Come Together: Integrating Perspective Taking and Perspectival Expressions

Julia Watson
Department of Computer Science
University of Toronto
(jwatson@cs.toronto.edu)

Anna Kapron-King
Department of Computer Science
University of Toronto
(annakk@cs.toronto.edu)

Jai Aggarwal
Department of Computer Science
University of Toronto
(jai@cs.toronto.edu)

Barend Beekhuizen
Department of Language Studies
University of Toronto, Mississauga
(barend.beekhuizen@utoronto.ca)

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

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

Abstract

Conversational interaction involves integrating the perspectives of multiple interlocutors with varying knowledge and beliefs. An issue that has received little attention in cognitive modeling of pragmatics is how speakers deal with the choice of words like *come* that are inherently perspectival. How do such lexical perspectival items fit into a speaker's overall integration of conversational perspective? We present new experimental results on production of perspectival words, in which speakers have varying degrees of certainty about their addressee's perspective. We show that the Multiple Perspectives Model closely fits the empirical data, lending support to the hypothesis that use of perspectival words can be naturally accommodated as a type of conversational perspective taking.

Keywords: language production; perspective taking; computational pragmatics; perspectival elements; modeling

Introduction

Because people have differing experiences, conversation occurs between interlocutors with differing knowledge and beliefs. Successful communication thus requires interlocutors to continually consider their own and each other's perspectives. In the domain of language production, speakers accommodate to their addressee's perspective, while also showing the influence of their own. For example, in describing an object whose name is unknown to the listener, speakers may use both the name known to the speaker *and* a description of the object (e.g. Isaacs & Clark, 1987; Heller et al., 2012). In other cases, object descriptions differentially reflect the knowledge of the listener depending on the speaker's goals (Yoon et al., 2012). Thus, managing 'conversational perspectives' – the knowledge each interlocutor brings to the interaction – is an inherent aspect of communication.

Research on conversational perspective has generally focused on the linguistic phenomena of reference: the production and comprehension of labels for objects. However, perspectives are also central in other linguistic domains. The best known lexical items that depend on a perspective for interpretation are deictic words such as *there*, *tomorrow*, and *you*, whose meaning depends on an anchoring space, time, and set of conversational participants, respectively (Kaplan, 1979). But other elements have also been argued to be perspectival, because their interpretation also depends on the perspective of an anchor, who may be one of the conversational partners or another attitude-holder. For instance, predicates of personal taste, such as the adjective *tasty*, depend on the perceptual experience of the anchor (Lasersohn, 2005; Kaiser & Lee, 2017). In another example, certain motion verbs, such as *come*, depend on the anchor's location relative to the direction of motion (Hockett, 1990; Oshima, 2006; Anderson & Dillon, 2019): A speaker may say *Rose is coming to the fire station*, if the speaker is located at the fire station, because from their perspective this is a 'coming' event in which the speaker is the anchor. Because in these cases the perspective holder – or the anchor – is part of the lexical semantics of the word, we refer to such elements as involving 'lexical perspective'.

The current project aims to ask whether lexical perspective can be treated as a kind of conversational perspective. We approach this problem from the theoretical stance of the Multiple Perspectives Model (MPM; Heller et al., 2016; Mozuraitis et al., 2018), a probabilistic model of pragmatics previously used in modeling reference. Taking the view that perspective taking is the driving force of communication, MPM proposes that speakers and listeners are always simultaneously integrating both of their perspectives. Our prediction is that the same approach to weighing conversational perspectives can account for behavior observed with elements that have been argued to involve lexical perspective, integrating these two views on perspective.

Inspired by Anderson (2020), we conducted a production experiment to elicit descriptions of motion events, examining how the production of *come* is affected by the location of the speaker and the listener. To further investigate the interaction of lexical and conversational perspectives, we also manipulated the speaker's certainty about the location of the listener. We find that, even in the presence of uncertainty, the speaker balances their perspective with their listener's, suggesting that speakers cannot simply "turn off" perspective taking.

MPM is shown here to naturally capture the phenomenon of lexical perspective using its core mechanism of integrating two conversational perspectives, accurately modeling the rates of *come* in our experiment. In some ways this is similar to Anderson & Dillon (2019) (see also Anderson, 2020), who extended the Rational Speech Acts (RSA) model from Goodman & Frank (2016) to accommodate both the speaker's and the addressee's perspectives. However, while the model of Anderson & Dillon (2019) weighs perspectives, as in MPM, the authors did not connect their model of perspectival expressions to general perspective taking behavior.

