AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Learning, Memory, and Cognition

ISSN: 0278-7393

2025, Vol. 51, No. 10, 1662–1681
https://doi.org/10.1037/xlm0001465

The Relative Importance of Language, Gaze, and Gesture in Deictic Reference

Gozdem Arikan, Peter Boddy, and Kenny R. Coventry
School of Psychology, University of East Anglia

When people communicate, they use a combination of modalities—speech, gesture, and eye gaze—to engage and transmit information to an addressee. Spatial deictic communication is a paradigmatic case, with spatial demonstratives (*this/that*) frequently co-occurring with eye gaze and pointing gestures to draw the attention of an addressee to an object location (e.g., *this cup*, *that chair*). Yet the effectiveness of these individual modalities in guiding attention has not been established. In two experiments, we manipulated pointing, gazing, and spatial demonstratives to establish their relative and combined effectiveness in directing attention to a specific referent. Participants saw an image (Experiments 1 and 2) or a short video clip (Experiment 2) with a person (agent) sitting behind a table, describing (*this, that*), gazing, and/or pointing at the items placed proximally or distally relative to the agent. All three modalities individually affected which of the two objects participants thought the person in the picture was referring to. However, pointing was the dominant cue to referent choice, with demonstratives on their own acting as a relatively weak spatial deictic cue. Overall, the effect of spatial demonstratives (Experiments 1 and 2) and gaze (Experiment 2) on attention to a referent was enhanced when coupled with pointing, both when targeting the distal and proximal positions. The results help to illuminate why spatial deictic communication is usually multimodal, with individual modalities contributing different communicative functions in the act of spatial deixis.

Keywords: deictic communication, demonstratives, pointing, eye gaze

Spatial deixis is paradigmatic of communication and serves to manipulate an interlocutor's focus of attention toward a referent (Diessel, 2006; Diessel & Coventry, 2020; Kita, 2003; Levinson, 1983). This can be conveyed through the modalities of language use (e.g., the spatial demonstratives *this/that/here/there*), eye gaze, and pointing gestures (Iverson et al., 1994; Liszkowski et al., 2012).

In practice, spatial demonstratives, eye gaze, and gesture often work in tandem during spatial deictic communication (Carpenter et al., 1998; Talmy, 2018). For example, Todisco et al. (2021) analyzed the spatial deictic communications of caregivers and infants during naturalistic storybook interactions, finding that infants use demonstratives, pointing, and eye gaze synchronously in communication. Moreover, some linguists have argued that it is obligatory to point or gaze when using specific spatial demonstratives in some languages (e.g., *Goemai*, Hellwig, 2003; *Kilivila*, Senft, 2004; *Yucatec*, Bohnemeyer, 2018; *Warao*, Herrmann, 2018; *Tiriyó*, Meira, 2018). This begs the question as to why such deictic cues usually co-occur. Using multiple cues is more effortful than employing single cues, and therefore one can question if multiple cues confer a greater tendency for speaker and hearer to establish joint attention to the speaker's intended referent compared to single cues alone. Spatial demonstratives, eye gaze, and gesture are all assumed to be powerful deictic tools in communication, but their relative strength in deictic communication has yet to be understood. Moreover, while one of the key functions of demonstratives is to direct the attention of a hearer to an object in space, just how effective demonstratives are as spatial deictic cues has not been established.

The main objectives of the current work are to establish the relative importance of language, gesture, and eye gaze as cues guiding attention to a potential referent. In doing so, we compare congruent and incongruent modalities in spatial reference choice, while also probing dynamic versus static signaling for reference. The end goal is to understand just how effective individual and multiple cues are in deictic reference and, in doing so, to illuminate why deictic communication is invariably multimodal.

This article was published Online First March 31, 2025.

L. Robert Slevc served as action editor.

Gozdem Arikan https://orcid.org/0000-0001-5097-6560
Kenny R. Coventry https://orcid.org/0000-0003-2591-7723

All data, analyses, and study materials are available to the public on the Open Science Framework at https://osf.io/96djz/. The authors report no disclosures. This study was funded through a PhD studentship awarded to Gozdem Arikan at the University of East Anglia.

Gozdem Arikan played a lead role in data curation, formal analysis, and methodology and an equal role in conceptualization, writing–original draft, and writing–review and editing. Peter Boddy played a supporting role in formal analysis and methodology. Kenny R. Coventry played an equal role in conceptualization, methodology, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to Gozdem Arikan or Kenny R. Coventry, School of Psychology, University of East Anglia, Research Park, Norwich NR4 7TJ, United Kingdom. Email: gozdemarikan@gmail.com or k.coventry@uea.ac.uk

Prior to presenting a series of experiments manipulating deictic cues across modalities, we briefly review evidence for the importance of individual deictic cues before considering their relative importance.

Single and Multiple (Modal) Deictic Cues

Demonstratives

The function of spatial demonstratives—words including *this*, *that*, *here*, and *there* in English—has been much debated in recent years (see Diessel & Coventry, 2020, for a recent review). It is generally accepted that demonstratives serve two closely related functions: indicating spatial location and coordinating speaker–listener (joint) attention (Diessel, 2006). With regard to the indication of spatial location, many have argued that demonstratives contrast distance, with a proximal term used for an object (referent) near a speaker and a distal term for an object far from a speaker (Anderson & Keenan, 1985; Bühler, 1934; Diessel, 2014). In English (and a wide range of other languages), it has been shown that *this* and *that* (and their equivalents in other languages) map onto a more precise peripersonal and extrapersonal space distinction (Caldano & Coventry, 2019; Coventry et al., 2008, 2014, 2023; see also Kemmerer, 1999). For example, Coventry et al. (2008, 2014, 2023) systematically manipulated the distance an object was positioned in the sagittal plane on a table in front of participants and found that *this* was used significantly more when the object was reachable by the speaker while *that* was used more when the object was out of reach. Moreover, extending reach through the use of a tool led to a corresponding extension of the use of *this* to describe object location beyond the end of the hand to the end of the tool (Coventry et al., 2008). So, *this* can be regarded as a cue for an addressee to look for an object within the peripersonal space of the speaker while *that* should draw attention to a referent distal from the speaker. Indeed, such mapping has been confirmed via event-related potentials when a heard demonstrative is incongruent with the expected proximal–distal mapping (Stevens & Zhang, 2013).

While distance/reachability is important for demonstrative choice in English, it is also the case that demonstrative use across languages is affected by a wider range of parameters (Coventry & Diessel, 2024; Coventry et al., 2014; Diessel & Coventry, 2020; Levinson et al., 2018; Peeters et al., 2021), including the relative positions of speaker and hearer (e.g., Coventry et al., 2008; Jungbluth, 2003; Rubio-Fernandez, 2022), attention/gaze of the addressee (e.g., Küntay & Ozyürek, 2006; Rubio-Fernandez, 2022), and properties of the referent and environment, including referent visibility (e.g., Chandralal, 2010; Coventry et al., 2014), familiarity (Coventry et al., 2014), and elevation (Forker, 2020). Moreover, it is also recognized that languages have "default" demonstratives that can be used simply to serve an attention function. In English, *that* is the dominant demonstrative form, and in past studies, it has been shown that *this* is seldom used outside peripersonal space, while *that* can be used irrespective of the distance an object is located for the speaker (e.g., within peripersonal space; see Diessel & Monakhov, 2023; Levinson, 2018, for discussion). Exactly how these multiple parameters and functions of demonstratives affect attention to a referent has yet, however, to be elucidated.

Gestures and Eye Gaze as Unimodal Cues

Gestures can be defined as "visible action when it is used as an utterance or as part of an utterance" (Kendon, 2004, p. 7). Included in such gestures are pointing with the arm, or other body parts, but also shrugs and nods. Pointing is a ubiquitous form of indicating (Clark, 2003; Cooperrider et al., 2021; Kendon, 2004; Talmy, 2018). It appears to be universal across cultures (Cooperrider et al., 2018; Eibl-Eibesfeldt, 1989) and developmentally is among the earliest forms of communication (Bates, 1979; Capirci & Volterra, 2008; Liszkowski et al., 2012). Infants as young as 4–6 months are capable of covertly shifting their attention to the intended reference of a pointing hand (Bertenthal et al., 2014). Moreover, pointing gestures are extremely frequent during spatial communication (Kita, 2003). Furthermore, Cooney et al. (2018) had shown that people are extremely accurate at judging the intended location when a conspecific is pointing with arm outstretched and index finger extended.

There is also much evidence for the power of eye gaze in directing attention. Sensitivity to eye gaze appears early in development, with evidence that infants at 12 months are able to accurately discriminate the object that someone is looking at (e.g., Phillips et al., 2002). Multiple studies with adults show that people automatically shift attention as a function of eye gaze; when people observe someone looking in a particular direction, their attention is shifted to the same location in space (e.g., Driver et al., 1999; Edwards et al., 2015; Friesen & Kingstone, 1998; Hietanen, 1999; Langton & Bruce, 1999, 2000). Such automaticity is reflected in studies showing that a gaze incongruent with a cued target location results in a cost responding to that location, even when participants are told with 100% certainty where the target will appear (Galfano et al., 2012). In deictic communication, Hanna and Brennan (2007) had shown that participants, in a spontaneous dialogue task, use the gaze of their interlocutors to identify targets in a matching task before hearing an object name that disambiguated reference.

Cues in Combination

While language, gesture, and eye gaze are powerful (individual) cues to intended reference, it is less clear regarding their relative contribution in spatial deixis. Several studies have broadly examined the relationship between language and gesture (e.g., Arslan et al., 2024; S. D. Kelly et al., 2010, 2015; S. D. Kelly & Lee, 2012) and the relationship between language, gesture, and gaze (Holler et al., 2014). For example, Holler et al. (2014) presented videos of speech only or speech and gesture object-related messages (nouns), with the speaker either gazing at the participants or sideways to another (out of view) person. They found that participants addressed directly in the speech-only condition were quicker to select the correct object than those in the unaddressed speech-only condition. However, participants in the speech plus (iconic) gesture condition when unaddressed performed as well as those in addressed conditions. Holler et al. argued that speech processing suffers when a speaker's gaze is averted but gesture processing does not, therefore facilitating overall comprehension (the *cross-modal priming hypothesis*). While these results suggest the interplay between language, gaze, and gesture in object-noun processed subsequent recall, only a small number of studies have examined the relationship between pointing

and demonstratives in spatial deixis on the one hand and gaze and demonstratives in spatial deixis on the other.

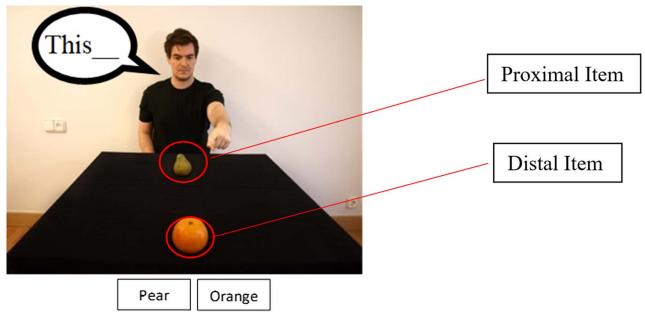
A few studies have examined the interaction between demonstratives and pointing gestures during interactive spatial description tasks. Bangerter (2004) has shown that, when participants gesture to describe object location, they employ short spatial descriptions, but when they do not gesture, descriptions are longer. Additionally, Bangerter found that gesture is more frequent when participants describe proximal locations rather than distal locations (see also Cooperrider, 2016), and Piwek et al. (2008) reported that distal demonstratives were preferred when not pointing over proximal demonstratives. So, while it has been often assumed that demonstratives and pointing are two sides of the same coin, patterns of co-occurrence are likely to be more nuanced.

The only study to directly compare the effects of gesture plus description to description or pointing alone was conducted by Bangerter and Louwerse (2005). Participants viewed video clips of a person describing and/or pointing to an array of objects on a computer monitor. Accuracy rates for correct identification of (location) targets by participants were similarly high when the person in the video gave a spatial description alone or pointed alone, but there was further improvement in target identification in the pointing plus spatial description condition. Using a rather different method, Chieffi et al. (2009) asked participants to read words aloud, with words (the Italian demonstrative equivalents of "here" and "there") on cards placed either close or far from them, while simultaneously pointing at their own body when the card was near and pointing toward a far location when the card was placed far. They found that the congruence of the word with direction of pointing affected both voice spectra and pointing kinematics, suggesting a bidirectional interaction between speech and gesture production systems.

In relation to the interplay between gaze and demonstrative use, García et al. (2017) employed a task where participants had to describe to each other where to place different plastic cups in an array of possible locations. When participants could not see each other's gaze (hidden behind goggles), participants used verbal deixis (demonstratives in particular) less than when they could see each other's gaze, suggesting that gaze may be fundamental to demonstrative use.

