Language universals rely on social cognition: Computational models of the use of *this* and *that* to redirect the receiver's attention

Marieke Woensdregt (marieke.woensdregt@mpi.nl)
Language and Computation in Neural Systems, Max Planck Institute for Psycholinguistics, The Netherlands

Julian Jara-Ettinger (julian.jara-ettinger@yale.edu)
Department of Psychology, Yale University, USA

Paula Rubio-Fernandez (paula.rubio-fernandez@ifikk.uio.no)
Department of Philosophy, Classics, History of Art & Ideas, University of Oslo, Norway

Abstract

Demonstratives—simple referential devices like *this* and *that*—are linguistic universals, but their meaning varies cross-linguistically. In languages like English and Italian, demonstratives are thought to encode the referent's distance from the producer (e.g., *that one* means "the one far away from me"), while in others, like Portuguese and Spanish, they encode relative distance from both producer *and* receiver (e.g., *aquel* means "the one far away from both of us"). Here we propose that demonstratives are also sensitive to the receiver's focus of attention, hence requiring a deeper form of social cognition than previously thought. We provide initial empirical and computational evidence for this idea, suggesting that producers use demonstratives to redirect the receiver's attention towards the intended referent, rather than only to indicate its physical distance.

Keywords: pragmatics; deictic communication; Theory of Mind; computational modelling

Introduction

Linguistic communication is a thoroughly social phenomenon: It requires the producer and receiver of a message to consider each other's mental states in order to make themselves understood (Brown-Schmidt, Yoon, & Ryskin, 2015; Grice, 1957; Rubio-Fernández, 2020; Sperber & Wilson, 1986). The question remains, however: how deep do the demands that language places on social cognition run? Is it only a matter of pragmatics (taking into account context), or does grammar also hinge on social cognition? Here we address that question by investigating a unique class of words that is present in all of the world's languages, and emerged early on in the evolution of language: demonstratives (e.g., *this* and *that*; Diessel, 2003). Key to our investigation, demonstratives serve two related functions: (1) they indicate the location of a referent relative to the *deictic center* (e.g., the speaker's position in English), and (2) they coordinate the interlocutors' joint focus of attention (Diessel, 2006, 2012a, 2012b).

While demonstratives are a universal tool for joint attention (Diessel, 1999, 2003), they exhibit great cross-linguistic variability: depending on the language, demonstratives may indicate not only the distance, but also the altitude, familiarity, position, reachability or visibility of a referent, from the perspective of the producer, the receiver, or both (Levinson, 2018). Thus, demonstratives can have different meanings (i.e., different *semantics*), despite always being used to establish joint attention (i.e., similar *pragmatics*). Here, we investigate how the meaning of different demonstrative systems (Study 1) and their pragmatic use (Study 2) hinge on social cognition.

Linguistic typology distinguishes between *distance-oriented* and *person-oriented* demonstrative systems (Diessel, 2013). In distance-oriented systems, demonstratives indicate the distance of a referent from the producer's position. For example, in Italian, the proximal form *questo* is used when the referent is close to the speaker, and the distal form *quello* when it is far away from the speaker. By contrast, in person-oriented systems, demonstratives indicate the distance of a referent not only from the producer's position, but also from the receiver's. For example, in Spanish (which has three demonstratives), the proximal form *este* is used when the referent is close to the speaker, the medial form *ese* when it is far from the speaker but close to the listener, and the distal form *aquel* when it is far away from both.

From the point of view of social cognition, distance-oriented and person-oriented systems make different perspective-taking demands. Thus, in languages with person-oriented systems, producers must monitor the receiver's spatial location to accurately use demonstratives (e.g., depending on whether the listener is close or far from the referent, a Spanish speaker may use *ese* or *aquel*). By contrast, in distance-oriented systems, producers always use demonstratives from their own, egocentric perspective (e.g., an Italian speaker would refer to a far-away object as *quello*, regardless of the listener's position).

Regarding the pragmatics of demonstratives, here we tested a novel prediction: given their universal function to establish joint attention, producers of all languages should be sensitive to their receiver's focus of attention when using demonstratives. Thus, if a listener is looking further away from the referent, an Italian speaker may use the proximal form 'questo' (gloss: *Look over here!*), even if the object is not particularly close to her. Reversely, if the listener is looking closer, the speaker may use the distal form 'quello' (gloss: *Look over there!*).