Providing an integrated account of (lexical) perspectival expressions and (conversational) perspective taking con-

tributes to both research areas, which have been largely separate up to this point. First, research on perspective taking has mainly focused on reference, and addressing how perspectives are managed in use of perspectival expressions provides a new domain for testing theories of perspective taking. Second, work on perspectival expressions has typically focused on a small subset of lexical expressions that are considered perspectival, and has not considered how their use fits in with other linguistic phenomena related to perspective. In addition to providing an integrated account of conversational and lexical perspective, our results using MPM suggest that perspective taking in communication is similar for both referring expressions and (at least some) perspectival elements, bringing us a step closer to understanding how perspective taking plays out across linguistic domains.

Production Experiment

We conducted an online production experiment to examine how the verb used to describe a motion event depends on the (depicted) location of the conversational partners. Participants were asked to communicate to a listener the destination of a third party character named Rose. We focus on use of the verb *come*, a perspectival element whose meaning depends on having an appropriate anchor at the destination of the motion event. Like Anderson (2020), we first manipulate whether the speaker and the listener are – or are not – at the destination.

We further examine the interaction of lexical and conversational perspectives by introducing cases where the speaker is uncertain of the listener's location. This manipulation contrasts with other work on reference production in which the speaker has uncertainty about things the listener can see that the speaker cannot (Hawkins et al., 2021). In the reference task, the speaker risks producing an ambiguous label if they do not consider the uncertain information (what items are in the listener's visual field) when formulating a description. In our task, the speaker can always communicate Rose's motion towards the destination without considering the uncertain information (the listener's location); for example, the speaker could say *Rose is walking to the* DESTINATION. This experimental manipulation thus complements and extends other work on uncertainty in perspective taking by posing a situation in which the speaker could safely ignore their uncertain knowledge without risk to successful communication, and seeing whether they do so.

If, as we theorize, lexical perspective can be subsumed by conversational perspective, then use of lexical perspectival elements like *come* should also exhibit perspective weighing behavior. In keeping with this, we predict that, when the speaker is certain about the listener's location, they will use *come* the most when both interlocutors are at the destination, and least when neither are.[1] We further predict that speakers will continue to engage in such perspective weighing even when they have uncertainty regarding the listener.

[1] Anderson (2020) made similar predictions, but their results differed from ours, as we return to below.

Method

Participants We report data from 119 participants on Prolific (paid £2.92), who self-reported as monolingual English speakers residing in the US. Sixteen participants were excluded because they described the situations incorrectly, and nineteen because they used either no verb or the same verb on all trials.

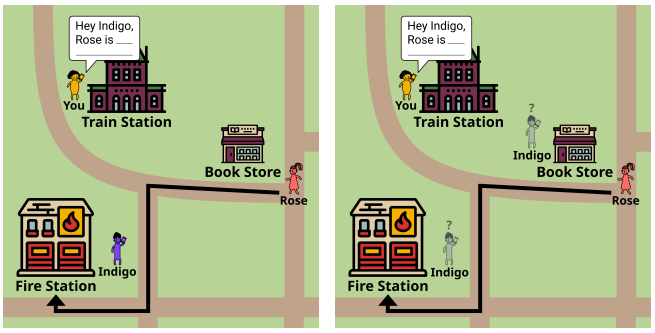


Figure 1: Sample maps in the Listener-Certain condition (left) and Listener-Uncertain condition (right).

Materials and design Thirty-two (32) maps displayed three (varying) locations and three (constant) figures: "you" (the participant) and Indigo (the listener), depicted as communicating over phones, and Rose, with a line to her intended destination (Figure 1). Each map displayed two large and one small location. The large locations (e.g., stadium, hospital) were chosen such that two people could plausibly be at the same place without being able to see each other.[2]

The design was a 2x2x2 within-subjects design. The first two factors manipulated the speaker's location (SPK: destination vs. not) and listener's location (LIS: destination vs. not), yielding four conditions, named for who is at the destination: Both, Speaker, Listener, and None. The third factor manipulated the speaker's certainty about the listener's location (UNC: certain vs. uncertain). In the Certain conditions, Indigo was depicted in purple in a single location; in the Uncertain conditions, Indigo was transparent, and appeared in two possible locations, with a question mark (Figure 1).

A list design cycled the 32 maps through the 8 conditions, so that each participant saw each map once. Across participants, each map occurred in all eight conditions. Item order was randomized for each participant, while ensuring the same condition appeared no more than twice in a row.