Our goal in the experiments that follow was to examine the extent to which language, gesture, and gaze guide attention to a referent in spatial deictic communication for the first time. In a series of online experiments, participants saw images of a person (agent) referring to one of two objects placed in front of them (on the sagittal plane). The task for participants was to identify the object they thought the agent was referring to (Figure 1). Both experiments manipulated whether, in addition to language, the agent only pointed, only gazed, or pointed and gazed at one of the two objects placed in front of them (Figure 2), while manipulating the congruence of verbal (demonstratives) and nonverbal cues (pointing and gaze; Figure 3). We had two primary goals. First, we wanted to test the relative strengths of individual cues versus multiple congruent cues on reference choice. Given that spatial deictic communication usually involves multiple cues, we predicted that multiple cues (e.g., Figure 2a) might be more effective cues to a reference than single cues (e.g., Figure 2b and 2c). This would be consistent with the findings of Bangerter and Louwerse (2005) with respect to pointing and spatial description. Second, to investigate the relative strengths of cues, we wanted to pit cues against each other by manipulating their congruency (e.g., Figure 2e). A further motivation comes from the findings of Holler et al. (2014). In their study, the multimodality of face-to-face conversation was probed by manipulation of shared gaze (speaker's gaze either directed at the listener or averted away from the listener) and co-speech gesture (speaker's gestures depicting object shape, size, or function). The results suggested that, when the speaker did not share gaze with the listener, the presence of a gesture alleviated the effects of averted gaze. Therefore, we predicted that pointing would dominate reference choice when the gaze is on a different referent. In addition to the primary goals, a secondary goal was also to compare cues and combinations of cues in cases where the nonverbal cues are static (shown in photographs) versus dynamic (shown in videos). This both served as a test of ecological validity (videos are more naturalistic) and probed potential differences between the presentation of textual (written) versus spoken demonstratives.

Figure 1
Example of the Task



Note. Participants picked one of the two nouns to indicate which item they thought the person in the image was referring to. The person(s) imaged for the purpose of this research consent to the use of their identifiable pictures in this article (figures) and to the sharing of these pictures with other researchers for research purposes. See the online article for the color version of this figure.

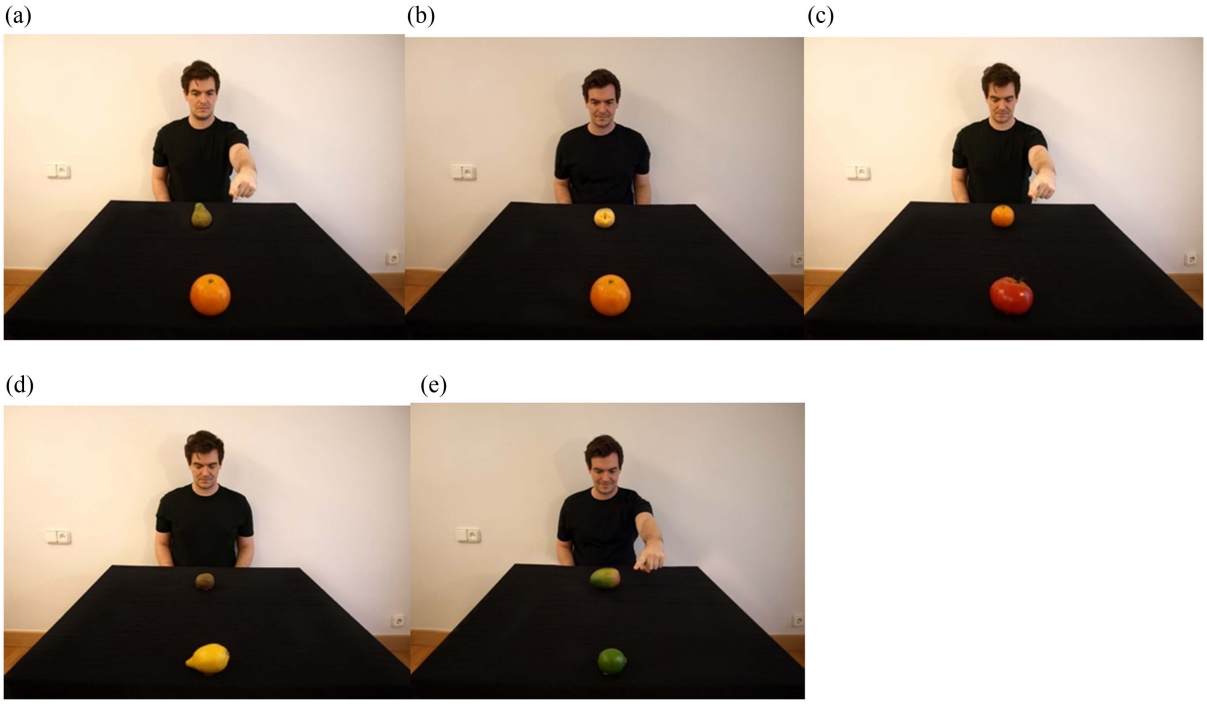
Transparency and Openness

The design of the study and the analysis plan for the acquired data were not preregistered. The raw and processed data, analysis files, and examples of stimuli (full set of images and videos are available on request) are available online at https://osf.io/96djz/. This study received ethical approval from the University of East Anglia, School of Psychology Ethics Committee prior to the commencement of data collection.

Experiment 1

The goal of Experiment 1 was to examine the impact of single versus multiple deictic cues and the congruence of cues on referent choice (Figures 1 and 2). Participants saw images of two objects on a table in front of the agent using language (*this, that*) with or without (congruent or incongruent) pointing and eye gaze to refer to one of the objects. Participants had to decide which of the two objects (a near object and a far object) they thought the agent was referring to.

Figure 2
Examples of Different Conditions



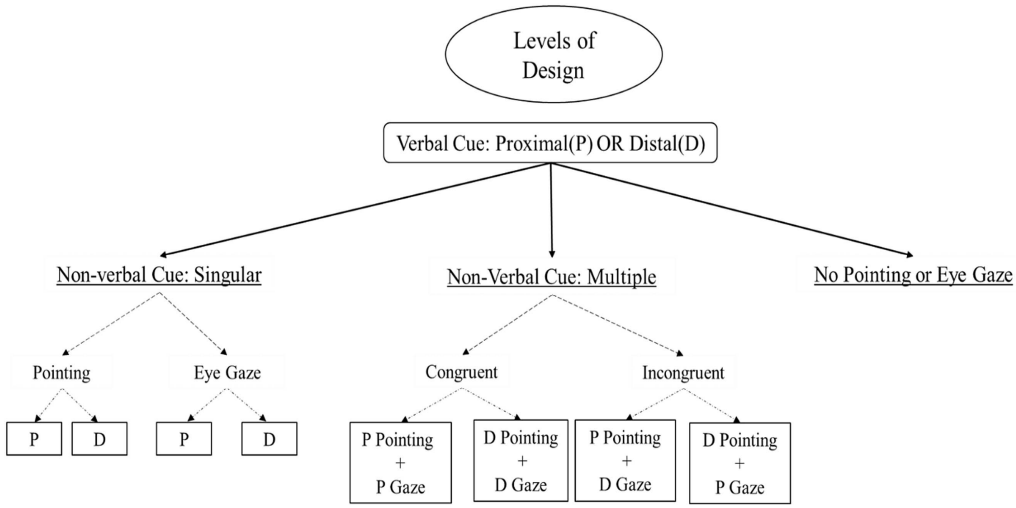
Note. (a) Pointing + eye gaze congruent. (b) Eye gaze only. (c) Pointing only. (d) No pointing or gaze. (e) Incongruent eye gaze and pointing. The person(s) imaged for the purpose of this research consent to the use of their identifiable pictures in this article (figures) and to the sharing of these pictures with other researchers for research purposes. See the online article for the color version of this figure.

Method

Participants

Participants were University of East Anglia students participating in the study for course credit. A power analysis was conducted a priori for sample size estimation, using G*Power (Version 3.1; Faul et al., 2009). The effects and other coefficients were determined based on previously published articles. Research manipulating similar factors (although not directly using the same methodology) such as the relation of referent distance and demonstrative words (Coventry et al., 2014) and gaze and gesture availability/congruency on language processing and use (Cooperrider, 2016;

Figure 3
Overview of Design (With Example Conditions)



Holler et al., 2014; S. D. Kelly et al., 2010; S. Kelly et al., 2015; Perea García et al., 2017) reported medium and medium-to-large effects and differing but mostly strong correlations among these factors. We assumed a similar effect size, and we did not differentiate between levels or their interactions since the experiments are based on the notion of multimodality of language. We followed sample size recommendations of Brysbaert (2019) and Faul et al. (2009). To detect a medium effect ($f = 0.25$) at a significance criterion of $\alpha = .05$ for a power of 80%, a sample size of 36 participants for this experiment was indicated. Brysbaert (2019) argued these measures to be reliable guide when setting up new research, in the absence of similar studies reporting reliable and consistent effect (size).

Participants who did not speak English as their first language were excluded from analyses, leaving a total number of 33 participants ($M_{age} = 21.9$, $SD_{age} = 8.18$; 29 females) after exclusion.

Stimuli and Design

The experiment was presented on the online experimentation platform Gorilla (https://gorilla.sc/). Stimuli on each trial consisted of images of the agent (male) sitting behind a table, pointing, and gazing at one of two items placed on a sagittal plane: one positioned proximally (close) and one distally (far) relative to the agent. The agent always pointed with his left arm. The task was to complete a two-word sentence with a missing word, placed in a speech bubble (Figure 1). The position of the speech bubble was counterbalanced, where half of the participants saw the speech bubble to the left of the agent and half on the right. Response options were positioned under the image, and participants were asked to click and select from one of the two options (proximal or distal item). The presentation order of the object choices was counterbalanced (with the proximal item appearing on the left 50% of the time). Participants could only complete the task on computers (use of mobile phones and tablets was not permitted), thus ensuring the screen size would be large enough to see the images clearly.

There were three main manipulations in the experiment: the spatial demonstrative (verbal cue) presented in the two-word sentence (either *this*, targeting the proximal item, or *that*, targeting the distal item), pointing arm, and eye gaze, in a 2 × 3 × 3 repeated measures design (see Figure 3). The pointing arm varied across three levels: no pointing (the arms of the agent resting on his leg) and pointing to the item closer or farther relative to the agent. The eye gaze of the agent also varied across three levels: no gaze (eyes closed) and gaze targeting the proximal or distal object.

At the beginning of the experiment, participants were introduced to the task. Once informed consent was given, participants completed an example trial, making it clear that their task was to select the missing word the person in the picture would utter according to what object they thought the person was referring to, via a button option beneath the image (Figure 1). Before the example trial, there was a familiarization stage during which participants saw the items they would see through the trials and their associated labels (all fruits with similar rounded shapes). The position of the items on the table was counterbalanced. Each item appeared in each position (near and far), and combinations were unique for each trial.

Each participant completed four blocks with 18 trials in each block, totaling 72 experimental trials, with four data points for each level of the design per participant. The order of blocks and trials within each block was randomized between participants. At the end of the task, participants completed a gaze detection task (with six trials) to check that they were accurate in detecting the gaze of the agent.

Data Preparation

Every time a participant picked the proximal item, they were given a score of 1 and 0 for the distal item. The data were then analyzed in terms of the likelihood of picking the proximal item.

Analysis Model

Statistical analyses were performed in Jamovi (Version 2.2; R Core Team, 2021; The jamovi project, 2021) and analyzed using generalized mixed models (Gallucci, 2019). Due to the binomial nature of the outcome (1 and 0), models had a binomial distribution and logit link function. The outcome was computed as the likelihood of participants choosing the proximal object. We included random coefficients with participants as intercepts and predictors and predictor interactions as random slopes, as long as the model converged, and the model fit was not singular. The random effects structure was formed considering the convergence and theoretical relevance (Barr et al., 2013). The final random coefficient structure included participants and demonstrative seen. Fixed effect factors included pointing, eye gaze, verbal cue, and all possible interactions. Models were estimated using log likelihood (odds ratio, *OR*). The main effect and interaction between the predictors reported in the text are from fixed effects omnibus tests. The fixed effects coefficients were assessed using Wald's *Z* statistics. Fixed effects parameter estimates are reported in tables, and all factors were coded with Helmert's contrast (these are not reported in the text but in tables in the Appendix; Table A1 for Experiment 1, Table A2 for Experiment 2, Table A3 for Experiment 2 static trials, and Table A4 for Experiment 2 dynamic trials). For interpretability, inverse odds are reported when the *OR* is less than 1 (1/*OR*). Unless stated otherwise, the graphs and percentages reported are based on estimated marginal means. Significant interactions and contrasts within a factor were then further investigated with post hoc analyses with Bonferroni correction.

Results

We first examined the gaze detection task to ensure participants were able to identify where the person in the image was looking. The mean gaze detection task score was 5.18 (range of total score = 4–6, $SD = 1.16$) out of a possible perfect score of 6. On that basis, data from all participants were included in the analyses.

The model indicated significant effects of all three factors (conditional $R^2 = .735$), pointing, $\chi^2(2) = 528.98$, $p < .001$; eye gaze, $\chi^2(2) = 146.19$, $p < .001$; and spatial demonstrative, $\chi^2(1) = 10.33$, $p = .001$, and significant interactions between pointing and gaze, $\chi^2(4) = 107.89$, $p < .001$, and pointing and demonstrative seen, $\chi^2(2) = 10.29$, $p = .006$. There was no significant interaction between demonstratives and eye gaze, $\chi^2(2) = .78$, $p = .67$, and the three-way interaction was also not significant, $\chi^2(4) = .88$, $p = .93$.

