

# Journal of Experimental Psychology: Human Perception and Performance

## **Spatial (Mis-)Interpretation of Pointing Gestures to Distal Referents**

Oliver Herbort and Wilfried Kunde

Online First Publication, August 31, 2015. <http://dx.doi.org/10.1037/xhp0000126>

### CITATION

Herbort, O., & Kunde, W. (2015, August 31). Spatial (Mis-)Interpretation of Pointing Gestures to Distal Referents. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication. <http://dx.doi.org/10.1037/xhp0000126>

# Spatial (Mis-)Interpretation of Pointing Gestures to Distal Referents

Oliver Herbolt and Wilfried Kunde  
Julius-Maximilians-Universität Würzburg

Pointing gestures are a vital aspect of human communication. Nevertheless, observers consistently fail to determine the exact location to which another person points when that location lies in the distance. Here we explore the reasons for this misunderstanding. Humans usually point by extending the arm and finger. We show that observers interpret these gestures by nonlinear extrapolation of the pointer's arm-finger line. The nonlinearity can be adequately described as the Bayesian-optimal integration of a linear extrapolation of the arm-finger line and observers' prior assumptions about likely referent positions. Surprisingly, the spatial rule describing the interpretation of pointing gestures differed from the rules describing the production of these gestures. In the latter case, the eye, index finger, and referent were aligned. We show that the differences in the production and interpretation of pointing gestures accounts for the systematic spatial misunderstanding of pointing gestures to distant referents. No evidence was found for the hypotheses that action-related processes are involved in the perception of pointing gestures. How participants interpreted pointing gestures was independent of how they produce these gestures and whether they had practiced pointing movements before. By contrast, both the production and interpretation seem to be primarily determined by salient visual cues.

*Keywords:* pointing, gestures, gesture interpretation, gesture production, deictic reference

*Supplemental materials:* <http://dx.doi.org/10.1037/xhp0000126.supp>

Humans commonly point to guide their social partners' attention (Butterworth, 2003; Tomasello, Carpenter, & Liszkowski, 2007). Nevertheless, observers often fail to determine the exact distal location to which another person points (Bangerter & Oppenheimer, 2006; Butterworth & Itakura, 2000; Schmidt, 1999). As an example, consider the pointing person in the inset of Figure 1. Most observers judge that the person is pointing to the vicinity of Position C. However, she has been instructed and believes herself to point to Position A. In everyday life, similar misunderstandings occur frequently when a nonsalient object is identified exclusively by a pointing gesture. For example, it is difficult to communicate the location of a star in the night sky, a roll in a bakery's breadbasket, or an animal hidden in the wild solely by pointing. In such situations, lengthy verbal communication or other contextual cues are necessary to identify the location of interest. Because of these difficulties, humans point less frequently once other modes of communication develop (Pechmann & Deutsch, 1982). Likewise, humans rely more on verbal description when the accuracy requirements for pointing are high (Bangerter, 2004).

In this article, we address how an observer extracts the distal location implied by another person's pointing gestures and why

these gestures are often misinterpreted.<sup>1</sup> To address the accuracy of the spatial information conveyed by the gestures alone, we reduced the role of supplementary verbal or contextual information as much as possible. Of course, such information would be necessary to fully understand a gesture, which, for example, implies which object feature the pointer wants to highlight (cf. Wittgenstein, 1953) or how the pointer expects the observer to act on the object (Liebal, Behne, Carpenter, & Tomasello, 2009). With this focus on spatial interpretation of pointing gestures, we addressed the following three objectives.

As a first objective, we asked whether pointing is an extrapolation of a vector defined by the pointer's posture. Whereas some experiments suggested that this is unlikely (Butterworth & Itakura, 2000), others provided evidence in favor of this hypothesis (Bangerter & Oppenheimer, 2006; Wnuczko & Kennedy, 2011). However, at least one of two critical aspects makes the interpretation of previous studies difficult.