We investigated the demands that demonstrative use poses on social cognition in two studies. In Study 1, we develop two computational models of demonstrative use, one distance-oriented and one person-oriented, which we tested in two languages with distance-oriented systems (English and Italian) and two languages with person-oriented systems (Portuguese and Spanish). In Study 2, we developed a variant of the base-

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

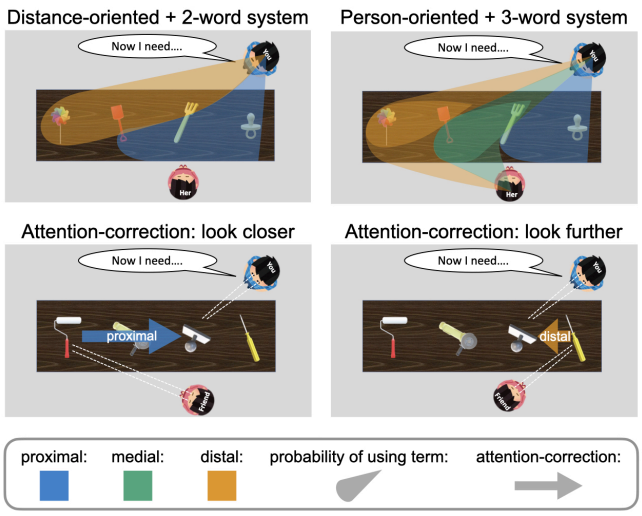

Figure 1: Sample experimental trials with added visualisations of the distance-oriented and person-oriented models (top) and the attention-correction mechanism (bottom). Coloured cones indicate the probability that the pragmatic speaker will produce each demonstrative for a given referent. The attention-correction arrows indicate for which demonstrative (proximal/distal) the production probability increases when the listener is looking in the wrong direction.

line models including an *attention-correction* mechanism (i.e. demonstratives are used flexibly depending on the receiver's attention), which we tested in the same four languages.

Study 1: Monitoring listener spatial location

We tested the typological analysis of English and Italian demonstratives (as two-way distance-oriented systems) and Portuguese and Spanish (as three-way person-oriented system) using two novel computational models of demonstrative use and an online experiment.For the languages in our sample, both languages with a two-way demonstrative contrast (e.g. *questo/ quello*) are distance oriented, whereas both languages with a three-way distinction (e.g. *este/ ese/ aquel*) are person oriented. However, these features do not necessarily co-occur. For example, Turkish has a three-way system (*bunal şunal ona*) but is distance oriented (Ozyurek, 1998).

Computational framework

For clarity, we explain our computational model in the context of our experimental setup (shown in Figure 1). Here, a speaker and a listener stand on opposite sides of a table with four objects. The speaker asks the listener for one of the objects using a demonstrative.

Our computational model is a hybrid of the *Incremental Collaborative Efficiency* (ICE) framework (Jara-Ettinger & Rubio-Fernandez, 2020) and the *Rational Speech Act* (RSA) framework (Frank & Goodman, 2012; Goodman & Frank, 2016). That is, our model is structured around two key ideas: First, the ICE framework posits that speakers go beyond producing messages that have sufficient information to recover the intended referent; instead, speakers also aim to help the listener identify the referent quickly and efficiently (Rubio-Fernandez, Mollica, & Jara-Ettinger, 2021). In the context of physical reference, this implies that speakers take into account the listener's expected visual search for the referent. That is, among two equally-informative utterances, speakers will prefer whichever helps the listener locate the referent faster. Second, the RSA framework posits that interlocutors engage in recursive social reasoning to derive the pragmatic meaning of words. Our model can therefore be thought of as having three layers: a simple speaker that produces demonstratives that best support visual search in literal listeners; a pragmatic listener that adjusts their visual search by considering why the speaker selected the demonstrative they did; and a pragmatic speaker that produces demonstratives by reasoning about the expected visual search of a pragmatic listener.