As expected, the nearer object was selected as the referent when *this* was presented more than when *that* was presented ($OR = 3.19$, $SE = 1.15$, $z = 3.21$, $p < .001$), with participants selecting the nearer object 77% of the time when they saw *this* and 51% of the time when they saw *that*. Again, as expected, participants were more likely to

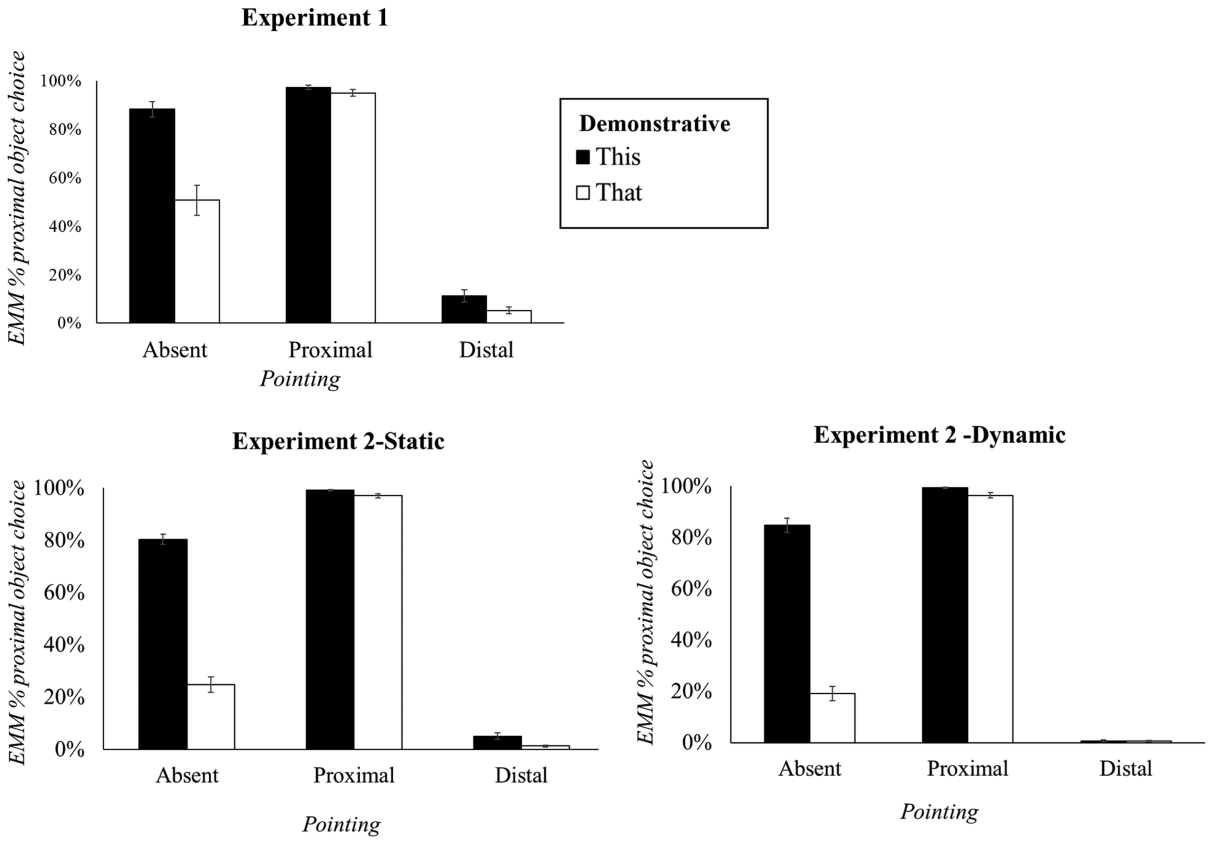
pick the nearer object when the agent was pointing at it as opposed to when they were pointing at the far object, $OR = 313.31$, $SE = 80.43$, $z = 22.39$, $p < .001$, with the selection of the nearer object when the agent was pointing at it 96% of the time and only 8% of the time when they were pointing at the far object. Similarly, when the agent was pointing at the nearer object, participants were around 10 times more likely to pick the proximal item than when there was no pointing, $OR = 9.48$, $SE = 2.29$, $z = 9.33$, $p < .001$. And again, as expected, the overall choice of referent was affected by the object the agent was looking at. When the agent was gazing at the proximal object, participants selected it 83% of the time. On the other hand, when he gazed at the distal object, participants selected the proximal object only 21% of the time. Furthermore, participants were nine times less likely to select the proximal object when the gaze was on the distal item compared to conditions when there was no gaze, $OR = 9.09$, $SE = 0.02$, $z = -9.58$, $p < .001$, and 12 times more likely to select the proximal object when the imaged person was gazing at the proximal object compared to when gazing at the distal object, $OR = 12.88$, $SE = 2.92$, $z = 11.28$, $p < .001$. The contrast of proximal gaze with its absence did not yield significance, $OR = 1.36$, $z = 1.39$, $p = .49$.

As mentioned above, there were also two significant interactions: between pointing and spatial demonstrative seen and between pointing and eye gaze. The interaction between demonstrative and pointing (see Figure 4) reveals two findings. First, the choice of referent was more clear-cut when demonstrative was combined with (congruent) pointing. When participants saw *this* in combination with proximal pointing, they picked the proximal item 97% of the time. When they saw *that* with distal pointing, they picked the proximal item only 1% of the time. Compared to trials with no pointing, participants were five times more likely to choose the closer item when *this* was combined with proximal pointing, $OR = 4.89$, $SE = 1.83$, $z = 4.25$, $p < .001$, and in trials when distal pointing was combined with seeing the word *that*, participants were 20 times less likely to choose the proximal item compared to trials when participants only saw *that*, $OR = 20$, $SE = 0.02$, $z = -9.43$, $p < .001$.

Second, when pointing and demonstrative are incongruent, the choice of the referent is determined by pointing and not by demonstrative. When a distal verbal cue (*that*) was paired with proximal pointing, participants were more likely to pick the proximal item (95% of the time) compared to trials in which both demonstratives and pointing referred congruently to the distal position, $OR = 342.07$, $SE = 124.27$, $z = 16.06$, $p < .001$, or trials when there was a distal demonstrative in the absence of pointing (51% of the time), $OR = 18.38$, $SE = 5.60$, $z = 9.56$ $p < .001$. When the proximal demonstrative, *this*, was paired with distal pointing, participants were less likely to pick the nearer object (11% of the time) compared to trials when both demonstratives and pointing

Figure 4

Bar Charts Illustrating the Interaction Between Spatial Demonstratives and Pointing



Note. Values represent the EMM obtained from the statistical model. Error bars represent the standard error. EMM = estimated marginal means.

target the near item congruently, $OR = 333$, $SE = 001$, $z = -15.63$, $p < .001$, or trials when there was a proximal demonstrative in the absence of pointing (88% of the time), $OR = 50$, $SE = 0.005$, $z = -12.86$, $p < .001$.

There was also a significant interaction between pointing and eye gaze (Figure 5). In the absence of pointing, proximal gaze led to a 95% chance of picking the proximal referent compared to a 10% chance of picking the proximal referent when the gaze was distal, $OR = 215.03$, $SE = 89.10$, $z = 12.96$, $p < .001$. There was also an effect of gaze when pointing is distal, with more frequent choice of the far referent with (congruent) distal gaze (only 3% choice of the proximal object) compared to proximal gaze (at 18%), $OR = 7.32$, $SE = 2.80$, $z = 5.21$, $p < .001$. Critically, when there is a proximal point, there is no effect of gaze, $OR = 1.35$, $SE = 0.27$, $z = -0.82$, $p = 1.00$. In other words, a proximal point appears to be a dominant cue to referent choice. Congruent proximal or distal pointing and gaze did not lead to a significant shift in choice of referent compared to pointing alone without gaze ($OR = 2.94$, $SE = 0.14$, $z = -2.63$, $p = .31$, and $OR = 2.44$, $SE = 0.20$, $z = -1.87$, $p = 1$, respectively) or gaze alone without pointing (compared to trials with both pointing and gaze; $OR = 1.04$, $SE = 0.41$, $z = -0.11$, $p = 1.00$, and $OR = 3.45$, $SE = 0.12$, $z = -2.96$, $p = .11$). Therefore, pointing is the dominant cue to referent choice.

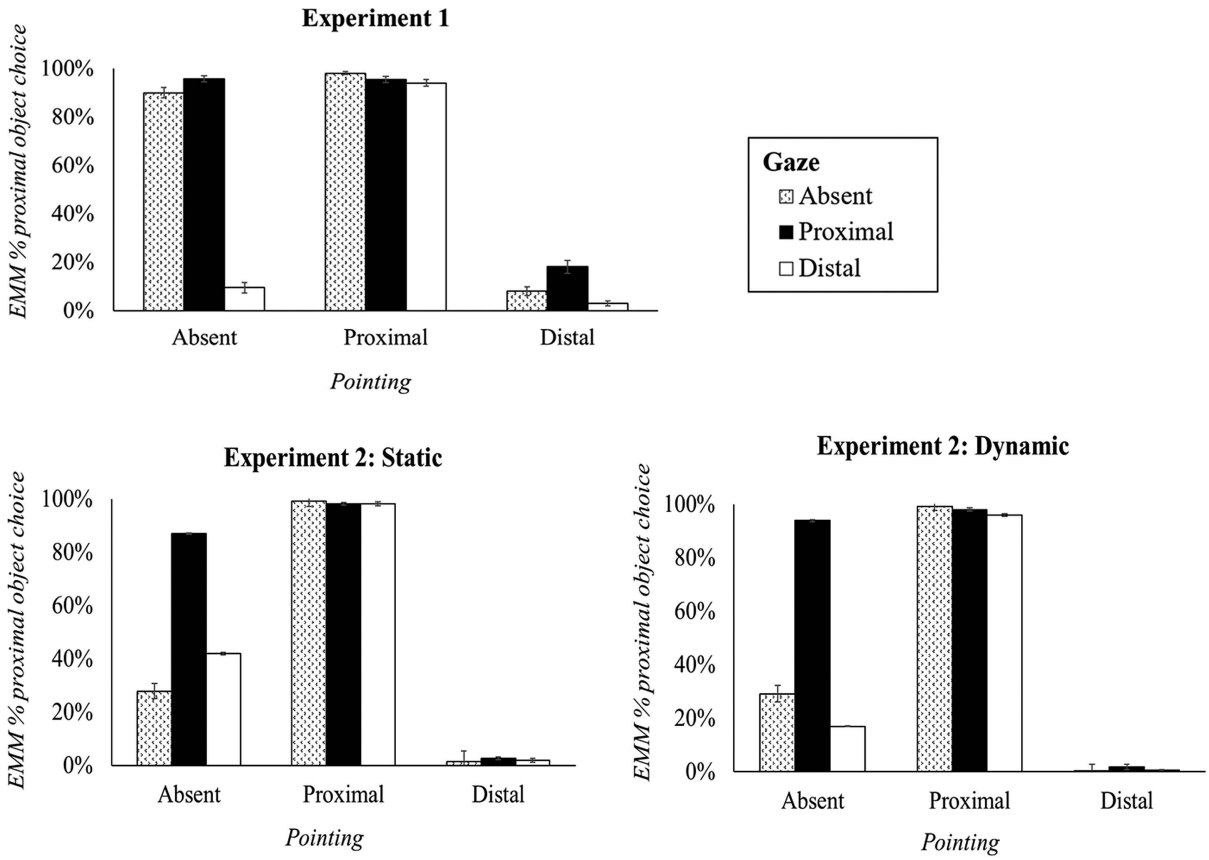
Discussion

Consistent with past studies, participants' choice of referent was influenced by all individual deictic cues, with main effects of language, gesture, and gaze. Results are consistent with previous literature showing that demonstratives are mapped onto perceptual space, as participants were more likely to pick the object within the speaker's reach when the demonstrative *this* was uttered by the speaker and the farther referent when *that* was uttered (Anderson & Keenan, 1985; Caldano & Coventry, 2019; Coventry et al., 2008, 2014, 2023). Regarding the role of gesture in spatial deixis, pointing toward the nearer item led to an increased likelihood of picking the proximal item as referent, compared to trials without any pointing or with distal pointing. In relation to gaze as a cue, it was only effective when targeted toward the farther item, with participants more likely to pick the distal item compared to the no gaze condition (proximal gaze did not affect referent choice compared to trials without gaze).

However, the pattern of interactions offers new insights into the interplay between deictic cues in referent selection. We asked if more deictic cues make it increasingly clear which object the speaker is referring to. Consistent with the results of Bangerter and Louwerse (2005), demonstrative plus (congruent) pointing led to an increased tendency to choose (or likelihood of choosing) the referent

Figure 5

Bar Charts Illustrating the Interaction Between Gaze and Pointing



Note. Values represent the estimated marginal means obtained from the statistical model. Error bars represent the standard error. EMM = estimated marginal means.

cued by the two modalities compared to language alone, both for proximal and distal deictic cues. However, the absence of an interaction between demonstrative and gaze suggests that adding a congruent gaze to linguistic reference for an object did not increase the tendency of choosing this target, indicating that language alone is sufficient.

The pattern of interactions also shows the relative strengths of deictic cues when they are in conflict (incongruent cues). When language and pointing are incongruent, pointing dominates referent choice, overriding any influence of demonstrative. And when gaze and pointing conflict, gaze affects referent choice when pointing is distal but does not affect referent choice in the presence of a proximal point, partially consistent with the previous findings of Holler et al. (2014). Overall, while individual cues affect reference choice, pointing is the strongest predictor of reference in the data when pitted against demonstratives and eye gaze.