The first aspect pertains to the ambiguity of the single referent locations that were identified in previous experiments. In several studies, candidate referent locations were presented as a number line on the ground, or on a vertical or horizontal pole at a single distance from the pointer (Bangerter & Oppenheimer, 2006; Butterworth & Itakura, 2000; Wnuczko & Kennedy, 2011). Participants then judged the pointed-at location on that number line. Thus, only a single referent location was extracted for each point-

---

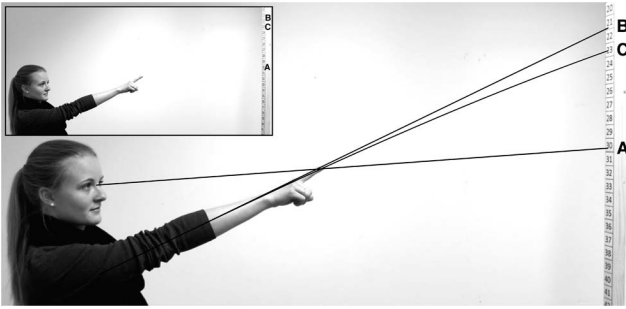
Oliver Herbolt and Wilfried Kunde, Department of Psychology, Julius-Maximilians-Universität Würzburg.

We thank Albrecht Sebald, Georg Schüssler, Hanna Mathew, and Felicitas Muth for assistance.

Correspondence concerning this article should be addressed to Oliver Herbolt, Department of Psychology, Julius-Maximilians-Universität Würzburg, Röntgenring 11, 97070 Würzburg, Germany. E-mail: [oliver.herbolt@psychologie.uni-wuerzburg.de](mailto:oliver.herbolt@psychologie.uni-wuerzburg.de)

---

<sup>1</sup> Although pointing gestures may have very different forms and functions, here we refer to "pointing gestures" exclusively as pointing gestures toward distal referents, which are commonly executed with an almost fully extended arm (Wnuczko & Kennedy, 2011, cf. the present Experiment 1). With "interpretation of pointing gestures," we refer to the extraction of the location the gesture is directed at.



**Figure 1.** Misunderstanding pointing gestures. The figure illustrates the linear extrapolation of the eye–finger line (A), the linear extrapolation of the arm–finger line (B), and an exemplar nonlinear, human extrapolation of the arm–finger line (C). Although most observers would think that the person in the inset is pointing to Position C, she was instructed (and believes herself) to point at Position A. The individual appearing here has consented for her likeness to be published in this article.

ing gesture. As a vector is defined by at least two points, this made it difficult to conclude whether the interpretation of pointing gestures corresponds to the extrapolation of a vector. Likewise, unless assumptions on the origin of the vector are introduced, the origin of the vector remains ambiguous. This ambiguity is further increased, because human attempts of vector extrapolation deviate systematically from geometric linear extrapolation (Salomon, 1947). For example, human vector extrapolation is often biased toward the horizontal or vertical axis (Bouma & Andriessen, 1968). As the magnitude of the bias is unknown a priori, multiple outcomes could be associated with the extrapolation of a single vector and vice versa. Figure 1 gives an example. The extrapolation of the arm–finger line could intersect the number line at different positions (B and C), depending on the assumed curvature of the extrapolation process. To resolve this ambiguity, we systematically varied the distance between the pointer and the area the pointing gesture was directed at. The multiple referent locations so identified for each pointing gesture allowed estimation of the nonlinearity of vector extrapolation, with the help of a computational model. Based on the estimation of the nonlinearity, participants' referent judgments could be associated with vectors defined by the pointing gesture.

