakers need to rely on their memory of the experimenter demonstrating the function of the object. Second, the visual setup in VWMH makes it highly unlikely that the addressee would nonetheless know what the hidden objects are. In Mozuiraitis et al. (2018), in contrast, speakers may notice that the visually-misleading object has some properties that are not consistent with their appearance (e.g., noticing that the Lego-crayon is not made of plastic), and may therefore entertain the possibility that the addressee could also notice these properties and figure out the that what looks like a Lego is really a crayon. In other words, this is a situation where there is more *uncertainty* about the addressee's perspective. Finally, the VWMH setup requires speakers to attribute absence of knowledge to their addressees, a level I Theory of Mind mismatch, whereas the Mozuraitis et al. (2018) setup requires speakers to attribute to the addressee *different* knowledge, a level II Theory of Mind mismatch. As the latter is more complex, it stands to reason that it will lead to less weight on the addressee's perspective.

Modelling, Because the experimental manipulation of perspectives in Mozuraitis et al. (2018) was between-participants, the production patterns in the two perspectives came from different participants than the pattern predicted by the model. In other words, these results were population-level modelling. Thus, for the VWMH data, we used the overall means from the Visual control conditions as the speaker's perspective, the overall means from the Linguistic controls conditions as the addressee's perspective, and combine them to achieve the means in the Audience Design conditions. We then find, based on both sets of control condition, a single value of *P(a)* that best predicts the Audience Design behavior in both conditions.

Results. The best fit is obtained with $P(a) = 0.916$, and the model yields values in the ranges consistent with the 95% confidence intervals for the two Audience Design conditions at $0.908 < P(a) < 0.924$. This result matches our prediction that *P(a)* in VWMH will be higher than the *P(a)* obtained in Mozuraitis et al. (2018), where the upper end of their range was 0.64.

That is, these modelling results demonstrate that the addressee's perspective is weighed far more in the situation in VWMH in which the cues to the addressee's perspective are more salient. This is the first piece of evidence that the weighing of perspectives depends on situational cues.

Question 2: Comparing across individuals

The claim of the simultaneity approach is that a speaker weighs the probability of each potential referring expression in the context of both their own and their addressee's perspectives in order to determine the form of the referring expression to be produced. Above we modelled the data at the population level. Our second goal is to examine whether each speaker can be modeled individually as using a single *P(a)* across all the trials. Because of the within-participant design of VWMH, where each participant contributed to both the Linguistic and Visual control conditions and to the Audience Design conditions, we can model these data at the individual level (this was not possible in Mozuraitis et al., because the type of knowledge mismatch they employed drove them to employ a between-participants design).

Modelling. In modelling the subject-level human data from VWMH, we model data from eighteen participants. Two additional participants were excluded, because they made corrections or edited their initial referring expression (e.g., by adding or correcting the adjective) on more than 40% of the trials (The remaining 18 participants made such corrections on 15% of the trials or less).

We split the trials for each speaker, across all six conditions, into two equal-sized groups. To avoid order effects, we took every other trial for each of the six conditions; for ease of reference we'll call these halfl and

half2. We then fit *P(a)* to the Audience Design conditions of half1 based on the two perspectives, which were derived from the Linguistic and Visual control conditions for half1. Next, we combine the two perspectives in half2 (derived from the Linguistic and Visual control conditions of half2) using the weight *P(a)* we got from half1, to predict the pattern of production in the Audience Design conditions in half2. To ensure that there is no bias in one half, we also did the reverse: derive *P(a)* from half2 and use it to predict half1.

Results. To examine our model, we examine how well the predicted response rates correlate with the observed response rates for each subject in VWMH, for each half of the data. When predicting half2 based on half1, we find a very high correlations of r=0.931 (95% CI: 0.819, 0.974) – see Figure 4. We also find a very high correlation when predicting half1 based on half2: r=0.919 (95% CI: 0.791, 0.969) (for space considerations, this is not plotted here). The fact that the *P(a)* that was fit from each half of the data can be combined with the behavior in the control conditions of the other half to predict the behavior in that Audience Design condition of that other half demonstrates that each participant is using a consistent weighing of *P(a)* throughout the experiment.