At the highest level, our computational framework associates different demonstratives with different patterns of visual search in a context-sensitive manner. We achieve this by assigning *association strengths* to referents, which capture the degree to which a demonstrative applies to an object given its location (e.g., *"this"* is strongly associated with objects close to the speaker). We assume a visual search strategy that prioritizes looking at strongly-associated objects (i.e., look at objects that are most likely to be the intended referent). This visual search strategy allows the speaker to estimate the expected time the listener will need to identify the correct referent (quantified in number of fixations), and select the demonstrative that minimizes listener search time. (All code used is available at: https://github.com/mariekewoensdregt/demonstratives_model).

Distance-oriented semantics The baseline distance-oriented model captures a system where the semantics of demonstratives are sensitive only to the position of the speaker. Below we describe for each demonstrative term how we define its association strengths under the distance-oriented model. First, for proximal demonstratives (e.g., *this* in English or *este* in Spanish), objects closer to the speaker have a higher association strength. Put more formally, the association strength of a referent with the proximal demonstrative is *inversely* proportional to the referent's distance from the speaker: $-|pos_r - pos_S|$ (where pos_r is the referent's position, and pos_S the speaker's position). Conversely, for distal demonstratives (e.g., *that* in English or *aquel* in Spanish), objects farther from the listener have a higher association strength. Thus, put formally, the association strength of a referent with the distal demonstrative is *directly* proportional to the referent's distance from the speaker: $|pos_r - pos_S|$.

Finally, medial demonstratives (e.g., *ese* in Spanish) encode intermediate distances. One possible way to implement this is with an association strength function that peaks around the central distribution of objects. Alternatively, however, this

meaning of medial demonstratives can emerge through pragmatic reasoning: medial demonstratives cover the intermediate region, because speakers would be wiser to use the proximal or distal demonstratives on the edges of the scene. We therefore set medial demonstratives as a uniform association strength function, and allow pragmatic reasoning to naturally constrain its meaning to the central region of the space.

Person-oriented semantics The baseline person-oriented model captures a system where the semantics of demonstratives are sensitive to the position of both the speaker and the listener. First, typological analyses posit that in person-oriented systems with three demonstratives, proximal demonstratives signal proximity to the speaker (Diessel, 2013). The association strength function of the proximal demonstrative is therefore identical to the one from the baseline distance-oriented model described above. Second, in person-oriented systems, medial demonstratives (e.g. *ese* in Spanish) signal referents that are far from the speaker but close to the listener. Put formally, a referent's association strength is defined by adding up its distance from the speaker (i.e. farther from speaker yields higher association strength) to the *inverse* of its distance from the listener (i.e. closer to listener yields higher association strength): $|pos_r - pos_S| - |pos_r - pos_L|$ (where pos_L stands for the listener's position). Finally, for distal demonstratives (e.g., *aquel*) a referent has high association strength if it is far from *both* the speaker and the listener. Thus, put formally, a referent's association strength is defined by adding up its distance from the speaker (i.e. farther from speaker yields higher association strength) to its distance from the listener (i.e. farther from listener yields higher association strength): $|pos_r - pos_S| + |pos_r - pos_L|$.

Production behaviour Below we describe how the context-sensitive semantics defined above (in terms of the referents' association strengths) get turned into visual search cost estimates, which are then used by the speaker to determine which demonstrative to produce in a given situation.

Simple speaker Given the context-sensitive semantics specified above, we normalize the association strengths that the different referents have for the simple listener ($A_L(r|w)$) to a common scale in the [0,1] range, and transform them into a probability distribution through softmax. The resulting distribution represents the probability of the listener fixating on each potential referent upon hearing a given demonstrative:

$$P_{L_{simple}}(r_{fixation}|w) \propto e^{A_L(r|w)/\tau_L} \quad (1)$$

where τ_L is a rationality parameter that modulates the influence of context-sensitive semantics on visual search. When τ_L is low, the listener will always fixate on objects in strict order of association strength. As τ_L increases, the listener performs a more noisy visual search.