Procedure Participants were given the following instruction at the beginning of the experiment: "Rose has told you (the yellow figure) her destination, and you need to relay the information to Indigo". Participants were also instructed on how to interpret the Uncertainty maps. On each trial, the par-

[2] This allows for the situation where the speaker may be uncertain about whether the listener is at the same location as the speaker. In these cases, we depicted the speaker and listener on opposite sides of a large building.

ticipant was shown one map, and prompted with "Hey Indigo, Rose is" followed by a text completion box that ended with a period (to encourage single sentence responses).

Annotation We used NLP tools to automatically annotate responses. First, we identified the first destination mentioned: either with a direct place name (e.g., *fire station*), or with a pronoun referring to one of the characters (e.g., *Rose is meeting me* refers to the location of the speaker). Then, we extracted the verb (e.g., *coming*) or other predicate (e.g., *on the way*) that immediately preceded this destination. If the destination was not expressed, we took the last verb in the response. If the response contained no predicate (e.g., *Rose is **at the fire station***), the response was labelled `no-verb`.[3]

Results

Figure 2 plots, for several motion verbs, the mean proportion of use across the eight experimental conditions. While this provides a rich data set for studying motion verbs, here we focus on *come* as a well-studied verb that is presumed to involve lexical perspective. We note, first, that the rates of *come* are higher when (at least) one interlocutor is present at the destination, and those rates drop in some of the Uncertainty conditions (specifically, Both and Listener).[4]

We first examine these patterns statistically by fitting a 2x2x2 mixed-effects logistic regression model; fixed effects were sum coded. The dependent variable was whether or not the verb *come* was used. The main effect of SPK (speaker at destination) was significant ($\beta = 0.97, SE = 0.30, z = 3.28, p = .001$), indicating that *come* was used more when the speaker was at the destination. The main effect of LIS (listener at destination) was also significant ($\beta = 0.66, SE = 0.13, z = 5.22, p < .001$), indicating that the listener being at the destination also led to an increase in using *come*. The main effect of Certainty was not significant ($p = .88$), but the Uncertainty $\times$ LIS interaction was significant ($\beta = -0.37, SE = 0.13, z = -2.93, p = .003$), indicating that using *come* was affected by uncertainty about the location of the listener. The SPK $\times$ LIS interaction was also significant, ($\beta = -0.35, SE = 0.13, z = -2.75, p = .006$), an effect carried by the difference between the Certainty and Uncertainty conditions (the three-way interaction was not significant).

For our modeling, we needed to compare the rates of *come* across the four conditions: Both, Speaker, Listener, and None. To this end, we fitted a separate mixed-effects logistic regression model for each of the Certainty and Uncertainty conditions. Fixed effects were coded using repeated contrasts (Schad et al., 2020); the models are summarized in Table 1. In the Certainty conditions, all contrasts were significant, matching our hypotheses: the use of *come* was more likely: (1) when Both interlocutors were at the destination compared with just the Speaker;[5] (2) when the Speaker was at the destination compared with just the Listener; and (3) when the Listener was at the destination compared with None of the characters.

The pattern in the Uncertainty conditions was different: (1) Both did not differ from Speaker, and (3) Listener did not differ from None, but (2) there was a difference between the Speaker and the Listener conditions. At first glance, this suggests that when the location of the listener is uncertain,

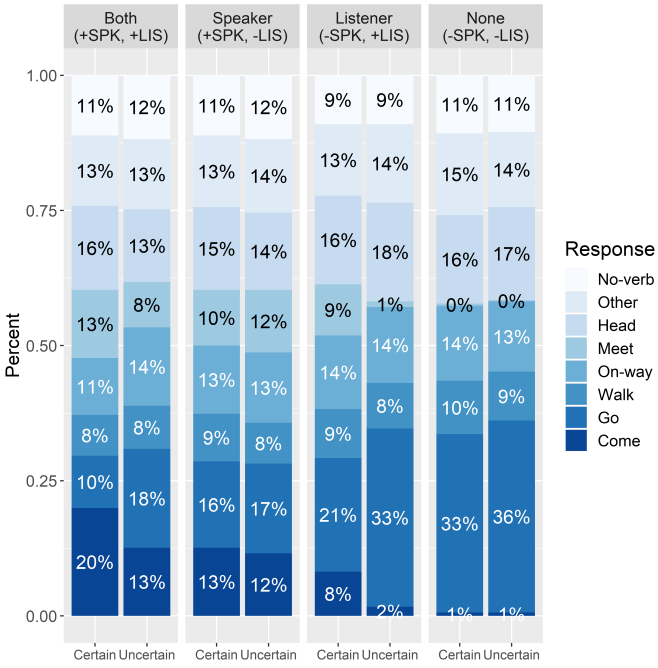


Figure 2: Verb use across the eight conditions.