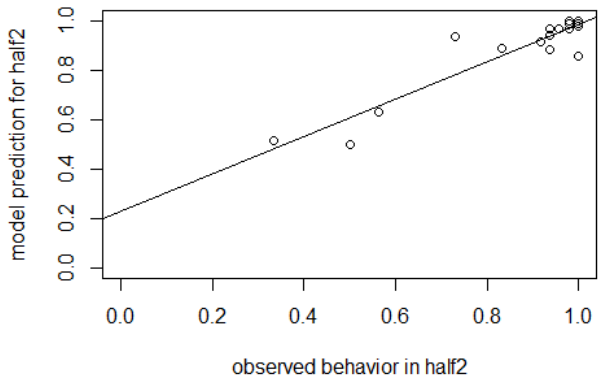

Figure 4: the correlation between observed behavior in half2 and those predicted by the model.

More generally, this successful individual-level modelling provides strong support for the claim of the simultaneity approach that speakers combine their own perspective with the addressee's perspective in tailoring referring expressions.

Question 3: Comparing across referential situations

In the modelling above, we show that each participant can be seen as using a consistent setting of *P(a)* across all their trials in the experiment. Note that this result was achieved when using both Advisable and Obligatory conditions to fit *P(a)* and then predicting both the Advisable and Obligatory conditions. however, those results do not indicate whether speakers use the same setting of *P(a) across the two referring conditions*, or whether they use one setting of *P(a)* in the Advisable condition, and a different setting of *P(a)* in the Obligatory condition.

Recall that the two referring conditions differ in how lack of adaptation to the addressee would affect referential success. In the Advisable condition, lack of adaptation (i.e., being egocentric and saying *the small vase*) should nevertheless allow the addressee to pick the correct object (as the addressee can see only one vase). But in the obligatory condition, lack of adaptation (i.e., being egocentric and saying *the medium vase*) could possibly confuse the addressee who can only see two vases, and may therefore lead to referential failure (recall that VWMH used four different absolute sizes, and thus the adjective *medium* did not correspond to a specific size of objects in their experiment). A different sensitivity to the consideration of referential success would be reflected in the simultaneity model as a different setting of *P(a)*, with a higher *P(a)* when there is a risk of referential failure. This would be similar to having a global consideration of referential success, as has been widely assumed in the literature. The simultaneity approach does *not* encode a global consideration of referential success. Instead, the relative weighing is hypothesized to be affected by general aspects of the situational context and possibly the individuals.

Modelling. To test the model with respect to the potential influence of referential success, we again fit the model's value for *P(a)* on half the data, and use that setting to combine the data from the two control conditions in the other half and predict the values in the Audience Design condition. But here we split the data into the Advisable and Obligatory conditions. This enables us to test directly whether subjects are using a consistent setting of *P(a)* across the entire experiment, or whether they instead adapt their weighing of the addressee's perspective depending on factors specific to each of the referring conditions, namely based on their consideration of which situation will lead to referential success and which may lead to referential failure. The simultaneity approach, in contrast, posits that the same *P(a)* should fit the data either direction, as the weighing is chosen for a particular situation.

Results. We fit *P(a)* based on the control conditions (Visual control → speaker's perspective; linguistic control → addressee's perspective) in the Obligatory trials for each subject, and use that *P(a)*, along with the control conditions for the Advisable condition, to predict the Audience Design response rate for that subject in the Advisable trials. In other words, we use the weighing of perspectives derived from the Obligatory data to predict the Advisable data. In this case, we find a very high correlation between the predicted values and the human data: r=.985 (95% confidence interval: .959 to .994); this is plotted in Figure 5, top panel.

We also did the reverse, namely, fitting *P(a)* based on Advisable conditions, and then using the fit *P(a)* value, as well as the Obligatory control conditions, to predict the rates in the Audience Design condition for that subject in the Obligatory trials. Here again we find a very high correlation to the human data: r=.932 (95% CI: .823, .975); see Figure 5, bottom panel. Using *P(a)* fit from each condition of the data achieves an excellent fit to the other condition, supporting the view that each participant is using a consistent weighing of *P(a)* for both the Obligatory and Advisable trials.