Through this process, our simple speaker forms a belief about how the listener will search for the referent after hearing a demonstrative. Because this belief is probabilistic, we compute the expected visual search (obtained via Monte Carlo simulations with $n = 1000$ samples) associated with each demonstrative. Finally, the speaker produces utterances approximately rationally: trying to minimise the simple listener's search cost for the intended referent:

$$P_{S_{simple}}(w|r, pos_S, pos_L) \propto e^{-C_{L_{simple}}(r|w)/\tau_S} \quad (2)$$

where $C_{L_{simple}}(r|w)$ is the expected visual search cost given a demonstrative w to identify referent r, and τ_S is the speaker's rationality. This parameter is analogous to the one from Eq. 1, but now modulates speaker behaviour (rather than listener visual search). When τ_S is low, the speaker always selects the demonstrative associated with the lowest expected search cost. As τ_S increases, the speaker's behaviour becomes more noisy: more likely to choose suboptimal demonstratives.

Pragmatic speaker Finally, we added a layer of recursive social reasoning, using the RSA framework (Frank & Goodman, 2012; Goodman & Frank, 2016). This pragmatic speaker assumes as their audience a pragmatic listener, who in turn assumes they receive utterances produced by the simple speaker described above.

The pragmatic listener's fixation probabilities (Eq. 3) , are based on the simple speaker's production probabilities (Eq. 2), in order to infer the probability of referent r given that the speaker produced word w. The pragmatic listener fixates on referents approximately rationally: trying to maximise the probability that the referent they look at is indeed the simple speaker's intended referent, given the demonstrative received, again modulated by the listener's rationality parameter τ_L:

$$P_{L_{prag}}(r_{fixation}|w, pos_S, pos_L) \propto e^{P_{S_{simple}}(w|r,pos_S,pos_L)/\tau_L} \quad (3)$$

Analogously to the simple speaker, the pragmatic speaker assumes that the pragmatic listener's visual search is expressed in the probabilities given by Eq. 3. This enables the pragmatic speaker to estimate the listener's expected visual search for the referent (implemented via Monte Carlo simulations with $n = 100$ samples). The pragmatic speaker then produces utterances approximately rationally: trying to minimise the pragmatic listener's search cost, modulated by rationality parameter τ_S:

$$P_{S_{prag}}(w|r, pos_S, pos_L) \propto e^{-C_{L_{prag}}(r|w)/\tau_S} \quad (4)$$

Experiment 1: Manipulating listener position

In Experiment 1, we obtained explicit judgments about demonstrative use in four languages, which we then compare to our baseline computational models.

Methods

Participants 200 native speakers of English (from the UK), Italian (from Italy), Portuguese (Peninsular), and Spanish (Peninsular) ($n = 50$ per language) were recruited through Prolific and performed our task in Qualtrics.

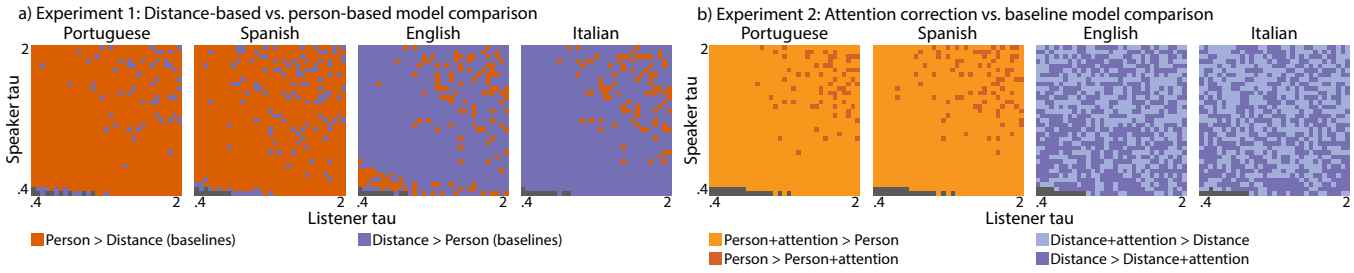

Figure 2: Model comparisons for Experiments 1-2. a) Comparison between *person-oriented* and *distance-oriented* models against data from Experiment 1. The person-oriented model best fit data from Portuguese and Spanish speakers, while the distance-oriented model best fit data from English and Italian speakers for a wide range of parameters. b) Evaluation of whether the attention correction mechanism adds additional explanatory value against data from Experiment 2. Our results reveal clear evidence of attention correction for Portuguese and Spanish, and ambiguous evidence for English and Italian.