One might be tempted to conclude that pointing is the dominant cue, with the other nonverbal cue—gaze—in second place, with demonstratives the weakest cue to reference. However, exophoric use (context-dependent use of demonstratives, referring to a position in the current, shared environment of the interlocutors) of demonstratives is more common in heard/spoken form rather than written form, and it is possible that participants in the first experiment did not attend to the word in the speech bubble as much as they might do when they hear a demonstrative (relative to nonverbal cues). Moreover, using still images omits information about the dynamics of gaze and pointing that may influence action perception and henceforth behavior (Haxby et al., 2020; Welke & Vessel, 2022). For these reasons, Experiment 2 attempted to replicate the findings of Experiment 1 using both dynamic and static images, with language presented auditorily.

Experiment 2

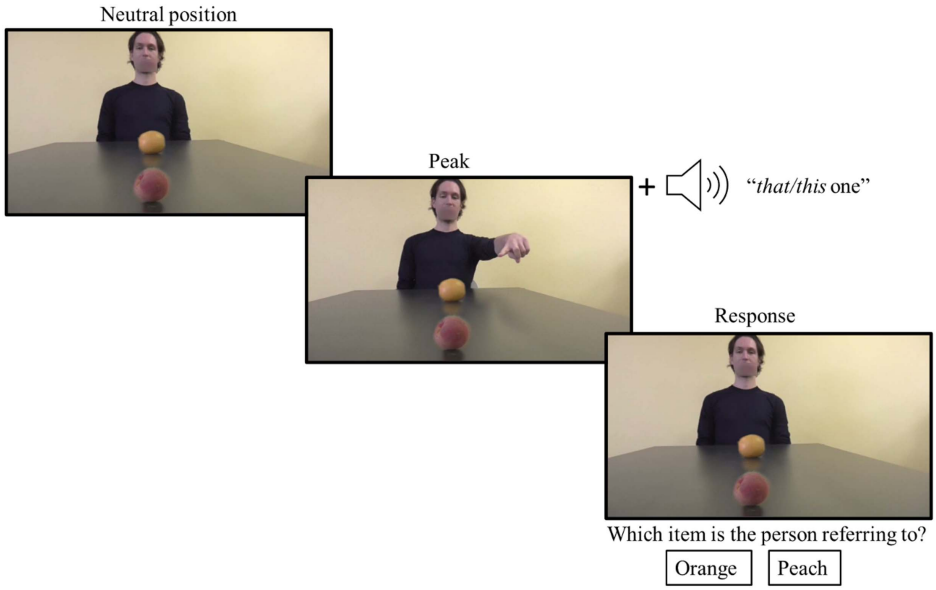
In Experiment 2, participants saw images (static trials) and 3-s videos (dynamic trials) of a person pointing, gazing, and verbally labeling ("this one" or "that one") one of the two potential referents. In contrast to Experiment 1, both images and videos were coupled with audio clips.

The design for Experiment 2 comprised 36 levels; adding "dynamic" and "static" conditions doubled the design compared to Experiment 1 (the rest of the factors/levels were the same as in the previous experiment). Participants saw 18 videos and 18 stills per block, with a total of 36 trials in each block. As in the previous experiment, there were four blocks of experimental trials and a block of gaze detection trials. There was a total of 12 gaze detection trials: six with images and six with videos. As in Experiment 1, participants could only complete the task on computers (use of mobile phones and tablets was not permitted).

The videos in the dynamic condition started with the agent gazing straight ahead, arm resting on his leg. Then, he pointed and looked at one of the two potential referents, labeling the item with "this one" or "that one" (audio clips), and retracted his arm and diverted his eye gaze to the neutral position (Figure 6). The images consisted of stills from the videos at the peak of labeling where the recorded person was pointing, gazing, and verbally labeling one of the two possible referents. Participants were asked to answer "Which item is the person referring to?" by picking the

Figure 6

Example Stimuli From Experiment 2, Static Condition



Note. The trials started with person gazing straight, followed by an image with the agent pointing and/or gazing coupled with audio, "this/that one." The response option was presented in the neutral position. The mouth of the imaged person was masked to avoid extra information being conveyed. The person(s) imaged for the purpose of this research consent to the use of their identifiable pictures in this article (figures) and to the sharing of these pictures with other researchers for research purposes. See the online article for the color version of this figure.

name of one of the two referents. The demonstratives were not written but only heard.

Method

Participants

Participants were a combination of students at the University of East Anglia ($N = 6$), completing the experiment for course credit, and participants from the Prolific recruitment platform ($N = 67$), who were compensated for their time. The participants from the two platforms were all aged between 18 and 35. The sample was not controlled for level of education or current employment status. A priori power analysis indicated around 72 participants for this experiment. Similar to the previous experiment, we ran a prior *F* test with repeated measures analysis of variance within factors, as it was the closest to our design for the data collection. We did not assume differing levels of power for each factor and their interactions, again due to the multimodal nature of demonstrative use. Therefore, once the level of design increased to 36 conditions from 18, the resulting sample size recommendation came to 72. There was a total of 73 participants (female = 46; $M_{age} = 22.6$, $SD_{age} = 3.45$).

Data Preparation

Every time a participant picked the proximal item, they were given a score of 1 and a score of 0 if they selected the distal item. There was a total of 114 data points per participant.

Analysis Model

The same analysis model (generalized mixed model) as Experiment 1 was used but with the addition of stimuli type to the model. In the final model, random components included participants as intercept, demonstratives, pointing, and gaze. Random components were kept to maximal as long as the model converged, and the model fit was not singular.

Results

For the gaze detection trials, the highest possible score for the gaze detection task was 12, with a mean score of 9.86 overall (range of total score = 6–12, $SD = 1.98$), with all participants scoring above chance. Participants were better at detecting gaze in dynamic trials (range = 3–6, $M = 5.30$, $SD = 0.95$) compared to static trials (range = 2–6, $M = 4.56$, $SD = 1.31$), $t(72) = 5.47$, $p < .001$.

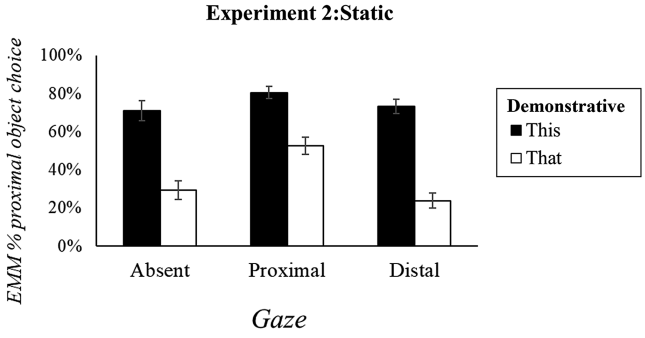
The model testing the effects of stimuli type (videos, stills), pointing, eye gaze, and spatial demonstrative (conditional $R^2 = 0.84$) produced significant main effects of spatial demonstrative, $\chi^2(1) = 61.54$, $p < .001$; pointing, $\chi^2(2) = 543.42$, $p < .001$; and eye gaze, $\chi^2(2) = 43.26$, $p < .001$. There were significant interactions between pointing and demonstrative, $\chi^2(2) = 48.26$, $p < .001$, and between pointing and gaze, $\chi^2(4) = 162.82$, $p < .001$. In the full model, there was also a significant interaction between eye gaze and demonstratives, $\chi^2(2) = 2.43$, $p = .30$. In addition, however, there were also a number of effects/interactions involving stimuli type (static vs. dynamic): a main effect of stimuli type, $\chi^2(1) = 16.56$, $p < .001$; interactions between stimuli type and pointing, $\chi^2(2) = 10.04 = 5$, $p < .001$, and stimuli type and gaze, $\chi^2(2) = 22.48$, $p < .001$; and significant three-way interactions between stimuli type, pointing, and eye gaze, $\chi^2(4) = 12.15$, $p = .002$; stimuli type, pointing, and demonstratives, $\chi^2(2) = 9.71$, $p = .008$; and stimuli type, eye gaze, and demonstratives, $\chi^2(2) = 7.96$, $p = .02$.

The interaction of pointing and demonstrative replicated the results of Experiment 1 (see Figure 4). However, there was a slight difference in Experiment 2 regarding the interaction of pointing and eye gaze. In addition to the pattern found in Experiment 1, in this experiment, congruent pointing and eye gaze targeting the distal position also conferred an advantage over trials without pointing, $OR = 32.7$, $SE = 0.01$, $z = -10.70$, $p < .001$.

Overall, the main effect of stimulus type was also significant, $\chi^2(1) = 16.56$, $p < .001$. During dynamic trials (therefore when stimuli were videos), participants were significantly less likely to pick the proximal item, $OR = 1.62$, $SE = 0.07$, $z = -4.07$, $p < .001$. However, given the multiple interactions involving dynamic and static stimuli, to further understand how the nature of the stimuli affects referent choice, we ran separate analyses for the two stimulus types (see Appendix). The results for the static stimuli largely replicate the results from Experiment 1, with significant main effects as before of demonstrative, $\chi^2(1) = 54.54$, $p < .001$; pointing, $\chi^2(2) = 410.06$, $p < .001$; and gaze, $\chi^2(2) = 32.57$, $p < .001$, and interactions between demonstrative and pointing, $\chi^2(2) = 16.01$, $p < .001$, and between pointing and gaze, $\chi^2(4) = 96.58$, $p < .001$, all yielding the same patterns as Experiment 1 (see Figures 4 and 5). In addition, however, the static stimuli in this experiment also produced a significant interaction between demonstratives and eye gaze, $\chi^2(2) = 7.22$, $p = .03$ (Figure 7). Comparing the effect of demonstrative alone to the congruent demonstrative and gaze condition, there was no significant difference between the single versus congruent proximal, $OR = 1.33$, $SE = 0.20$, $z = -1.05$, $p = 1.00$, or distal cues, $OR = 1.69$, $SE = 0.47$, $z = 1.90$, $p = .86$. Therefore, demonstratives alone were enough in guiding attention. However, when demonstrative and gaze were incongruent, there was a significant effect of (incongruent) gaze when the demonstrative was distal compared to both the single demonstrative condition, $OR = 2.67$, $SE = 0.66$, $z = 3.98$, $p = .001$, and the congruent demonstrative plus gaze condition, $OR = 3.55$, $SE = 0.81$, $z = 3.98$, $p < .001$. When the demonstrative was proximal, however, incongruent gaze did not affect the choice of referent

Figure 7

Bar Charts Illustrating the Interaction Between Spatial Demonstratives and Gaze in Experiment 2 Static Trials



Note. Values represent the estimated marginal means obtained from the statistical model. Error bars represent the standard error. EMM = estimated marginal means.

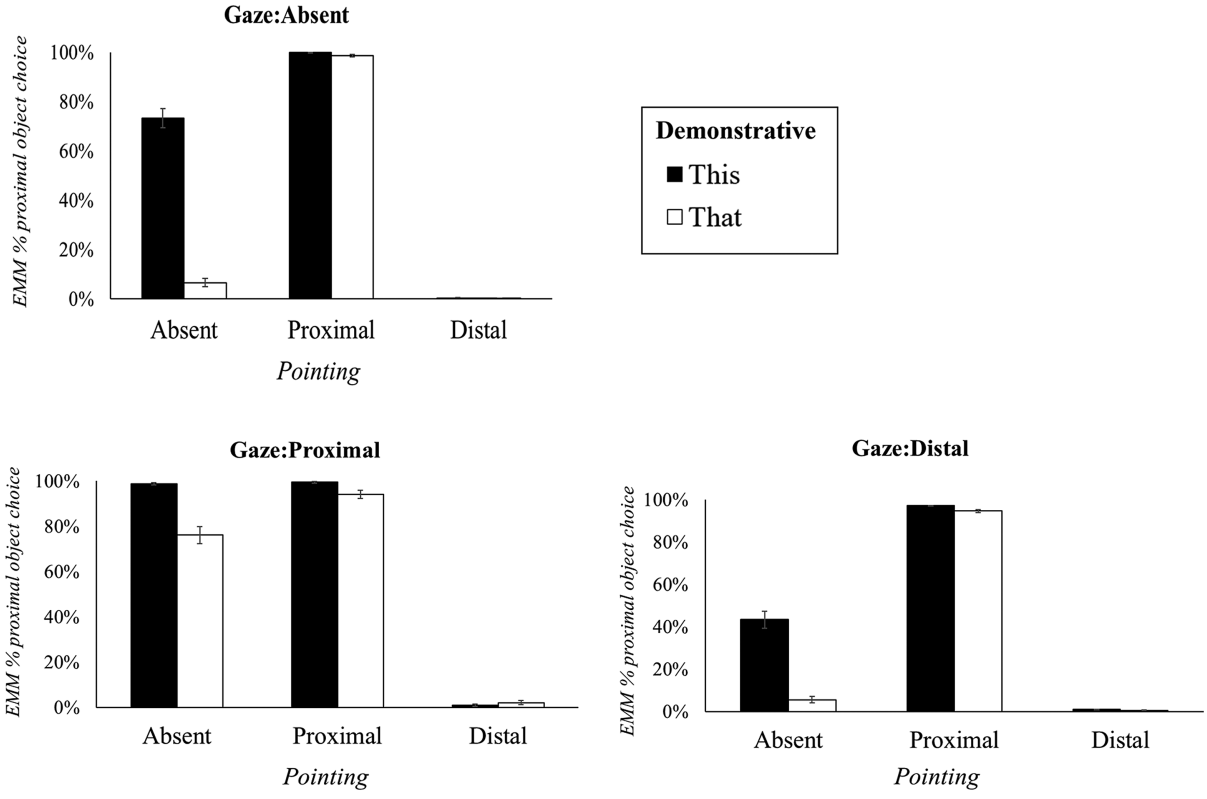
compared to either the single demonstrative condition, $OR = 1.12$, $SE = 0.29$, $z = .43$, $p = 1.00$, or the congruent demonstrative plus gaze condition, $OR = 1.52$, $SE = 0.33$, $z = 1.88$, $p = .894$. In other words, gaze only impacted referent choice when the demonstrative was distal.