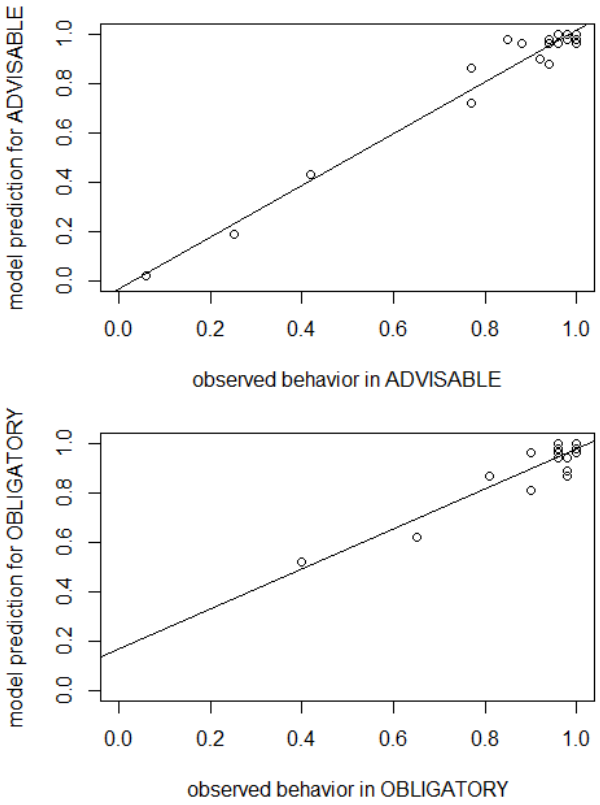

Figure 5: the correlation between observed values and those predicted by the model. Top: values in ADVISABLE predicted based on *P(a)* fit from the OBLIGATORY condition. Bottom: values in OBLIGATORY predicted based on *P(a)* fit from the ADVISABLE condition.

Discussion

We used the probabilistic model of Mozuraitis et al., (2018) to model human data from the production experiment of VWMH. First, the within-participants design of VWMH allowed for individual-level modelling, demonstrating the generality of the simultaneity approach and providing stronger support for it.

More interestingly, this allowed modelling under a different knowledge mismatch than Mozuraitis et al., (2018). Modelling results reveal that speakers in VWMH weighed the perspective of the addressee more than speakers in the Mozuraitis study. We attribute this difference to the cues to the mismatched knowledge available in each of the two situations: with visual co-presence, this information is perceptually available, is associated with less uncertainty, and is a Level I Theory of Mind knowledge mismatch that is easier to attribute to one's partner.

Finally, modelling the two referential situations reveals that speakers are not sensitive to a global consideration of referential success in determining the weighing of perspectives. This is a surprising result as the literature on reference has generally assumed that such a global consideration plays a crucial role in the production of referring expressions. This finding is, however, predicted by the simultaneity approach which assumes the weighing to be determined by cues in the situational context, and possibly cues related to the individual speakers.

Acknowledgments

We are extremely grateful to Flora Vanlangendonck and her colleague for sharing their data with us. We acknowledge support from SSHRC of Canada. to D. Heller and from NSERC of Canada to S. Stevenson.

References

Brown, P., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, *19*, 441-472.

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).

Donnellan, K. (1966). Reference and definite descriptions. *Philosophical Review*, *75*, 281-304

Gorman, K. S., Gegg-Harrison, W., Marsh, C. R., & Tanenhaus, M. K. (2013). What's learned together stays together: Speakers' choice of referring expression reflects shared experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39(3)*, 843-853.

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*, 43-61.

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.

Mozuraitis, M., Stevenson, S. & Heller, D. (2018). Modelling reference production as the probabilistic combination of multiple perspectives. *Cognitive Science*.

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

Wardlow Lane, L., Groisman, M., & Ferreira, V. (2006). Don't talk about pink elephants! *Psychological Science*, *17(4)*, 273–277.

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.

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.

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.