Materials and Procedure Stimuli consisted of 16 displays showing a speaker (figure labelled "You") and a listener ("Her") on opposite sides of a table with four objects (Fig. 1). The speaker appeared at the top right-end of the table across trials, while the listener's position varied parametrically with each object position. The speaker's attention indicated the target object in each trial, illustrated through both body orientation and line of gaze (represented by dashed lines). Target position was counterbalanced across trials, fully crossed with the listener's position in a 4x4 design.

Participants were asked to adopt the role of the speaker and complete a request for the target object (*"Now I need..."*), by choosing a demonstrative out of two or three options, depending on the language (e.g. *this*/ *that* in English; *este*/ *ese*/ *aquel* in Spanish). Participants were asked to imagine that they were in the physical situation depicted in each display and had to select the expression they would be more likely to use.

Results of Study 1

Our model enables us to generate quantitative predictions about distance-oriented and person-oriented systems, for languages with either two or three demonstratives (i.e., with what probability the pragmatic speaker would produce the various demonstratives in a given situation). However, these models require that we set the τ_L and τ_S parameters. We therefore began by computing model predictions for a range of these parameters ($[0.4, 2.0]$ in steps of 0.05τ), and performing model comparison via Bayes factors (using a uniform prior over models). As Figure 2a shows, our computational models enabled us to extract which demonstrative system is used in each language, in a robust manner, as parameter setting had little effect on our results. This model-based analysis and our experiment confirmed past typological analyses. The behavioural data of our English and Italian participants is best explained by the distance-oriented model, suggesting that speakers of these languages use demonstratives to mark the relative distance of referents from their own position. The data for Portuguese and Spanish, by contrast, is best explained by the person-oriented model, suggesting speakers of these languages use demonstratives to mark the distance of referents relative to both the speaker's and listener's position. Table 1 shows mean Bayes factors and their corresponding evidence strength (Lee & Wagenmakers, 2014).

Lang.:	**Distance-oriented:** #; mean BF; evidence	**Person-oriented:** #; mean BF; evidence
Port.	55; $9.577e+05$; extreme	1013; 0.012; very strong
Span.	114; $2.085e+09$; extreme	954; 0.016; very strong
Eng.	930; $2.041e+18$; extreme	141; 0.182; moderate
It.	980; $3.980e+35$; extreme	91; 0.107; moderate

Table 1: Number of parameter settings that favour each model for Experiment 1, with corresponding mean Bayes factors and evidence strength. For Portuguese and Spanish, the majority of parameter settings fall in the column where the person-oriented model fits best, whereas for English and Italian the majority falls in the distance-oriented column (cf. Fig. 2a).

Figure 3 shows the proportions with which participants used the various demonstratives in particular sample trials of interest, alongside the corresponding model predictions. To generate these model predictions, we used maximum likelihood estimation (through grid approximation in steps of 0.05τ) to identify the best parameter setting. These trials illustrate that in a distance-oriented language like English (top panel of Figure 3), demonstrative choice is sensitive to the referent's position (relative to the speaker), but not to the listener's position. That is, when the target is in Position 2 (i.e. one position away from the speaker), both the behavioural data and the model predictions show that the proximal and distal terms are roughly equally likely to be selected. However, when the target is in Position 4 (i.e. furthest away from the speaker), the distal term is preferred (both in the model and the behavioural data). Crucially, *within* these two situations (target in Position 2 and target in Position 4), neither the behavioural data nor the model predictions distinguish

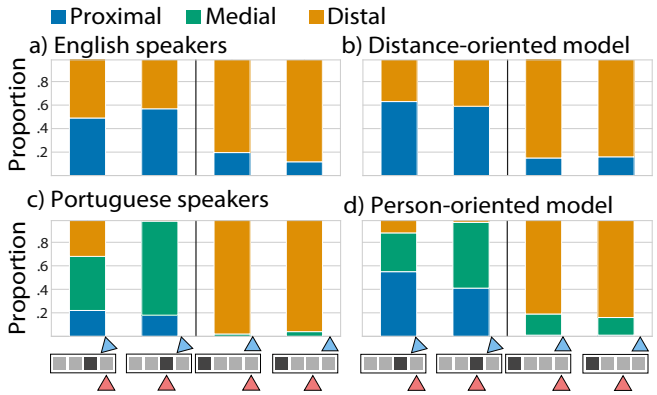

Figure 3: Experiment 1 sample trials. Top row (a-b) shows 4 trials by English speakers and the distance-oriented model. Bottom row (c-d) shows the same 4 trials by Portuguese speakers and the person-oriented model. Bottom diagrams show abstract schematics of the trial: each square represent an object, with the target in black. The blue and red triangles represent the speaker and listener location, respectively.