The second aspect pertains to the fact that observers interpreted pointing gestures of other naive participants or confederates of the experimenter in previous studies (Bangerter & Oppenheimer, 2006; Butterworth & Itakura, 2000). As these gestures were neither controlled nor recorded (for an exception, see Wnuczko & Kennedy, 2011), it is difficult to relate the pointer's gesture to an observer's interpretation. Moreover, as the rules describing the production and interpretation of pointing gestures might differ, systematic biases might be introduced when pointing gestures of other humans were interpreted (Wnuczko & Kennedy, 2011). Here, we address this problem by presenting computer-generated images of pointers to participants. In sum, we improved the methodology of previous work by relying on well-defined pointing gestures and by accounting for the potential nonlinearity of human vector extrapolation.

As a second objective, we address why misunderstanding in the spatial interpretation of pointing gestures occur. It has been sug-

gested that differences in the rules that describe the production and interpretation of pointing result in such misunderstandings. Pointers seem to align the tip of their index finger with the referent in their visual field so that eye, index finger, and referent fall on a line (Line A in Figure 1; Bangerter & Oppenheimer, 2006; Taylor & McCloskey, 1988; Wnuczko & Kennedy, 2011). In contrast, it has been suggested that the interpretation of pointing gestures is based on the extrapolation of the pointer's arm and finger (Line B in Figure 1; Wnuczko & Kennedy, 2011). However, it is, at present, unknown whether these possible differences in the production and interpretation of pointing gestures can indeed account for the misinterpretation of pointing gestures. Here, we derive computational models for gesture interpretation and production, and test whether the models can account for the misunderstandings observable in the interaction of naive pointers and observers.

Third, according to the simulation theory of action understanding, observers rely at least partially on their own motor networks to understand the actions of others (Calvo-Merino, Glaser, Grèzes, Passingham, & Haggard, 2005; Fogassi et al., 2005; Gallese & Goldman, 1998; Gallese & Sinigaglia, 2011; Wolpert, Doya, & Kawato, 2003; for a critique, see Saxe, 2005). To test whether such mechanisms are involved in the understanding of pointing gestures as well, we assess whether gesture production and interpretation are related to each other on an actor-general and actor-specific level, and whether gesture production prior to gesture interpretation facilitates understanding. To our knowledge, this has not been previously examined in the context of pointing gestures.

In sum, our main hypotheses are as follows (cf. Figure 1). First, the interpretation of pointing gestures can be construed as the attempt to extrapolate a vector defined by the pointing gesture, most likely the vector defined by arm and index finger. Second, the extrapolation process is nonlinear (Figure 1, Line C). Third, pointing gestures are systematically misinterpreted because pointers align eye, index finger, and referent (Figure 1, Line A), but observers extrapolate the arm (Figure 1, Line C). Additionally, to test the possible involvement of motor processes in the interpretation of pointing gestures, we test whether an interaction between the production and interpretation of pointing gestures can be established.

The remainder of the article is structured as follows. Experiment 1 examines how pointing gestures to distal referents are spatially interpreted and produced, and how interpretation and production are related. Experiments 2a to 2d provide additional evidence for the conclusion that the spatial interpretation of pointing gestures is mainly based on the extrapolation of the pointer's arm and finger. Experiment 3 shows that the reported effects generalize to different observer perspectives. Experiment 4 reveals that the discrepancies between pointing gesture production and interpretation account for the misunderstandings of pointing gestures in a dyadic pointer–observer setting.

## Experiment 1

Experiment 1 was conducted to establish how participants interpret pointing gestures, how they produce these gestures, and whether gesture production and interpretation are related on an actor-specific or actor-general level. Participants were tested in two tasks. In a *gesture interpretation task*, participants estimated where on a vertical pole a computer-generated figure—with vari-