Table 1: Repeated contrasts analysis in the Certainty and Uncertainty conditions. Significant effects ($p < 0.05$) are bolded.

Fixed Effect	Certainty			
	β	SE	z	p
1. BothVsSpeaker	**0.941**	**0.235**	**4.012**	**<0.001**
2. SpeakerVsListener	**0.723**	**0.262**	**2.766**	**0.005**
3. ListenerVsNone	**3.013**	**0.632**	**4.767**	**<0.001**
	Uncertainty			
1. BothVsSpeaker	0.167	0.250	0.668	0.504
2. SpeakerVsListener	**2.686**	**0.434**	**6.188**	**<0.001**
3. ListenerVsNone	1.046	0.702	1.490	0.136

[3]Two alternative annotation schemes – one using the *first* verb of the response and another using the *last* verb of the response – yielded the same qualitative and statistical results.

[4]We further note that the combined rates of *come* and *go* are roughly similar across conditions, meaning that *go* trades off with *come*. However, we do not consider *go* here, since its status as a perspectival verb is a subject of debate in the literature. Most other verbs are not affected by our manipulations, with the exception of *meet*, which depends on someone being at the destination. We will explore patterns involving *go* and *meet* in future work.

[5]This contrasts with the results of Anderson (2020), who found no significant difference between the Both and Speaker conditions. The difference may arise due to aspects of the experimental design: our task uses maps that emphasize spatial delineation of venues and participant locations.

speakers privilege their own presence at the destination more than the listener's; we discuss the implications of this for our proposal of perspective mixing in the modeling section below.

Computational Modeling with MPM

MPM is a probabilistic approach to modeling pragmatics based on the fundamental assumption that language users simultaneously weigh both their own and their interlocutors' (conversational) perspectives (Heller et al., 2016). Thus far, evidence for this approach has been found in the production and comprehension of referring expressions in situations where the speaker and listener have different knowledge about objects, such as objects occluded from the view of one interlocutor. In such cases, both perspectives must be taken into account to correctly model the patterns of how speakers formulate referring expressions (Mozuraitis et al., 2018; Heller & Stevenson, 2018), and of how listeners interpret them (Heller et al., 2016), including in dyadic interactions (Ryskin et al., 2020).

Here we explore whether the MPM mechanism of simultaneous perspective weighing can also account for how speakers use the verb *come* which depends on lexical perspective. If, as we hypothesize, the use of lexical perspectival expressions is governed by the same general mechanisms of (conversational) perspective weighing, then MPM – a model designed to capture conversational perspective taking – should be able to capture patterns in the use of verbs like *come* without having to add any new, special mechanisms specific to lexical perspective.

In addition, we use MPM to further examine whether and how speakers weigh perspectives in our task when they are uncertain about the listener's perspective. Other work has adopted the MPM perspective weighing approach in modeling speaker uncertainty about what the listener can see in a reference production task (Hawkins et al., 2021). As noted earlier, referential tasks are different because, in such a task, not taking uncertainty into account can potentially lead to communicative failure. In contrast, in our task, speakers could produce a motion verb that does not depend on perspective (e.g., *walk* or *head*). Our work here thus extends the understanding of how speakers deal with uncertainty. In the following subsections, we use MPM to test various hypotheses of how uncertain knowledge affects perspective weighing.

MPM and Probability Estimates

In language production (Mozuraitis et al., 2018), MPM models the choice of a linguistic expression e given the meaning m to be expressed, taken over all perspectives a, as:

$$P(e|m) = \sum_a P(e|m,a)\,P(a) \tag{1}$$

The standard case is dyadic communication where a refers to either the speaker's (S) or listener's (L) perspective:

$$\begin{aligned} P(e|m) &= P(e|m,a=S)\,P(a=S) + P(e|m,a=L)\,P(a=L) \\ &= \omega\,P(e|m,a=S) + (1-\omega)\,P(e|m,a=L) \end{aligned} \tag{2}$$

where we use ω as shorthand for $P(a=S)$. The ω parameter captures the extent to which a speaker weighs their own perspective, where a value of 1 indicates a speaker who is completely egocentric, and 0 indicates a speaker who has completely adapted to the listener's perspective. Finding the optimal value of ω allows us to assess how much each perspective is weighed: this is the value that provides the best fit of the model predictions (probabilities) to the experimental findings (proportions of expressions used by participants).