whether the listener is in Position 1 or 2. This shows that listener position does not affect demonstrative choice in English (and analogously in Italian). If we compare this to a person-oriented language like Portuguese (bottom of Figure 3), we see that both the behavioural data and the model *are* sensitive to listener position. That is, when the target is in Position 2 and the listener in Position 1, all three demonstrative terms can be used. However, when the listener moves to Position 2 (i.e. right in front of the target), we see that the medial term is preferred (suggesting that the medial demonstrative means "far from me *but close to you*" in Portuguese, and analogously in Spanish).

Study 2: Monitoring listener visual attention

The goal of Study 2 is to expand our computational model and experimental setting to investigate whether speakers monitor the listener's visual attention and use demonstratives flexibly in order to *redirect* the listener towards the intended referent.

Integrating attention-correction into our models

We developed an *Attention-correction model* extension, which follows the same structure as our baseline models but modifies the association strength function of the proximal and distal demonstratives. For proximal demonstratives, the attention-correction mechanism boosts the association strength of any referent closer to the speaker relative to the listener's *focus of attention*. Conversely, for distal demonstratives, the attention-correction mechanism boosts the association strength of any referent that is farther from the speaker relative to the listener's attention. Formally, we achieve this by simply adding a constant value of 1 to the association strength function for those referents that are (i) closer than the listener's attention for the proximal term, and (ii) further away than the listener's attention for the distal term. The rest of the model works in an identical manner, creating a speaker that now also uses demonstratives flexibly to direct the listener's attention, and the listener reacts accordingly.

Experiment 2: Manipulating listener attention

Experiment 2 consisted of a task similar to Experiment 1, with the difference that the stimuli now revealed (and parametrically varied) the listener's attention, enabling us to test if people use demonstratives as attention-redirection devices.

Methods

Participants 200 native speakers of English (from the UK), Italian (from Italy), Portuguese (Peninsular), and Spanish (Peninsular) ($n = 50$ per language) were recruited through Prolific and performed our task in Qualtrics.

Materials and Procedure The stimuli consisted of 18 trials, similar to those in Experiment 1. The two key differences were (i) that the listener was always positioned directly in front of the target object, and (ii) that the listener's attention varied parametrically with object position. Speaker and listener attention were indicated by their body orientation and line of gaze, and the target was always the object that the speaker was looking at. In half the trials, the speaker and listener perspectives were *misaligned* (i.e. looking at different objects), so the speaker would have to redirect the listener's attention. The procedure was the same as in Experiment 1. Of interest was whether participants selected different demonstratives in aligned- vs. misaligned-perspectives trials.

Results of Study 2

Figure 2b and Table 2 show the comparison between the baseline model we derived for each language in Experiment 1 and the same model with an attention-correction mechanism. Adding an attention-correction mechanism improved model fit for Portuguese and Spanish, for virtually any parameter setting. For English and Italian, however, which model comes out as most likely varies greatly across different parameter settings, with no clear pattern.

Lang.:	**Baseline:** ; #; mean BF; evidence	**Attention-correction:** #; mean BF; evidence
Port.	65; 0.126; moderate	999; $1.649e+144$; extreme
Span.	83; 0.087; strong	982; $9.533e+139$; extreme
Eng.	583; 0.068; strong	489; $1.315e+54$; extreme
It.	545; 0.053; strong	532; $4.950e+32$; extreme

Table 2: Number of parameter settings that favour each model for Experiment 2, with corresponding mean Bayes factors and evidence strength. For Portuguese and Spanish, the Attention-correction model fits best for the majority of parameter settings. English and Italian show no clear pattern.