ous combinations of head and arm orientations, and at various distances from the pole—was pointing.<sup>2</sup> In a *gesture production task*, we recorded the postures participants assumed when pointing at various vertical positions at distances of 1 m, 2 m, or 3 m. We analyzed participants' behavior with two candidate models that relate the pointer's posture to the estimated (interpretation task) or instructed (production task) referent. The eye–finger rule posits that the eye, index finger, and referent form a line, and the arm–finger rule posits that the shoulder, index finger, and referent fall on a line.<sup>3</sup> As human attempts of line extrapolation systematically deviate from linear, geometric extrapolation (Bouma & Andriessen, 1968; Salomon, 1947), we used Bayesian models, which construe line extrapolation as the Bayesian optimal combination of a geometric extrapolation process, and the observer's prior assumption about likely referent positions. The Bayesian models are briefly outlined in the method section and formally derived in the Appendix. We relied on Bayesian models because of their frequent application to describe the integration of different information sources, for example, in contour extrapolation (Singh & Fulvio, 2005), multisensory integration (Deneve & Pouget, 2004; Knill & Pouget, 2004), or spatial judgments (Cheng, Shettleworth, Huttenlocher, & Rieser, 2007).

To test whether the interpretation of pointing gestures can be described as the attempt to extrapolate the arm–finger or eye–finger line, we fitted corresponding models to participants' judgments. Additionally, we tested whether the vertical position of referent estimates change linearly as a function of the horizontal distance between pointer and referent.

To test whether gesture interpretation and production are governed by different rules, we compared the fits of the eye–finger and arm–finger model in the gesture interpretation and production task. Finally, to test for a possible involvement of the motor system in the interpretation of pointing gestures, we test whether gesture production prior to gesture interpretation affects the position and variability of estimates, and whether gesture interpretation and production are related on an actor-specific level.

## Method

**Participants.** Sixty-four<sup>4</sup> students (56 women, eight men; mean age = 20 years) of the University of Würzburg participated. According to the Handedness scale and Eyedness scale of the Lateral Preference Inventory (LPI; Coren, 1993), 59 participants were right-handed, 5 were left-handed, 39 were right-eyed, 22 were left-eyed, and 3 had no eye preference. Half of the participants performed the gesture production task before the gesture interpretation task; for the other half, the order was reversed. Participants gave informed consent and received course credit.

**Design, procedure, and data analysis for interpretation task.** The stimuli were rendered using a 3D modeling software with a resolution of 1280 × 800 pixels. They showed the side view of a man who was pointing at a vertical pole in an otherwise empty room (see Figure S1 of the online supplemental materials for exemplary stimulus). The pole was located either 1 m, 2 m, or 3 m to the right of the man (1 m in the virtual scene corresponded to 191 pixels in the stimulus); arm orientation could be  $-20^\circ$ ,  $-10^\circ$ , . . . ,  $20^\circ$  (negative angles denote downward points); and head orientation could be  $-20^\circ$ ,  $-10^\circ$ , . . . ,  $20^\circ$  (negative angles denote downward head orientations). The stimuli were projected 1 m in

front of the participant. The size of the entire projection measured 112 cm × 71 cm. The visual angle of the projected stimuli corresponded to the visual angle that would result from actually seeing the scene from a distance of 6 m. Each trial began with the presentation of a blank beige screen for 500 ms, followed by an image of the pointer and the pole. Participants clicked with a mouse on the position on the pole, where they thought the figure was pointing. The mouse cursor was constrained to move along the pole. After 10 training trials, four blocks of 75 different stimuli (5 head orientations × 5 arm orientation × 3 distances) were presented. Stimulus order was randomized. In two blocks, the initial cursor position was at the foot of the pole; in the other two, it was at the top of the pole. Blocks were presented in random order. Altogether, 300 trials were administered, separated by self-paced breaks every 25 trials.

Screen coordinates of participants' estimates were converted into the coordinate system of the virtual scene to enable a better comparison with the production task. Except for 54 trials of one participant that were lost because of a computer failure, data from all trials were analyzed.