Our critical hypothesis here is that the perspective weighing mechanism of MPM can be directly used (without any extensions) to account for lexical perspective. To test this, we use the setting of the perspective variable, $a \in \{S, L\}$, to also represent the setting of the *anchor* for a perspectival expression; namely, we treat conversational perspective and lexical perspective as the same, such that the anchor for a usage of *come* must be the conversational perspective taker. For our experiment, since we focus on modeling usage of the lexical perspectival verb *come*, we set the expression e to be *come* (a usage of the lemma *come*); a then is the available perspective/anchor for *come*.

We consider the meaning m to be those relevant aspects of the situation that license a felicitous use of *come*: because *come* needs a perspectival anchor at the destination, m is used to encode whether the speaker and/or the listener are at the destination. We represent the speaker as at, or not at, the destination as $dest(S)$ or $\neg dest(S)$, respectively, and represent the listener location analogously. Thus each meaning m is a pair of such specifications; for example, for the Speaker condition, $m = [dest(S), \neg dest(L)]$. However, given our critical assumption that the anchor for lexical perspective is equivalent to the conversational perspective taker (a in Eqn. (2)), the meaning m thus simplifies when considered from the perspective of either the speaker or the listener. That is, while the full meaning m for each experimental condition specifies the value of *dest* for both the speaker and the listener, only the value of $dest(a)$, where a is the current perspective taker, matters for a usage of *come*.

For example, to predict the rate of *come* in the Speaker condition of the experiment, we instantiate Eqn. (2) as:

$$\begin{aligned} &P(come|\text{Speaker}) \\ &= \omega\,P(come|m=[dest(S),\neg dest(L)], a=S) \\ &\quad + (1-\omega)\,P(come|m=[dest(S),\neg dest(L)], a=L) \qquad (3) \\ &= \omega\,P(come|dest(S)) \\ &\quad + (1-\omega)\,P(come|\neg dest(L)) \qquad (4) \end{aligned}$$

where Eqn. (3) can be simplified as in Eqn. (4) because of our assumption that the anchor of the perspectival expression, *come*, must be the same as the conversational perspective taker (a). This means that, here, from the speaker's perspective, when $a = S$, the only part of the meaning that matters is $dest(S)$, and from the listener's perspective, when $a = L$, the only part of the meaning that matters is $\neg dest(L)$.

We can analogously express the probability of *come*

in the other three experimental conditions (Both, Listener, and None) as an ω-weighted combination of $P(come|dest(S))$ or $P(come|\neg dest(S))$ with $P(come|dest(L))$ or $P(come|\neg dest(L))$. The question now is, how do we estimate each of these probabilities?

First, we assume that the speaker's preference for using *come* is the same regardless of which perspective they are reasoning from, their own (S) or the listener's (L). That is, $P(come|dest(S)) = P(come|dest(L))$ and $P(come|\neg dest(S)) = P(come|\neg dest(L))$. This means we need only estimate two values, $P(come|dest)$ and $P(come|\neg dest)$, where (bare) *dest* indicates whether the perspective taker is at the destination.

Since there is no norming data for determining reasonable estimates of these two probabilities, we consider some of our experimental conditions as base values of the probabilities of interest, and use those to estimate the values of the probabilities in other conditions (cf. Mozuraitis et al. (2018)). In particular, we estimate these probabilities from the behavioral data in the Both-Certain and None-Certain conditions, and use these two values to derive the predictions for the other six conditions:

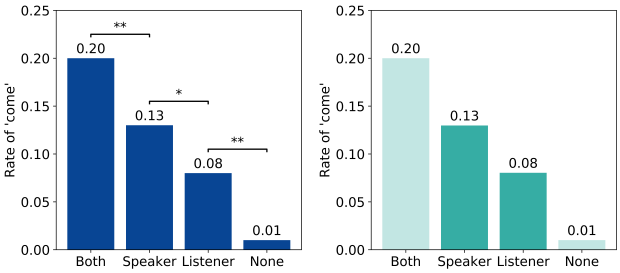
$$\begin{aligned} P(come|\text{Both}) &= \omega\, P(come|dest) + (1-\omega)\, P(come|dest) \\ &= P(come|dest) \\ P(come|\text{None}) &= \omega\, P(come|\neg dest) + (1-\omega)\, P(come|\neg dest) \\ &= P(come|\neg dest) \end{aligned}$$