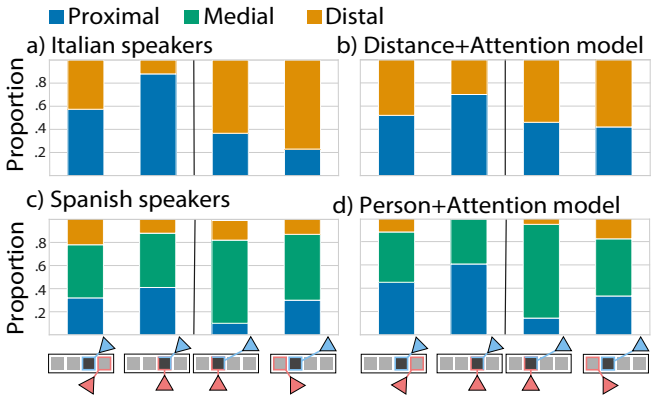

Figure 4: Experiment 2 sample trials. Top row (a-b) shows 4 trials by Italian speakers and the distance+attention model. Bottom row (c-d) shows the same 4 trials by Spanish speakers and the person+attention model. Bottom diagrams show abstract schematics of the trial. Each agent's attention is depicted through a solid line from the agent to an object, as well as a coloured outline corresponding to the colour of the agent.

Figure 4 shows sample trials alongside model predictions from Experiment 2. Here we can compare two cases of misalignment in Italian and Spanish: When the listener's attention is focused closer to the speaker than the target object (target in Position 2 and listener attention on Position 1, counting from right to left), speakers of both languages are more likely to use the distal term compared to when speaker and listener attention are aligned (on Position 2). Furthermore, languages with person-oriented systems like Spanish also show that when the listener's attention is focused further away than the target object (target in Position 3 and listener attention on Position 4), speakers are more likely to select the proximal demonstrative than when speaker and listener perspectives are aligned (target and listener attention on Position 3).

Discussion

We set out to investigate how language use recruits social cognition by focusing on demonstratives—a unique class of words that is present in all of the world's languages and whose use is pervasive in everyday social interaction (Diessel & Coventry, 2020). To better understand how deep the demands that demonstratives place on social cognition are, we started by distinguishing the semantics of different demonstrative systems (i.e. their grammatical meaning) and their pragmatics (i.e. how they are used to establish joint attention with the receiver). We predicted that social cognition could be recruited at both levels, depending on the language.

We developed two novel computational models of demonstrative use, one for distance-oriented systems (which indicate referent distance from the producer's position) and another for person-oriented systems (which indicate distance from both producer and receiver). These models are based on the ICE framework (Jara-Ettinger & Rubio-Fernandez, 2020)—which rests on the assumption that speakers produce helpful referential expressions by considering the listener's visual search, and on the RSA framework (Frank & Goodman, 2012; Goodman & Frank, 2016)—which captures pragmatic inferences through recursive social reasoning. In Study 1, these baseline models and an online demonstrative-choice experiment were used to test typological analyses of four different demonstrative systems. The results confirmed that English and Italian have distance-oriented systems, whereas Portuguese and Spanish have person-oriented systems.

The use of computational models to test typological analyses of different languages has the potential to make an important contribution to both linguistics and typology since there is not always a consensus on how to characterise a given demonstrative system. For example, there is a longstanding debate on the nature of the Spanish system (Peeters, Krahmer, & Maes, 2021), which our model predictions and human data seem to resolve in favor of the person-oriented description. Regarding social cognition, our results also confirm that different demonstrative systems place different perspective-taking demands on their users. That is, person-oriented systems require that producers monitor the receiver's spatial location to accurately use demonstratives, while distance-oriented systems do not require switching perspectives.

In Study 2, an attention-correction mechanism was incorporated into the baseline models to investigate whether producers of all languages use demonstratives flexibly depending on whether their perspective is aligned with the receiver's or not. Being sensitive to the receiver's focus of attention would be efficient since demonstratives are used to establish joint attention across languages (Diessel, 2006). Model comparison confirmed that Portuguese- and Spanish-speakers use demonstratives flexibly to redirect the listener's attention to the intended referent. In English and Italian, while evidence in favour of the attention-correction model is stronger than for the baseline model, there is no clear model that fits best across the range of parameter settings. The human data, however, does suggest that English- and Italian-speakers use demonstratives flexibly depending on the listener's attention focus.