For the dynamic stimuli, results also in the main replicated Experiment 1, with significant main effects as before of demonstrative, $\chi^2(1) = 27.42$, $p < .001$; pointing, $\chi^2(2) = 274.90$, $p < .001$; and gaze, $\chi^2(2) = 115.62$, $p < .001$, and interactions between demonstrative and pointing, $\chi^2(2) = 35.40$, $p < .001$, and between pointing and gaze, $\chi^2(4) = 124.98$, $p < .001$, all yielding the same patterns as seen in Experiment 1 (see Figures 4 and 5). However, there was also a significant three-way interaction between demonstrative, pointing, and gaze, $\chi^2(4) = 10.69$, $p = .03$ (Figure 8). In the absence of pointing, demonstrative and gaze affect referent choice for dynamic stimuli. For each level of gaze, *this* is associated with a greater choice of proximal referent compared to *that* (all contrasts $p <$ .001). Moreover, in the absence of pointing, congruent gaze and demonstrative together lead to a higher tendency to choose the targeted referent; proximal gaze and proximal demonstrative were associated with a greater selection of the proximal referent compared to the proximal demonstrative alone, $OR = 30.04$, $SE = 13.39$, $z = 7.63$, $p < .001$. When cues (gaze and demonstrative) are incongruent, proximal gaze was the dominant cue even when the linguistic cue (demonstrative) was distal ($M = .76$, $SE = 0.04$) compared to trials when both cues were congruently targeting distal position ($M = 0.06$, $SE = 0.01$), $OR = 54.10$, $SE = 16.20$, $z = 13.32$, $p < .001$. The three-way interaction also implied that, compared to trials without any nonverbal cue, having a combination of all three cues increased the strength of cueing to a target position (all contrasts $p < .001$; Figures 8 and 9). However, as long as the pointing arm was targeting the proximal item, participants would pick the proximal item, even when the gaze was distal and participants heard *that*, $OR = 1981.06$, $SE = 1395.69$, $z = 10.75$, $p < .001$ (compared to trials in which all cues target the distal position). The same trend is the case when pointing is distal; participants are cued by pointing even when gaze and demonstratives are proximal, $OR = 2000$, $SE = 0.0004$, $z = 12.87$, $p < .001$.

The individual analyses above reveal differences in patterns of interactions for static and dynamic stimuli that drive the two- and three-way interactions of stimuli type with pointing, demonstrative, and gaze in the full model. Overall, the most noticeable difference between static and dynamic cues was the increased influence of (distal) gaze on object referent choice. When the agent was not pointing, gaze toward the distal position was a stronger cue in dynamic trials than in static trials, $OR = 3.57$, $SE = 0.05$, $z = -7.22$ (Figure 5). There was also an effect of incongruent distal gaze on demonstratives heard. In dynamic trials, a combination of hearing

Figure 8

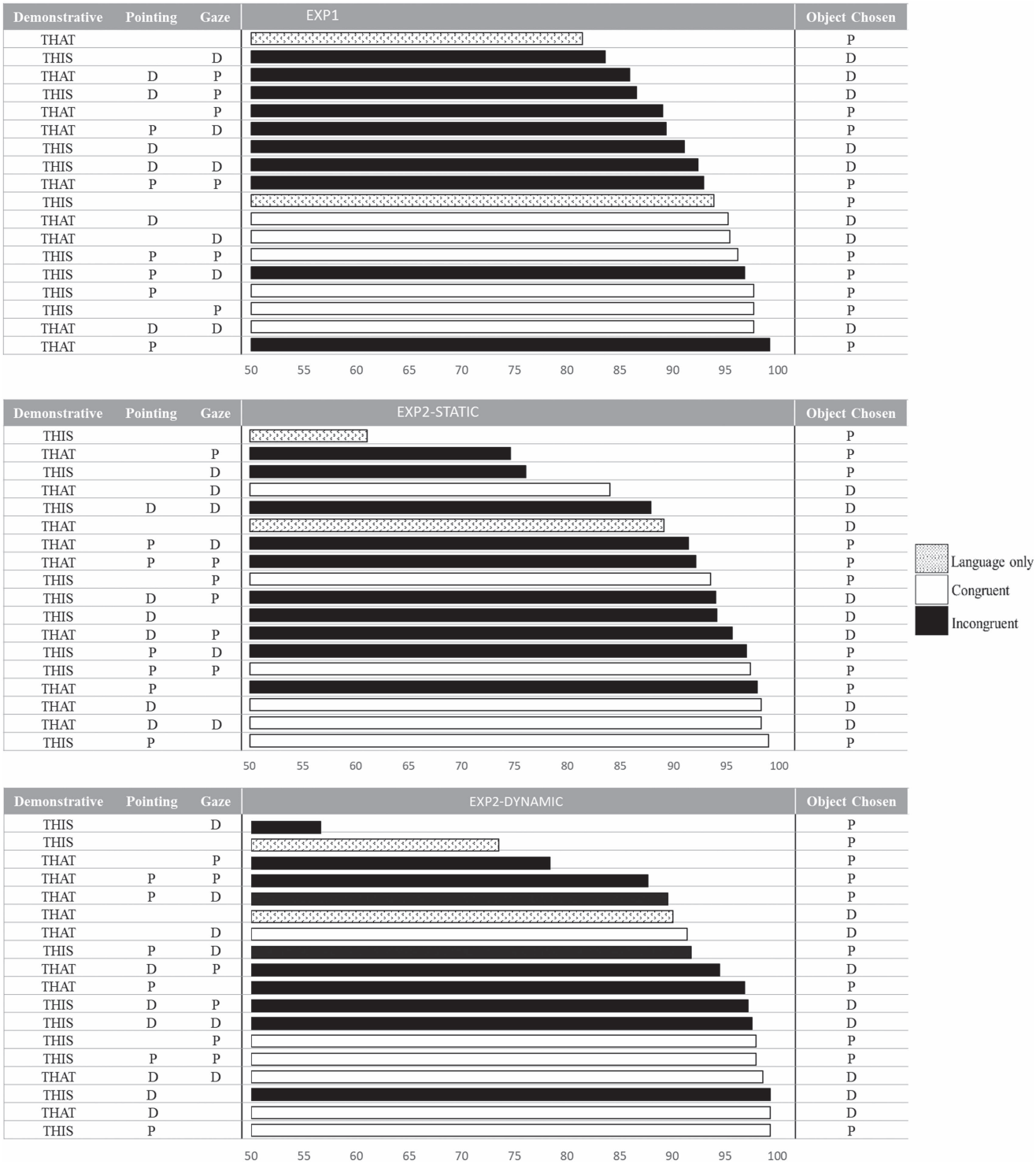
Bar Charts Illustrating the Three-Way Interaction Between Spatial Demonstratives, Pointing, and Gaze in Experiment 2 Dynamic Trials



Note. Values represent the estimated marginal means obtained from the statistical model. Error bars represent the standard error. EMM = estimated marginal means.

Figure 9

Ranked Tendency of Choosing a Targeted Reference for Each Level of Design Across the Two Experiments



Note. Experiment 2 is further divided into static and dynamic trials. At the end of each bar (on the right), it is indicated which one of the potential two referents, proximal or distal, participants were most likely to choose at a given level of the design. Data represent likelihood scores calculated from raw data (not from the estimated marginal means from the model). D = distal; P = proximal; EXP = experiment.

the word *this* with distal gaze led to a decreased likelihood of choosing the proximal item, $OR = 5$, $SE = 0.04$, $z = -7.35$ (Figure 8), compared to static trials. Overall, the effects of all cues are stronger in the dynamic compared to static trials. However, the data showed a similar pattern of effects for all three cues, both in Experiment 1 and Experiment 2 dynamic and static trials (Figures 4 and 5).

Discussion

Results of Experiment 2 from all three analyses are broadly consistent with those of Experiment 1 regarding the significant effect of the three types of deictic cues individually and patterns of two-way interactions of pointing with gaze and with heard demonstrative. In addition, however, there are some specific differences as a function of stimuli type.

Results again are consistent with the view that demonstratives map onto perceptual space with the demonstrative *this* indicating an object within reach/near the speaker and *that* indicating a farther referent (Anderson & Keenan, 1985; Caldano & Coventry, 2019; Coventry et al., 2008, 2014, 2023). Moreover, pointing and gaze were also associated with directing the participants to a particular referent, again consistent with the effectiveness of nonverbal cues as indicators of target position. Further replicating Experiment 1, pointing was the dominant cue, compared to both demonstratives and eye gaze. However, in Experiment 2, pointing and gaze together had an advantage over seeing one cue only, but only when targeting the distal position. This effect held even when data from static and dynamic stimuli were analyzed individually.

In the individual analyses separating static and dynamic cues, the pattern of interactions was overall consistent with the results of Experiment 1, but with two main differences. First, in the static trials, there was a significant interaction between gaze and demonstrative, indicating that gaze affected referent choice when the distal demonstrative was uttered, but not when the demonstrative was proximal. The presence of this interaction in this experiment (and the absence of it in Experiment 1) can potentially be explained by hearing the demonstratives rather than seeing them in a speech bubble, which is a more naturalistic way to perceive the verbal spatial reference. Second, in dynamic trials, there was a three-way interaction between pointing, eye gaze, and demonstrative heard. This three-way interaction suggests that, in the absence of pointing, eye gaze and demonstratives in combination provide a clearer cue to referent than hearing demonstratives alone. Furthermore, targeting with pointing, gaze, and demonstratives toward the same position leads to a stronger effect, as opposed to hearing spatial demonstratives alone.

Overall, the differences between stimuli types suggest that seeing the action of referring, which includes the trajectory of both eye gaze and pointing, and hearing spoken words uttered lead to an increased tendency to select a specific target referent when the cues are in combination (Figure 9). Static and dynamic trials mainly differed in terms of introducing the motion of referring as a factor. While the trends in choices based on the two types of stimuli were mostly mirrored, with the introduction of dynamic trials, there is an increased strength of all cues, and gaze and demonstratives in particular. It is of note that the gaze detection task produced differences in the accuracy of the gaze detection task for static and dynamic stimuli. People are simply better at detecting gaze as a motion rather than fixated statically at the target position, which offers a possible reason for the increased effectiveness of gaze in particular in the dynamic scenes.

General Discussion

In two studies, our goal was to unpack the strength and relative effectiveness of pointing gesture, eye gaze, and spatial demonstratives in guiding attention to a specific position in space by manipulating the number and congruency of cues participants were exposed to on a given trial. In two experiments, we tested the potential advantages of seeing multiple cues over single cues on reference choice and the relative strength of cues by pitting cues against each other (using incongruent cues). Furthermore, we examined cues in the context of static versus dynamic deictic stimuli.

Figure 9 plots the likelihood of choosing a specific referent, in terms of the overall percentage of trials where the same referent was selected. As can be seen, all cues and combinations of cues lead to above chance levels of selection of a specific referent (proximal or distal), but a number of general patterns emerge. First, as the results of the experiments demonstrate, and as one can see in Figure 9, demonstratives on their own, while above chance cues to a specific referent, fare relatively worse than cases where there are multiple congruent cues. While supporting the mapping between demonstratives and perceptual space (Anderson & Keenan, 1985; Caldano & Coventry, 2019; Coventry et al., 2008, 2014, 2023), it is certainly not the case that *this* is associated with participants always selecting the proximal referent and *that* with choice of distal referent in the absence of other cues. A likely reason for demonstratives not being the most reliable source for referent choice is the inherent vagueness and flexibility in the use of demonstrative forms. Demonstratives not only refer to the objects in space but also have other functions (Dixon, 2003). For example, it is well established that the use of demonstratives is affected by a range of factors, including the relative positions of speaker and hearer (e.g., Coventry et al., 2008; Jungbluth, 2003; Rubio-Fernandez, 2022) and the attention/gaze of the addressee (e.g., Küntay & Ozyürek, 2006; Rubio-Fernandez, 2022). Furthermore, the choice of demonstratives is influenced by a range of nonspatial factors such as object visibility, familiarity, and ownership (Chandralal, 2010; Coventry et al., 2014; Skilton & Peeters, 2021). Given the multiple parameters associated with these terms, it is understandable in our data why on their own they are not perfect; they serve to communicate more than mere object reference.

Second, in general, one can see that congruent multiple cues (the white bars in Figure 9) are associated with a higher tendency to choose a specific target referent compared to either demonstratives on their own (the hatched bars) or multiple incongruent cues (the black bars), consistent with previous findings from Bangerter and Louwerse (2005). The increased decisiveness of multiple congruent cues over single or fewer cues may help to explain why multiple cues seem to occur in deictic communication (Carpenter et al., 1998; Talmy, 2018), emerging synchronously relatively early in language development (Todisco et al., 2021).

Third, the pattern of results is also informative regarding the relative strengths of spatial deictic cues. While demonstratives are relatively vague cues in guiding attention, the dominance of

pointing over gaze and demonstratives was found consistently for both experiments and for both static and dynamic stimuli. In all cases when language and eye gaze were incongruent with pointing (see Figure 9), pointing most determined referent choice. While demonstratives are relatively vague pointers to a specific referent, the dominance of pointing over gaze and demonstratives was found consistently for both experiments and for both static and dynamic stimuli.