We thus estimate $P(come|dest)$ from the observed mean proportion of *come* responses in the Both-Certain condition (0.20), and $P(come|\neg dest)$ from the observed proportion of *come* responses in the None-Certain condition (0.01).[6]

Modeling Certainty Trials

To test whether lexical perspective can be treated as a special case of conversational perspective taking, we need to examine whether combining perspectives as described above can provide a good fit to our experimental data. We first examine the Certainty conditions: we ask whether by using estimates based on the empirical data in Both-Certain and None-Certain, we can find a value of the perspective weighting parameter ω that predicts the values of the Speaker-Certain and Listener-Certain conditions.

Searching over values 0–1 in increments of 0.01, we find an optimal ω of 0.63 that minimizes the sum of squared error ($SSE < 0.0001$) between the model's predicted probabilities of *come* and the empirical proportions of *come* in the Certainty trials.[7] Figure 3 shows our model's predictions



(a) Experimental results (b) Model predictions ($SSE < 0.0001$)

Figure 3: Certainty conditions: (a) Human data, and (b) modeling results. In (b), conditions used to estimate base probabilities are shown in a lighter color, since they are *not* model predictions.

compared to the human values, highlighting the remarkable fit MPM achieves despite its simplicity. Importantly, this excellent fit depends on mixing perspectives, and cannot be achieved by considering only a single perspective, either the speaker's or the listener's. This result supports our hypothesis that the use of lexical perspectival expressions can be naturally accommodated within MPM's general mechanism of integrating conversational perspectives.

Modeling Uncertainty Trials

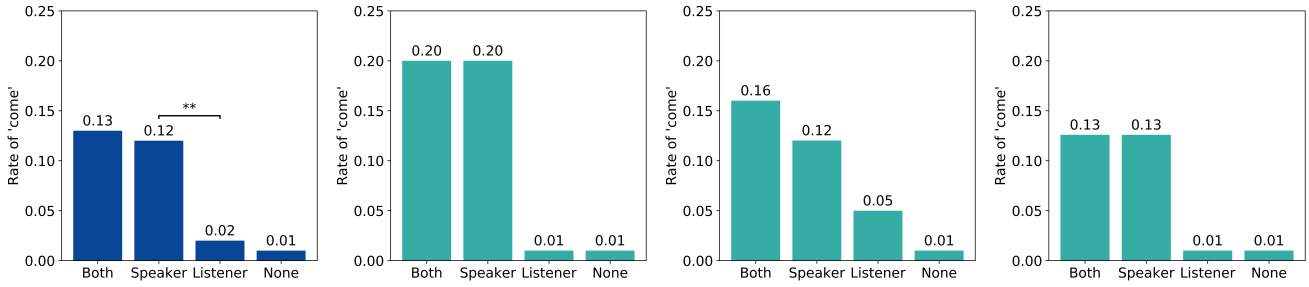
Our uncertainty manipulation enables us to explore the impact of the speaker's uncertain knowledge about the listener's perspective, in a task where (in contrast to earlier work) simply avoiding perspectival expressions that depend on the uncertain knowledge could be an effective communicative strategy. In line with the MPM view that all interaction involves perspective weighing, our hypothesis was that speakers would continue to use *come* in a way that takes the listener's perspective into account even in this situation.

The Egocentric Model (Figure 4b). We must first test whether, contrary to our hypothesis, speakers simply behaved egocentrically, ignoring the listener's perspective because they lack reliable information about it, instead only considering their own. This is implemented in MPM by setting $\omega = 1$, which allots all the weight to the speaker's perspective. Under this assumption, the model's predictions are equivalent to $P(come|dest)$ in the Both-Uncertain and Speaker-Uncertain conditions, and $P(come|\neg dest)$ in the Listener-Uncertain and None-Uncertain conditions; see Figure 4b. This model predicts a much higher rate of *come* in Both and Speaker than we find in the experimental results, incurring an *SSE* of 0.0114, a much poorer fit than in the Certainty conditions above.

It may seem surprising that the Egocentric Model yields a higher rate of *come* in the Speaker-Uncertain condition than what we found in our earlier modeling of the Certainty conditions (see Figure 3b). How can the rate of *come* be higher

[6] We use Both-Certain and None-Certain as base conditions because each involves an ω-weighted sum of the same value ($P(come|dest)$ or $P(come|\neg dest)$, respectively), simplifying to 1 times that value. Similar results are obtained when we instead solve the system of equations for $P(come|dest)$ and $P(come|\neg dest)$ using the Speaker-Certain and None-Certain conditions, showing the robustness of the approach to how these probabilities are estimated.