We consider two possible reasons for the mixed results of Study 2. First, it is possible that simply by having two demonstratives at their disposal (instead of three), English- and Italian-speakers reveal less attention correction than Portuguese- and Spanish-speakers. Second, post-hoc analyses showed that for the 2-way distance-oriented models, the differences in model predictions between the baseline and attention-correction variants are very small for most trials compared to the 3-way person-oriented models. We are therefore planning to run a third experiment including more critical trials to try to address these open questions.

In conclusion, the results of our studies confirm that using demonstratives—one of the building blocks of human language—requires social cognition. Future studies should explore the implications that cross-linguistic differences and universals in demonstrative use may have for human social cognition and its development.

Acknowledgments

This research was supported by funding from the Research Council of Norway (275505) awarded to PRF, and NSF (BCS-2045778) awarded to JJE.

References

Brown-Schmidt, S., Yoon, S. O., & Ryskin, R. A. (2015). People as Contexts in Conversation. In *Psychology of Learning and Motivation* (Vol. 62, pp. 59–99). Elsevier. doi: 10.1016/bs.plm.2014.09.003

Diessel, H. (1999). The morphosyntax of demonstratives in synchrony and diachrony. *Linguistic Typology*, *3*(1), 1–49. doi: 10.1515/lity.1999.3.1.1

Diessel, H. (2003). The relationship between demonstratives and interrogatives. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, *27*(3), 635–655. doi: 10.1075/sl.27.3.06die

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*, *17*(4), 463–489. doi: 10.1515/COG.2006.015

Diessel, H. (2012a). Deixis and demonstratives. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *An international handbook of natural language meaning* (Vol. 3, pp. 2407–2431). Berlin: Mouton de Gruyter.

Diessel, H. (2012b). Where do grammatical morphemes come from? On the development of grammatical markers from lexical expressions, demonstratives, and question words. In K. Davidse et al. (Eds.), *Grammaticalization and language change: New reflections.* Amsterdam: John Benjamins.

Diessel, H. (2013). Distance contrasts in demonstratives. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from `https://wals.info/chapter/41`

Diessel, H., & Coventry, K. R. (2020). Demonstratives in Spatial Language and Social Interaction: An Interdisciplinary Review. *Frontiers in Psychology*, *11*, 3158. doi: 10.3389/fpsyg.2020.555265

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

Goodman, N. D., & Frank, M. C. (2016). Pragmatic Language Interpretation as Probabilistic Inference. *Trends in Cognitive Sciences*, *20*(11), 818–829. doi: 10.1016/j.tics.2016.08.005

Grice, H. P. (1957). Meaning. *The Philosophical Review*, *66*(3), 377–388.

Jara-Ettinger, J., & Rubio-Fernandez, P. (2020). The social basis of referential communication: Speakers construct reference based on listeners' expected visual search. *PsyArXiv*. doi: 10.31234/osf.io/fzuvh

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press.

Levinson, S. C. (2018, July). Introduction: Demonstratives: Patterns in Diversity. In S. Levinson, S. Cutfield, M. Dunn, N. Enfield, S. Meira, & D. Wilkins (Eds.), *Demonstratives in Cross-Linguistic Perspective* (First ed., pp. 1–42). Cambridge University Press. doi: 10.1017/9781108333818.002

Ozyurek, A. (1998). An analysis of the basic meaning of Turkish demonstratives in face-to-face conversational interaction. In S. Santi, I. Guaitella, C. Cave, & G. Konopczynski (Eds.), *Oralite et gestualite: Communication multimodale, interaction: Actes du colloque ORAGE 98* (pp. 609–614). Paris: L'Harmattan.

Peeters, D., Krahmer, E., & Maes, A. (2021). A conceptual framework for the study of demonstrative reference. *Psychonomic Bulletin & Review*, *28*(2), 409–433. doi: 10.3758/s13423-020-01822-8

Rubio-Fernández, P. (2020, July). Pragmatic markers: The missing link between language and Theory of Mind. *Synthese*. doi: 10.1007/s11229-020-02768-z

Rubio-Fernandez, P., Mollica, F., & Jara-Ettinger, J. (2021). Speakers and listeners exploit word order for communicative efficiency: A cross-linguistic investigation. *Journal of Experimental Psychology: General*, *150*(3), 583.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (First ed.). Blackwell Publishing.