One can consider a number of possible reasons why pointing may be a stronger cue to a specific reference than gaze on the one hand and language on the other. With respect to the comparison between pointing and gaze, there are three possible reasons (not mutually exclusive) why pointing may be a stronger cue to reference than gaze. First, pointing may be a more precise and less ambiguous cue than gaze while also being more perceptually salient. Pointing may grab the most attention of viewers due to the perceptual salience of the action itself (compared to gaze), both in terms of the size of a gesturing arm and the availability of trajectory information (only Experiment 2). With static images, gaze occupies only a fraction of the image space compared to the pointing arm, making the pointing arm much easier to identify and process. Moreover, with dynamic scenes, again the movement of the arm occupies more points changing in location, with the pointing moving nearer toward the referent than the gaze, which remains bound to the head. So, there is less projection required on behalf of participants from the speaker to the referent with pointing compared to eye gaze. The trajectory (prestroke, stroke, poststroke hold and retreat) of the referential gesture overall is rich and elaborate in giving the listener cues for reference, making it a reliable and strong cue (Kendon, 1980; McNeill, 1992; Norris, 2011; Shattuck-Hufnagel et al., 2016). Overall, pointing improves the perceptual accuracy of locating a referent (Kowadlo et al., 2010). Cooney et al. (2018) have shown that people are extremely accurate at judging the intended location when a conspecific is pointing with arm outstretched and index finger extended. And pointing may indeed be a more precise cue than gaze (Butterworth & Itakura, 2000; Cooperrider et al., 2018), with listeners in turn focusing on the cue that offers the least ambiguous attentional nudge to a referent. Indeed, in our studies, we checked that participants could identify in the different gaze conditions which of the two potential referents on the table they thought the speaker was looking at. While performance was well above chance, it was nevertheless not perfect with static scenes, suggesting that gaze may be a less reliable visual cue than gesture (we did not include a gesture detection task as it seemed obvious which object the gesture was pointing toward in pilot work).

A second possibility is that pointing is a more overt communicative signal than gaze. As Cooperrider (2023) discussed, pointing is used when there is the intention of a speaker to direct the addressee's attention, but in contrast, gaze has a much wider use beyond signaling referent location, including signaling the amount of cognitive load involved in thinking at a given point in time and also signaling changes in the visual environment that may not be directly related to the communication at hand. Gesture is also executed slightly earlier than gaze, making gestural information available earlier and longer than other modalities. People tend to begin gesturing milliseconds before speech (Church et al., 2014) and hold the gesture after speech, making it available for longer than language. There is then an argument that gesture may grab the attention of the hearer/viewer first and hold it longest, giving more time for participants to locate the intended referent.

A third possibility relates to the nature of the tasks we used—the use of photographs and video images—where processing of gaze information in three-dimensional space is perhaps limited relative to gaze information in more real interactive situations. The stimuli used manipulated eye gaze in a very isolated manner, for example, restricted head movement and no head orientation information, which could have hindered the effect of eye gaze and decreased its salience. Previous research suggests head and gaze orientation and even eyebrow movement together with eye gaze influence the processing of an utterance (Hanna et al., 2020; Hömke et al., 2022; Langton et al., 2000). Future work would do well to examine cues in more realistic three-dimensional settings, for example, using virtual reality rather than still images/films, which can also be embedded in a more naturalistic interactive setting. Moreover, such studies might also benefit from eye-tracking participants to establish exactly how much time people spend processing different spatial deictic cues and if attention to cues is predictive of referent choice.

With respect to the relative strength of pointing and language, we can consider several possible explanations. First, pointing occurs early in development and may therefore take precedence on developmental grounds compared to demonstrative reference. Pointing is an important landmark/turning point in the development of language. There is a large body of research suggesting that deictic gestures (any type of gesture used when capturing and guiding attention) are a universally prelinguistic behavior closely tied to the emergence of the first words of children (Carpenter et al., 1998; Iverson & Goldin-Meadow, 2005). Children at around 9–12 months of age engage in gesture use while communicating (Liszkowski, 2010), for example, with their caregivers, before they start talking. Furthermore, the use of gesture precedes, predicts, and, overall, closely relates to (efficiency in) vocabulary learning and the emergence of language use (Carpenter et al., 1998; Iverson & Goldin-Meadow, 2005; Liszkowski, 2010). While it is unclear why pointing is universally present in infants prelinguistically, the "gesture-first" hypothesis maintains that gesture preceded language in the development of communication (e.g., Arbib, 2012; Corballis, 2010; Tomasello, 2008), and as such, the precedence of gesture developmentally could link to its precedence diachronically. Moreover, there is some evidence that children (at 2 years of age) already prioritize pointing over words (Grassmann & Tomasello, 2010), consistent with the "gesture-first" hypothesis.

Another reason why pointing may be more salient than language is that demonstratives serve a wide range of functions aside from merely communicating the location of a referent. As we discussed above, whether one uses *this* or *that* is affected by a wide range of parameters in addition to (relative) referent position (Coventry & Diessel, 2024; Coventry et al., 2014; Diessel & Coventry, 2020; Levinson et al., 2018; Peeters et al., 2021). Moreover, even as cues to the distance of a location, the fact that *that* as the "default" demonstrative in English can be used irrespective of the distance an object is located for the speaker (e.g., within peripersonal space; see Diessel & Monakhov, 2023; Levinson, 2018, for discussion) makes demonstratives a less reliable cue to a specific referent than more "overt" pointing (and gaze) cues.

Taken together, there are compelling reasons why pointing in our data is the most persuasive cue to spatial deixis, especially when cues are incongruent. However, it is also the case that, in general, multiple (congruent) cues lead to an increased tendency to choose a target referent compared to fewer cues. Although there is evidence for the co-occurrence of multiple cues in the early use of deictic communication (e.g., Todisco et al., 2021), and in some languages it is has been observed that it is obligatory to point or gaze when using specific spatial demonstratives, there is also recent evidence suggesting that pointing gestures do not always accompany demonstratives. For example, Bangerter (2004) found that gesturing was more likely to occur when a referent was in close proximity to a speaker (see also Cooperrider, 2016), but codemonstrative pointing was less frequent when using distal demonstratives to refer to far locations (see also Piwek et al., 2008). If pointing improves the localization of a referent, as was the case in our studies, one can ask why people do not always point when using demonstratives. One key difference between these studies and ours is that the further object in our studies was still not especially distal to the speaker, whereas Bangerter's (2004) studies involved referents that were much further away from participants. Cooperrider (2016) suggested that the use of pointing in combination with distal referents may be predicted by the extent to which pointing can unambiguously "pick out" the intended referent (Lücking et al., 2015). Accordingly, one can predict that rerunning our studies with much more distal references would correspondingly reduce the dominance of pointing as a cue to the speaker's intended referent.

The use of multiple deictic cues also provides a means for a speaker to communicate much more information about an object being referred to than mere object location. Indeed, a mismatch between deictic cues affords communication of information about the connection and attitude of a speaker toward the object referred to. For instance, the use of *that* when pointing and looking at a proximal object, while drawing the attention of the hearer to the object, may also provide information that the object is not liked by the speaker or that the object is not a main attentional focus for the speaker and may be less likely to occur in future reference. The multifaceted nature of spatial deictic communication places it at the junction between spatial communication and social communication, making it informationally rich and a natural bridge between spatial understanding and theory of mind (Coventry & Diessel, 2024; Rubio-Fernandez, 2022). Indeed, broadening out cues to reference, it has long been noted that "common ground" is also essential to establish reference (Clark, 1996). For example, Clark et al. (1983), in a series of real-world experiments, have shown that factors including the salience of the potential referent (kept constant in our studies) and the assumed knowledge that speaker and hearer have about each other and what they know in common both affect spatial reference choice. While our studies strictly controlled these other factors, in more real-life demonstrative reference settings, such factors are important; future studies would do well to manipulate spatial deictic cues while also manipulating other factors prominent in real-word communicative settings.

In summary, results from two experiments suggest that pointing is the most dominant cue as a determinant of referent choice and that demonstratives on their own are rather imperfect cues to object location. When pointing is incongruent with either gaze or demonstratives, pointing is the strongest predictor of choice of referent participants think the speaker is referring to. Proximal cues (including the word *this*) led to a higher likelihood of picking the proximal referent individually, and the effect is strengthened when all three cues are present. The effect of demonstratives is also strengthened with the addition of a nonverbal cue. Future studies would do well to examine online measures of attention to cues, for example, with the use of eye tracking, to examine the primacy of focus on nonlinguistic cues to observe the trajectory of visual scanning when perceiving deictic cues.

References

Anderson, S. R., & Keenan, E. L. (1985). Deixis. In T. Shopen (Ed.), *Language typology and syntactic description* (Vol. 3, pp. 259–308). Cambridge University Press.

Arbib, M. A. (2012). *How the brain got language: The mirror system hypothesis*. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199896684.001.0001

Arslan, B., Ng, F., Göksun, T., & Nozari, N. (2024). Trust my gesture or my word: How do listeners choose the information channel during communication? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *50*(4), 674–686. https://doi.org/10.1037/xlm0001253

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, *15*(6), 415–419. https://doi.org/10.1111/j.0956-7976.2004.00694.x

Bangerter, A., & Louwerse, M. M. (2005). Focusing attention with deictic gestures and linguistic expressions. *Proceedings of the Annual Meeting of the Cognitive Science Society* (pp. 1331–1336).

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, E. (1979). Cognition and communication from 9 to 13 months: Correlational findings. In E. Bates (Ed.), *The emergence of symbols: Cognition and communication in infancy* (pp. 69–140). Academic Press. https://doi.org/10.1016/B978-0-12-081540-1.50009-0

Bertenthal, B. I., Boyer, T. W., & Harding, S. (2014). When do infants begin to follow a point? *Developmental Psychology*, *50*(8), 2036–2048. https://doi.org/10.1037/a0037152

Bohnemeyer, J. (2018). Yucatec demonstratives in interaction: Spontaneous vs. elicited data. In S. C. Levinson, S. Cutfield, M. Dunn, N. Enfield, & S. Meira (Eds.), *Demonstratives in cross-linguistic perspective* (pp. 176–205). Cambridge University Press. https://doi.org/10.1017/9781108333818.009

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, *2*(1), 1–38. https://doi.org/10.5334/joc.72

Bühler, K. (1934). *Sprachtheorie* (Vol. 2). Fischer.

Butterworth, G., & Itakura, S. (2000). How the eyes, head and hand serve definitive reference. *British Journal of Developmental Psychology*, *18*(1), 25–50. https://doi.org/10.1348/026151000165553

Caldano, M., & Coventry, K. R. (2019). Spatial demonstratives and perceptual space: To reach or not to reach? *Cognition*, *191*, Article 103989. https://doi.org/10.1016/j.cognition.2019.06.001

Capirci, O., & Volterra, V. (2008). Gesture and speech: The emergence and development of a strong and changing partnership. *Gesture*, *8*(1), 22–44. https://doi.org/10.1075/gest.8.1.04cap

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative

competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, *63*(4), i–vi, 1–143. https://doi.org/10.2307/1166214

Chandralal, D. (2010). *Sinhala*. John Benjamins. https://doi.org/10.1075/loall.15

Chieffi, S., Secchi, C., & Gentilucci, M. (2009). Deictic word and gesture production: Their interaction. *Behavioural Brain Research*, *203*(2), 200–206. https://doi.org/10.1016/j.bbr.2009.05.003

Church, R. B., Kelly, S., & Holcombe, D. (2014). Temporal synchrony between speech, action and gesture during language production. *Language, Cognition and Neuroscience*, *29*(3), 345–354. https://doi.org/10.1080/01690965.2013.857783

Clark, H. H. (1996). *Using language*. Cambridge University Press. https://doi.org/10.1017/CBO9780511620539

Clark, H. H. (2003). Pointing and placing. In S. Kits (Ed.), *Pointing: Where language, culture and cognition meet* (pp. 251–276). Psychology Press.