[7] The SSE is calculated only on the two predicted values, for Speaker and Listener, and not on the Both and None values we used as base conditions. (But note that including those does not change the SSE, since they are identical to the empirical values.)



(a) Experimental results

(b) Egocentric Model ($SSE = 0.0114$)

(c) Two Listeners Model ($SSE = 0.0018$)

(d) Non-Dest Listener Model ($SSE = 0.0002$)

Figure 4: Uncertainty conditions: (a) Human experimental results, (b–d) modeling results for the three models considered, where all probabilities are predictions of the model.

here when only considering that the speaker is present at the destination? This happens because, in the Egocentric Model, the speaker is not weighing in the perspective of a listener *not* at the destination, with its lower rates of *come*, as we did in the modeling of the Certainty conditions in Figure 3. Indeed, the poor fit of the Egocentric Model to the human data suggests that speakers are taking the listener perspective into account, as expected under MPM.

The Two Listeners Model (Figure 4c). One possibility for modeling the listener's perspective is to create *two separate* listener representations, one for each of Indigo's possible locations. This results in a model which combines three perspectives, one for the speaker and two for the listener, splitting the weight of the listener among these two perspectives. For example, in the Listener-Uncertain condition in Figure 1, we would have one listener perspective where the listener is at the destination (Indigo at the fire station) and another where the listener is *not* at the destination (Indigo at the book store). Here we find that the best ω is 0.58, similar to the $\omega = 0.63$ we saw for the Certainty conditions. This means that a weight of $\omega = 0.58$ is allotted to the speaker perspective, and the remaining weight of $(1-\omega) = 0.42$ is split evenly among the two listener perspectives (0.21 each). This model, with *SSE* of 0.0018, yields a much better fit to the human data than the Egocentric Model (see Figure 4c).

However, the Two Listeners Model's predictions are not consistent with the human data. Specifically, it predicts increased rates of *come* when the listener could be at the destination: Both-Uncertain is higher than Speaker-Uncertain, and Listener-Uncertain higher than None-Uncertain; see Figure 4c. The absence of such differences in the human data suggests that, while speakers are not ignoring the listener (as in the Egocentric Model), they are not considering the possibility that the listener is at the destination.

The Non-Destination Listener Model (Figure 4d). Our final model encodes the possibility that the speaker acts conservatively and, lacking certainty about the listener's location, assumes that the listener is *not* at the destination. This model weighs both the speaker's perspective and a *single* listener's perspective. Instead of splitting the listener's probability of $(1-\omega)$ between the two possible perspectives of the listener, as in the Two Listeners Model, here we allot the entire probability $(1-\omega)$ to one listener who is not at the destination. For example, this model of a conservative speaker would assume, for the situation in Figure 1, a listener perspective in which Indigo is at the bookstore (which is not the destination).

The model predictions in Figure 4d show a very strong qualitative fit to the empirical data, and we find the lowest *SSE* yet of the possible models of uncertainty, at 0.0002. Interestingly, the optimal ω, of 0.60, is very similar to what we found for the Certainty conditions. This result supports the MPM view that interlocutors are continually engaging in perspective weighing, while shedding light on how speakers integrate listener perspectives in the presence of uncertainty.

General Discussion

Our starting point was the proposal that lexical perspectives can be subsumed under the more general weighing of conversational perspectives. We began examining this question by asking how speakers produce motion verbs, some of which are known to be perspectival (e.g., *come*); our production experiment was inspired by the paradigm of Anderson (2020). We found that the presence of both interlocutors at the destination increased the rate of usage of the perspectival motion verb *come* over the cases when only one interlocutor was at the destination, revealing that the perspectives of both interlocutors was considered. These results differ from those of Anderson (2020); this may be due to design details. We demonstrated that the Multiple Perspectives Model (MPM) attains an excellent fit to our experimental data, providing initial support for the view that lexical perspectives can be subsumed under conversational perspectives.

A key novel aspect of our experiment was the introduction of uncertainty about the listener's location. An intuitive prediction is that, when a speaker is uncertain about the listener's perspective, they will exhibit more egocentric behavior (i.e., weighing their own perspective more). Importantly, even in these cases, the best fit of the model to the human data was achieved by integrating the speaker and listener perspectives. This result lends further support to the fundamental claim of

MPM that perspective integration is an inherent aspect of language use. Interestingly, we find that speakers assume a listener's perspective where the listener is *not* at the destination. This may be due to Grice's Maxim of Quality: When uncertain about the listener's location, the speaker conservatively behaves as though the listener is not at the destination, in order to not trigger a potentially false presupposition.