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior*, *22*(2), 245–258. https://doi.org/10.1016/S0022-5371(83)90189-5

Cooney, S. M., Brady, N., & McKinney, A. (2018). Pointing perception is precise. *Cognition*, *177*, 226–233. https://doi.org/10.1016/j.cognition.2018.04.021

Cooperrider, K. (2016). The co-organization of demonstratives and pointing gestures. *Discourse Processes*, *53*(8), 632–656. https://doi.org/10.1080/0163853X.2015.1094280

Cooperrider, K. (2023). Fifteen ways of looking at a pointing gesture. *Public Journal of Semiotics*, *10*(2), 40–84. https://doi.org/10.37693/pjos.2023.10.25120

Cooperrider, K., Fenlon, J., Keane, J., Brentari, D., & Goldin-Meadow, S. (2021). How pointing is integrated into language: Evidence from speakers and signers. *Frontiers in Communication*, *6*, Article 567774. https://doi.org/10.3389/fcomm.2021.567774

Cooperrider, K., Slotta, J., & Núñez, R. (2018). The preference for pointing with the hand is not universal. *Cognitive Science*, *42*(4), 1375–1390. https://doi.org/10.1111/cogs.12585

Corballis, M. C. (2010). Mirror neurons and the evolution of language. *Brain and Language*, *112*(1), 25–35. https://doi.org/10.1016/j.bandl.2009.02.002

Coventry, K. R., & Diessel, H. (2024). Spatial communication systems and action. *Trends in Cognitive Sciences*. Advance online publication. https://doi.org/10.1016/j.tics.2024.10.002

Coventry, K. R., Griffiths, D., & Hamilton, C. J. (2014). Spatial demonstratives and perceptual space: Describing and remembering object location. *Cognitive Psychology*, *69*, 46–70. https://doi.org/10.1016/j.cogpsych.2013.12.001

Coventry, K. R., Gudde, H. B., Diessel, H., Collier, J., Guijarro-Fuentes, P., Vulchanova, M., Vulchanov, V., Todisco, E., Reile, M., Breunesse, M., Plado, H., Bohnemeyer, J., Bsili, R., Caldano, M., Dekova, R., Donelson, K., Forker, D., Park, Y., Pathak, L. S., … Incel, O. D. (2023). Spatial communication systems across languages reflect universal action constraints. *Nature Human Behaviour*, *7*(12), 2099–2110. https://doi.org/10.1038/s41562-023-01697-4

Coventry, K. R., Valdés, B., Castillo, A., & Guijarro-Fuentes, P. (2008). Language within your reach: Near-far perceptual space and spatial demonstratives. *Cognition*, *108*(3), 889–895. https://doi.org/10.1016/j.cognition.2008.06.010

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

Diessel, H. (2014). Demonstratives, frames of reference, and semantic universals of space. *Language and Linguistics Compass*, *8*(3), 116–132. https://doi.org/10.1111/lnc3.12066

Diessel, H., & Coventry, K. R. (2020). Demonstratives in spatial language and social interaction: An interdisciplinary review. *Frontiers in Psychology*, *11*, Article 555265. https://doi.org/10.3389/fpsyg.2020.555265

Diessel, H., & Monakhov, S. (2023). Acquisition of demonstratives in cross-linguistic perspective. *Journal of Child Language*, *50*(4), 922–953. https://doi.org/10.1017/S030500092200023X

Dixon, R. M. (2003). Demonstratives: A cross-linguistic typology. *Studies in Language*, *27*(1), 61–112. https://doi.org/10.1075/sl.27.1.04dix

Driver, J., IV, Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition*, *6*(5), 509–540. https://doi.org/10.1080/135062899394920

Edwards, S. G., Stephenson, L. J., Dalmaso, M., & Bayliss, A. P. (2015). Social orienting in gaze leading: A mechanism for shared attention. *Proceedings: Biological Sciences*, *282*(1812), Article 20151141. https://doi.org/10.1098/rspb.2015.1141

Eibl-Eibesfeldt, I. (1989). *Human ethology*. Aldine de Gruyter.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Forker, D. (2020). *A grammar of Sanzhi Dargwa* (Vol. 2). Language Science Press.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, *5*(3), 490–495. https://doi.org/10.3758/BF03208827

Galfano, G., Dalmaso, M., Marzoli, D., Pavan, G., Coricelli, C., & Castelli, L. (2012). Eye gaze cannot be ignored (but neither can arrows). *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *65*(10), 1895–1910. https://doi.org/10.1080/17470218.2012.663765

Gallucci, M. (2019). *GAMLj: General analyses for linear models* [jamovi module]. https://gamlj.github.io/

García, J. O. P., Ehlers, K. R., & Tylén, K. (2017). Bodily constraints contributing to multimodal referentiality in humans: The contribution of a de-pigmented sclera to proto-declaratives. *Language & Communication*, *54*, 73–81. https://doi.org/10.1016/j.langcom.2016.10.007

Grassmann, S., & Tomasello, M. (2010). Young children follow pointing over words in interpreting acts of reference. *Developmental Science*, *13*(1), 252–263. https://doi.org/10.1111/j.1467-7687.2009.00871.x

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*(4), 596–615. https://doi.org/10.1016/j.jml.2007.01.008

Hanna, J. E., Brennan, S. E., & Savietta, K. J. (2020). Eye gaze and head orientation cues in face-to-face referential communication. *Discourse Processes*, *57*(3), 201–223. https://doi.org/10.1080/0163853X.2019.1675467

Haxby, J. V., Gobbini, M. I., & Nastase, S. A. (2020). Naturalistic stimuli reveal a dominant role for agentic action in visual representation. *NeuroImage*, *216*, Article 116561. https://doi.org/10.1016/j.neuroimage.2020.116561

Hellwig, B. (2003). *The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria)* [Doctoral dissertation, Radboud University Nijmegen]. Max Planck Series in Psycholinguistics.

Herrmann, S. (2018). Warao demonstratives. In S. C. Levinson, S. Cutfield, M. Dunn, N. Enfield, & S. Meira (Eds.), *Demonstratives in cross-linguistic perspective* (pp. 282–302). Cambridge University Press. https://doi.org/10.1017/9781108333818.014

Hietanen, J. K. (1999). Does your gaze direction and head orientation shift my visual attention? *Neuroreport: For Rapid Communication of Neuroscience Research*, *10*(16), 3443–3447. https://doi.org/10.1097/00001756-199911080-00033

Holler, J., Schubotz, L., Kelly, S., Hagoort, P., Schuetze, M., & Özyürek, A. (2014). Social eye gaze modulates processing of speech and co-speech gesture. *Cognition*, *133*(3), 692–697. https://doi.org/10.1016/j.cognition.2014.08.008

Hömke, P., Levinson, S. C., & Holler, J. (2022). *Eyebrow movements as signals of communicative problems in human face-to-face interaction*. PsyArXiv. https://doi.org/10.31234/osf.io/3jnmt

Iverson, J. M., Capirci, O., & Caselli, M. C. (1994). From communication to language in two modalities. *Cognitive Development*, *9*(1), 23–43. https://doi.org/10.1016/0885-2014(94)90018-3

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, *16*(5), 367–371. https://doi.org/10.1111/j.0956-7976.2005.01542.x

Jungbluth, K. (2003). Deictics in the conversational dyad. In F. Lnz (Ed.), *Deictic conceptualisation of space, time and person* (pp. 13–40). John Benjamins. https://doi.org/10.1075/pbns.112.04jun

Kelly, S., Healey, M., Özyürek, A., & Holler, J. (2015). The processing of speech, gesture, and action during language comprehension. *Psychonomic Bulletin & Review*, *22*(2), 517–523. https://doi.org/10.3758/s13423-014-0681-7

Kelly, S. D., & Lee, A. L. (2012). When actions speak too much louder than words: Hand gestures disrupt word learning when phonetic demands are high. *Language and Cognitive Processes*, *27*(6), 793–807. https://doi.org/10.1080/01690965.2011.581125

Kelly, S. D., Ozyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, *21*(2), 260–267. https://doi.org/10.1177/0956797609357327

Kemmerer, D. (1999). "Near" and "far" in language and perception. *Cognition*, *73*(1), 35–63. https://doi.org/10.1016/S0010-0277(99)00040-2

Kendon, A. (1980). *The relationship of verbal and nonverbal communication*. De Gruyter Mouton.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press. https://doi.org/10.1017/CBO9780511807572

Kita, S. (Ed.). (2003). *Pointing: Where language, culture, and cognition meet*. Psychology Press. https://doi.org/10.4324/9781410607744

Kowadlo, G., Ye, P. J., & Zukerman, I. (2010). Influence of gestural salience on the interpretation of spoken requests. In T. Kobayashi, K. Hirose, & S. Nakamura (Eds.), *Proceedings of the 11th annual conference of the International Speech Communication Association* (pp. 2034–2037). Causal Productions.

Küntay, A. C., & Ozyürek, A. (2006). Learning to use demonstratives in conversation: What do language specific strategies in Turkish reveal? *Journal of Child Language*, *33*(2), 303–320. https://doi.org/10.1017/S0305000906007380

Langton, S. R. H., & Bruce, V. (1999). Reflexive visual orienting in response to the social attention of others. *Visual Cognition*, *6*(5), 541–567. https://doi.org/10.1080/135062899394939

Langton, S. R. H., & Bruce, V. (2000). You must see the point: Automatic processing of cues to the direction of social attention. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 747–757. https://doi.org/10.1037/0096-1523.26.2.747

Langton, S. R. H., Watt, R. J., & Bruce, I. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, *4*(2), 50–59. https://doi.org/10.1016/S1364-6613(99)01436-9

Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511813313

Levinson, S. C. (2018). Demonstratives: Patterns in diversity. In S. C. Levinson, S. Cutfield, M. Dunn, N. Enfield, & S. Meira (Eds.), *Demonstratives in cross-linguistic perspective* (pp. 222–241). Cambridge University Press. https://doi.org/10.1017/9781108333818.002

Levinson, S. C., Cutfield, S., Dunn, M. J., Enfield, N. J., & Meira, S. (Eds.). (2018). *Demonstratives in cross-linguistic perspective* (Vol. 14). Cambridge University Press. https://doi.org/10.1017/9781108333818

Liszkowski, U. (2010). Deictic and other gestures in infancy. *Acción Psicológica*, *7*(2), 21–33. https://www.redalyc.org/articulo.oa?id=344030764003

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & de Vos, C. (2012). A prelinguistic gestural universal of human communication. *Cognitive Science*, *36*(4), 698–713. https://doi.org/10.1111/j.1551-6709.2011.01228.x

Lücking, A., Pfeiffer, T., & Rieser, H. (2015). Pointing and reference reconsidered. *Journal of Pragmatics*, *77*, 56–79. https://doi.org/10.1016/j.pragma.2014.12.013

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

Meira, S. (2018). Tiriyó: Non-contrastive exophoric uses of demonstratives. In S. C. Levinson, S. Cutfield, M. Dunn, N. Enfield, & S. Meira (Eds.), *Demonstratives in cross-linguistic perspective* (pp. 222–241). Cambridge University Press.

Norris, S. (2011). Three hierarchical positions of deictic gesture in relation to spoken language: A multimodal interaction analysis. *Visual Communication*, *10*(2), 129–147. https://doi.org/10.1017/9781108333818.011

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

Perea García, J. O., Ehlers, K. R., & Tylén, K. (2017). Bodily constraints contributing to multimodal referentiality in humans: The contribution of a de-pigmented sclera to proto-declaratives. *Language & Communication*, *54*, 73–81. https://doi.org/10.1016/J.LANGCOM.2016.10.007

Phillips, A. T., Wellman, H. M., & Spelke, E. S. (2002). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition*, *85*(1), 53–78. https://doi.org/10.1016/S0010-0277(02)00073-2

Piwek, P., Beun, R. J., & Cremers, A. (2008). 'Proximal' and 'distal' in language and cognition: Evidence from deictic demonstratives in Dutch. *Journal of Pragmatics*, *40*(4), 694–718. https://doi.org/10.1016/j.pragma.2007.05.001

R Core Team. (2021). *R: A language and environment for statistical computing (R packages retrieved from MRAN snapshot)* (Version 4.0) [Computer software]. https://cran.r-project.org

Rubio-Fernandez, P. (2022). Demonstrative systems: From linguistic typology to social cognition. *Cognitive Psychology*, *139*, Article 101519. https://doi.org/10.1016/j.cogpsych.2022.101519

Senft, G. (2004). Aspects of spatial deixis in Kilivila. In G. Senft (Ed.), *Deixis and demonstratives in Oceanic languages* (pp. 59–80). Pacific Linguistics.

Shattuck-Hufnagel, S., Ren, A., Mathew, M., Yuen, I., & Demuth, K. (2016). Non-referential gestures in adult and child speech: Are they prosodic. In *Proceedings from the 8th international conference on speech prosody* (pp. 836–839). Boston University.

Skilton, A., & Peeters, D. (2021). Cross-linguistic differences in demonstrative systems: Comparing spatial and non-spatial influences on

demonstrative use in Ticuna and Dutch. *Journal of Pragmatics*, *180*, 248–265. https://doi.org/10.1016/j.pragma.2021.05.001

Stevens, J., & Zhang, Y. (2013). Relative distance and gaze in the use of entity-referring spatial demonstratives: An event-related potential study. *Journal of Neurolinguistics*, *26*(1), 31–45. https://doi.org/10.1016/j.jneuroling.2012.02.005

Talmy, L. (2018). *The targeting system of language*. MIT Press. https://doi.org/10.7551/mitpress/9780262036979.001.0001

The jamovi project. (2021). *jamovi* (Version 2.2) [Computer Software]. https://www.jamovi.org

Todisco, E., Guijarro-Fuentes, P., Collier, J., & Coventry, K. R. (2021). The temporal dynamics of deictic communication. *First Language*, *41*(2), 154–178. https://doi.org/10.1177/0142723720936789

Tomasello, M. (2008). *Origins of human communication*. MIT Press. https://doi.org/10.7551/mitpress/7551.001.0001

Welke, D., & Vessel, E. A. (2022). Naturalistic viewing conditions can increase task engagement and aesthetic preference but have only minimal impact on EEG quality. *NeuroImage*, *256*, Article 119218. https://doi.org/10.1016/j.neuroimage.2022.119218

Appendix

Full Analysis Results From Experiments 1 and 2

Table A1

Summary of Mixed Logit Model Comparing Levels of Variables via Helmert's Contrast in Experiment 1

	Parameter estimate				Wald's test	
Predictor (fixed effect)	Log-odds (β)	*SE*	Odds ratio	[95% CI]	*Z*	*p* (β = 0)
Intercept	0.61	0.12	1.83	[1.46, 2.30]	5.18	<.001
Demonstrative 1	1.16	0.36	3.19	[1.57, 6.47]	3.21	.00
Pointing 1	4.00	0.22	54.50	[35.21, 84.37]	17.93	<.001
Pointing 2	−3.50	0.22	0.03	[0.02, 0.05]	−15.77	<.001
Gaze 1	1.43	0.19	4.20	[2.87, 6.13]	7.42	<.001
Gaze 2	−2.24	0.23	0.11	[0.07, 0.17]	−9.58	<.001
Demonstrative 1 × Pointing 1	−0.75	0.45	0.47	[0.20, 1.13]	−1.68	.09
Demonstrative 1 × Pointing 2	−1.15	0.44	0.32	[0.13, 0.76]	−2.59	.01
Demonstrative 1 × Gaze 1	0.13	0.39	1.13	[0.53, 2.42]	0.33	.75
Demonstrative 1 × Gaze 2	0.38	0.47	1.47	[0.59, 3.67]	0.82	.41
Pointing 1 × Gaze 1	−2.60	0.43	0.07	[0.03, 0.17]	−6.05	<.001
Pointing 1 × Gaze 2	3.38	0.52	29.45	[10.56, 82.13]	6.46	<.001
Pointing 2 × Gaze 1	−1.69	0.45	0.18	[0.08, 0.45]	−3.76	<.001
Pointing 2 × Gaze 2	3.38	0.52	29.45	[10.56, 82.13]	6.46	<.001
Demonstrative 1 × Pointing 1 × Gaze 1	0.66	0.86	1.94	[0.36, 10.42]	0.77	.44
Demonstrative 1 × Pointing 2 × Gaze 1	−0.30	0.90	0.74	[0.13, 4.29]	−0.34	.74
Demonstrative 1 × Pointing 1 × Gaze 2	−0.10	1.06	0.90	[0.11, 7.17]	−0.10	.92
Demonstrative 1 × Pointing 2 × Gaze 2	0.03	1.05	1.03	[0.13, 7.99]	0.03	.98

Random component	*SD*	Variance	ICC
Participant (intercept)	0.10	0.61	0.05
Demonstrative	1.77	3.14	

Note. The first level in contrast is the contrast of the proximal target with the average of rest of the levels, distal and/or absent. For pointing and the gaze, second level of contrast looks at the contrast of distal target to no pointing or gaze. Number of observations 2,376, grouped by participants ($N = 33$). *SE* = standard error; CI = confidence interval; ICC = intraclass correlation.

(*Appendix continues*)

Table A2

Summary of Mixed Logit Model Comparing Levels of Variables via Helmert's Contrast in Experiment 2

Predictor (fixed effect)	Parameter estimate				Wald's test	
	Log-odds (β)	*SE*	Odds ratio	[95% CI]	*Z*	*p* (β = 0)
Intercept	−0.04	0.09	0.96	[0.80, 1.15]	−0.43	.67
Type 1	−0.48	0.12	0.62	[0.49, 0.78]	−4.07	<.001
Demonstrative 1	1.81	0.23	6.13	[3.90, 9.64]	7.84	<.001
Pointing 1	6.29	0.28	539.10	[309.97, 937.59]	22.28	<.001
Pointing 2	−4.57	0.26	0.01	[0.01, 0.02]	−17.6	<.001
Gaze 1	1.30	0.20	3.67	[2.48, 5.43]	6.51	<.001
Gaze 2	−0.23	0.19	0.79	[0.54, 1.16]	−1.21	.23
Type 1 × Demonstrative 1	−0.24	0.24	0.79	[0.50, 1.25]	−1.02	.31
Type 1 × Pointing 1	0.22	0.26	1.25	[0.74, 2.09]	0.84	.40
Type 1 × Pointing 2	−0.86	0.27	0.42	[0.25, 0.72]	−3.17	.00
Type 1 × Gaze 1	0.85	0.23	2.34	[1.49, 3.69]	3.68	<.001
Type 1 × Gaze 2	−0.62	0.31	0.54	[0.30, 0.98]	−2.01	.04
Demonstrative 1 × Pointing 1	−0.27	0.29	0.76	[0.43, 1.35]	−0.93	.35
Demonstrative 1 × Pointing 2	−1.93	0.31	0.15	[0.08, 0.27]	−6.29	<.001
Type 1 × Demonstrative 1 × Pointing 1	0.82	0.53	2.27	[0.81, 6.36]	1.56	.12
Type 1 × Demonstrative 1 × Pointing 2	−1.66	0.54	0.19	[0.07, 0.55]	−3.06	.00
Demonstrative 1 × Gaze 1	−0.37	0.24	0.69	[0.44, 1.10]	−1.55	.12
Demonstrative 1 × Gaze 2	−0.15	0.31	0.86	[0.47, 1.58]	−0.50	.62
Type 1 × Demonstrative 1 × Gaze 1	0.59	0.46	1.80	[0.73, 4.46]	1.27	.20
Type 1 × Demonstrative 1 × Gaze 2	−1.31	0.61	0.27	[0.08, 0.90]	−2.14	.03
Pointing 1 × Gaze 1	−2.36	0.29	0.09	[0.05, 0.17]	−8.11	<.001
Pointing 1 × Gaze 2	−1.52	0.37	0.22	[0.11, 0.45]	−4.12	<.001
Pointing 2 × Gaze 1	−2.33	0.31	0.10	[0.05, 0.18]	−7.53	<.001
Pointing 2 × Gaze 2	0.62	0.38	1.86	[0.88, 3.92]	1.63	.10
Type 1 × Pointing 1 × Gaze 1	−0.91	0.52	0.40	[0.15, 1.11]	−1.75	.08
Type 1 × Pointing 1 × Gaze 2	−0.43	0.69	0.65	[0.17, 2.51]	−0.62	.53
Type 1 × Pointing 2 × Gaze 1	−0.55	0.53	0.58	[0.20, 1.65]	−1.02	.31
Type 1 × Pointing 2 × Gaze 2	1.69	0.71	5.41	[1.35, 21.63]	2.39	.02
Demonstrative 1 × Pointing 1 × Gaze 1	1.02	0.55	2.78	[0.95, 8.14]	1.86	.06
Demonstrative 1 × Pointing 2 × Gaze 1	−0.49	0.57	0.61	[0.20, 1.85]	−0.87	.38
Demonstrative 1 × Pointing 1 × Gaze 2	−0.51	0.72	0.60	[0.15, 2.48]	−0.70	.48
Demonstrative 1 × Pointing 2 × Gaze 2	0.47	0.72	1.60	[0.39, 6.54]	0.65	.52
Type 1 × Demonstrative 1 × Pointing 1 × Gaze 1	0.65	1.03	1.91	[0.25, 14.41]	0.63	.53
Type 1 × Demonstrative 1 × Pointing 1 × Gaze 2	−0.45	1.37	0.64	[0.04, 9.35]	−0.33	.74
Type 1 × Demonstrative 1 × Pointing 2 × Gaze 1	−0.72	1.07	0.49	[0.06, 3.95]	−0.67	.50
Type 1 × Demonstrative 1 × Pointing 2 × Gaze 2	0.85	1.41	2.34	[0.15, 37.09]	0.60	.55

Random component	*SD*	Variance	ICC
Intercept	0.41	0.17	0.005
Demonstrative 1	1.53	2.34	
Pointing 1	1.77	3.14	
Pointing 2	1.43	2.03	
Gaze 1	1.17	1.37	
Gaze 2	0.76	0.58	

Note. For stimuli type, the contrast compares dynamic to static stimuli. The first level in contrast in demonstratives, pointing and gaze are the contrast of the proximal target with the average of the rest of the levels, distal and/or absent. For pointing and gaze, the second level of contrast looks at the contrast of distal target to no pointing or gaze. Number of observations 10,512, grouped by participants (N = 73). *SE* = standard error; CI = confidence interval; ICC = intraclass correlation.

(*Appendix continues*)

Table A3

Summary of Mixed Logit Model Comparing Levels of Variables via Helmert's Contrast in Experiment 2, Static Trials

Predictor (fixed effect)	Parameter estimate				Wald's test	
	Log-odds (β)	*SE*	Odds ratio	[95% CI]	*Z*	*p* (β = 0)
Intercept	0.23	0.11	1.26	[1.02, 1.57]	2.12	.03
Demonstrative 1	1.76	0.24	5.81	[3.64, 9.27]	7.39	<.001
Pointing 1	5.92	0.33	370.78	[195.59, 702.88]	18.13	<.001
Pointing 2	−3.78	0.25	0.02	[0.01, 0.04]	−15.05	<.001
Gaze 1	0.80	0.14	2.22	[1.67, 2.95]	5.53	<.001
Gaze 2	−0.09	0.19	0.92	[0.63, 1.33]	−0.46	.64
Demonstrative 1 × Pointing 1	−0.64	0.40	0.53	[0.24, 1.16]	−1.59	.11
Demonstrative 1 × Pointing 2	−1.10	0.35	0.33	[0.17, 0.66]	−3.17	.00
Demonstrative 1 × Gaze 1	−0.65	0.29	0.52	[0.30, 0.92]	−2.26	.02
Demonstrative 1 × Gaze 2	0.40	0.38	1.49	[0.71, 3.12]	1.05	.29
Pointing 1 × Gaze 1	−1.91	0.34	0.15	[0.08, 0.29]	−5.61	<.001
Pointing 1 × Gaze 2	−2.01	0.31	0.13	[0.07, 0.24]	−6.52	<.001
Pointing 2 × Gaze 1	−1.67	0.45	0.19	[0.08, 0.46]	−3.67	<.001
Pointing 2 × Gaze 2	−0.08	0.39	0.92	[0.43, 2.00]	−0.20	.84
Demonstrative 1 × Pointing 1 × Gaze 1	0.99	0.68	2.70	[0.71, 10.22]	1.46	.15
Demonstrative 1 × Pointing 2 × Gaze 1	−0.52	0.62	0.60	[0.18, 2.00]	−0.84	.40
Demonstrative 1 × Pointing 1 × Gaze 2	−0.64	0.91	0.53	[0.09, 3.12]	−0.70	.48
Demonstrative 1 × Pointing 2 × Gaze 2	0.49	0.79	1.64	[0.35, 7.67]	0.63	.53

Random component	*SD*	Variance	ICC
Participant (intercept)	0.354	0.125	0.04
Demonstrative	1.348	1.817	
Pointing 1	1.637	2.68	
Pointing 2	1.251	1.565	

Note. The first level in contrast is the contrast of proximal target with the average of rest of the levels, distal and/or absent. For pointing and gaze, second level of contrast looks at contrast distal target to no pointing or gaze. Number of observations 5,256, grouped by participants (N = 73). *SE* = standard error; CI = confidence interval; ICC = intraclass correlation.

(*Appendix continues*)

Table A4

Summary of Mixed Logit Model Comparing Levels of Variables via Helmert's Contrast in Experiment 2, Dynamic Trials

	Parameter estimate				Wald's test	
Predictor (fixed effect)	Log-odds (β)	*SE*	Odds ratio	[95% CI]	*Z*	*p* (β = 0)
Intercept	−0.26	0.15	0.77	[0.58, 1.02]	−1.81	.07
Demonstrative 1	1.68	0.32	5.39	[2.87, 10.13]	5.24	<.001
Pointing 1	6.65	0.41	772.03	[347.10, 1717.18]	16.30	<.001
Pointing 2	−5.23	0.41	0.01	[0.00, 0.01]	−12.66	<.001
Gaze 1	1.67	0.18	5.30	[3.71, 7.58]	9.15	<.001
Gaze 2	−0.65	0.24	0.52	[0.32, 0.84]	−2.69	.01
Demonstrative 1 × Pointing 1	0.09	0.44	1.10	[0.46, 2.60]	0.21	.83
Demonstrative 1 × Pointing 2	−3.01	0.53	0.05	[0.02, 0.14]	−5.66	<.001
Demonstrative 1 × Gaze 1	−0.04	0.36	0.96	[0.47, 1.95]	−0.12	.90
Demonstrative 1 × Gaze 2	−0.76	0.49	0.47	[0.18, 1.22]	−1.56	.12
Pointing 1 × Gaze 1	−2.90	0.39	0.06	[0.03, 0.12]	−7.44	<.001
Pointing 1 × Gaze 2	−2.68	0.44	0.07	[0.03, 0.16]	−6.06	<.001
Pointing 2 × Gaze 1	−2.18	0.52	0.11	[0.04, 0.31]	−4.19	<.001
Pointing 2 × Gaze 2	1.61	0.59	4.98	[1.58, 15.76]	2.74	.01
Demonstrative 1 × Pointing 1 × Gaze 1	1.60	0.78	4.93	[1.07, 22.63]	2.05	.04
Demonstrative 1 × Pointing 2 × Gaze 1	−1.43	0.88	0.24	[0.04, 1.35]	−1.62	.11
Demonstrative 1 × Pointing 1 × Gaze 2	−1.07	1.04	0.34	[0.04, 2.63]	−1.03	.30
Demonstrative 1 × Pointing 2 × Gaze 2	1.40	1.17	4.04	[0.40, 40.29]	1.19	.23

Random component	*SD*	Variance	ICC
Participant (intercept)	0.37	0.137	0.04
Demonstrative	1.87	3.50	
Pointing 1	2.22	4.91	
Pointing 2	1.68	2.81	

Note. The first level in contrast is the contrast of proximal target with the average of rest of the levels, distal and/or absent. For pointing and gaze, second level of contrast looks at contrast of distal target to no pointing or gaze. Number of observations 5,256, grouped by participants ($N = 73$). *SE* = standard error; CI = confidence interval; ICC = intraclass correlation.

Received June 12, 2023
Revision received December 7, 2024
Accepted December 17, 2024 ■