In the current experiment, we considered cases where only the speaker's and the listener's perspectives were considered as possible anchors for lexical perspectival elements. However, a verb like *come* can also take a non-interlocutor as the perspective holder, as in *Felix thinks that Rose is coming to the fire station*, where Felix – but not the speaker or the listener – is at the fire station. While this possibility goes beyond the scope of the current study, we note that since the MPM formulation can accommodate any number of perspectives (cf. Eqn. (1)), it has the potential to be extended to model these cases as well.

Usage of perspectival lexical elements has previously been modeled using Perspectival RSA (PRSA; Anderson & Dillon, 2019; Anderson, 2020), which sought to account for perspectival expressions by mixing two RSA models: one based on truth-conditions from the speaker's perspective, and another based on truth-conditions from the listener's perspective. While PRSA is able to account for the empirical patterns similar to MPM, this extension of RSA was thought of as a specialized mechanism tailored to account for a limited set of lexical items. In contrast, MPM views perspective weighing as an inherent aspect of language: Lexical perspective is modeled through the very same perspective weighing mechanism used for all communication.

The finding that speakers weigh multiple perspectives in producing the verb *come* is particularly interesting because in our experiment, speakers are not required to use perspectival expressions in order to succeed at the task of communicating Rose's location to Indigo; instead, they could say *walk* or *on her way*. This contrasts with previous work on referential communication, where perspective taking was essential for communicative success. For example, Hawkins et al. (2021) found perspective combination in situations where speakers directed a listener to choose an object while being uncertain about the array of objects available to the listener. In such communicative situations, a rational approach would be for the speaker to attend to the listener's perspective to ensure that the object description picks out the intended object in the listener's array. In our task, however, including perspectival information is not required for successful communication, and so a rational approach might be for speakers to choose a different strategy under uncertainty. We nevertheless find that speakers weigh the two perspectives – their own and the listener's – using a similar weight across the certain and uncertain trials (i.e., about 0.60). This finding is similar to Heller & Stevenson (2018) who found that speakers used similar weights independent of how this might affect communicative success in two different referential situations. Taken together, these findings indicate that interlocutors do not "turn off" perspective integration even when it is not required for communicative success, supporting the idea that integrating multiple perspectives is an inherent aspect of communication.

References

Anderson, C. (2020). *Shifting the perspectival landscape*. Unpublished doctoral dissertation, UMass Amherst.

Anderson, C., & Dillon, B. W. (2019). Guess who's coming (and who's going): Bringing perspective to the rational speech acts framework. *Proc. of SCiL*, *2*(1), 185–194.

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

Hawkins, R. D., Gweon, H., & Goodman, N. D. (2021). The division of labor in communication: Speakers help listeners account for asymmetries in visual perspective. *Cognitive Science*.

Heller, D., Gorman, K. S., & Tanenhaus, M. K. (2012). To name or to describe: Shared knowledge affects referential form. *Topics in cognitive science*, *4*(2), 290–305.

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 *Proc. of CogSci.*

Hockett, C. F. (1990). Bring, take, come, and go. *Journal of English Linguistics*, *23*(1-2), 239–244.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *JEP:Gen*, *116*, 26.

Kaiser, E., & Lee, J. H. (2017). Experience matters: A psycholinguistic investigation of predicates of personal taste. In *Semantics and linguistic theory* (Vol. 27, pp. 323–339).

Kaplan, D. (1979). On the logic of demonstratives. *Journal of philosophical logic*, *8*(1), 81–98.

Lasersohn, P. (2005). Context dependence, disagreement, and predicates of personal taste. *Linguistics and philosophy*, *28*(6), 643–686.

Mozuraitis, M., Stevenson, S., & Heller, D. (2018). Modeling reference production as the probabilistic combination of multiple perspectives. *Cognitive science*, *42*, 974–1008.

Oshima, D. Y. (2006). Motion deixis, indexicality, and presupposition. In *Semantics and linguistic theory* (Vol. 16).

Ryskin, R., Stevenson, S., & Heller, D. (2020). Probabilistic weighting of perspectives in dyadic communication. In *Proc. of CogSci.*

Schad, D., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *JML*, *110*, 104038.

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.