**Design, procedure, and data analysis for production task.** For the production task, a column of 64 squares (4 cm × 4 cm) was attached to vertical pole (286 cm height, 4.4 cm width). The squares were numbered from 1 (284 cm above the floor) to 64 (32 cm above the floor). Each trial began with the participant standing in front of the pole with the arms beside the trunk. The participant then pointed at the referent announced by the experimenter. Once the participant assumed a steady pointing posture, the experimenter pressed a key. A short beep followed after 500 ms, upon which the participant lowered the arm again.

Before the data collection, four practice trials were administered. Then, three blocks followed, in which the pole was placed at different distances from the participant (1 m, 2 m, and 3 m relative to the participants' trunk). The order of blocks was randomized. In each block, participants had to point at five different targets, each of which was presented 10 times. The targets were selected with respect to the participant's shoulder height (0.8 m, 0.4 m, 0 cm,  $-0.4$  m,  $-0.8$  m). The targets were presented in random order.

The movements of the participants were recorded using an electromagnetic motion tracking system (Ascension 3D Guidance trakSTAR; Ascension Technology Corporation, Shelburne, VT). Sensors were attached below the tip of the participants' right index finger, on the forearm, on the shoulder, and close to the participants' right eyes. From the sensor data, the position of the tip of the index finger, the wrist, the elbow, the shoulder joint, and the eyes were computed and smoothed with a second-order Butterworth filter with a cutoff frequency of 5 Hz.

For each trial, the position of the eye, index finger, wrist, elbow, and shoulder were extracted at the moment with the lowest tan-

<sup>2</sup> Head and arm orientation were varied orthogonally because it was unclear how pointers' head and arm orientation are related in natural pointing gestures comparable with those used in our task. However, the independent manipulation of both factors allowed us to test for independent effects of head orientation and arm orientation on gesture interpretation.

<sup>3</sup> Please note that upper arm, forearm, and finger are usually aligned when pointing at distant locations (Wnuczko & Kennedy, 2011, cf. the present Experiment 1).

<sup>4</sup> As the experiment was part of a course requirement, the sample size was determined by the number of students in the course.

gential index finger velocity in the period 500 ms prior to 300 ms after the experimenter's key press. The offsets of the dominant eye, index finger tip, and forearm sensor were measured before the experiment. If there was no eye preference, we used the average of both eye positions. The offset for the shoulder sensor from the shoulder joint was calculated as follows: For each trial, the offset values that minimized the position variance of the shoulder joint were computed, and the participant-wise median of these values was used as the shoulder offset.

Trials were excluded from the analysis if the pointing posture could not be extracted (one trial or 0.01%), if the participant had already lowered the arm when the experimenter pressed the key (two trials or 0.02%), or if the tangential velocity of the index finger at the time point of posture extraction was two standard deviations above a participants mean (308 trials or 3%).

**Model comparison for gesture production and interpretation task.** For the production task, the vertical position of the extrapolation of the eye–finger vector and the arm–finger vector (operationalized as vector from shoulder to finger) at the position of the pole was compared with the vertical position of the instructed referent. For the interpretation task, the estimated position of the referent was compared with the predicted position based on the nonlinear extrapolation of the arm–finger or eye–finger line. Bayesian models were used to describe humans' systematic deviations from geometric extrapolation (Bouma & Andriessen, 1968; Salomon, 1947). In the following, the Bayesian model of pointing gesture interpretation is briefly summarized. It is based on four assumptions. First, participants engage in geometric extrapolation of the arm–finger line or eye–finger line. Second, the reliability of the extrapolation decreases with distance (Pavel, Cunningham, & Stone, 1992; Salomon, 1947). Third, participants assume a priori that the referents are normally distributed, centered on shoulder height. The shoulder height was intentionally chosen because it defines the origin of the arm–finger line and because it corresponded to the mean of all referents estimates in an experimental session. The assumption of a normal-distributed prior at shoulder height was independently asserted in a control experiment reported in Text 1 of the online supplemental materials. Fourth, participants integrate the geometric extrapolation and a priori information optimally according to Bayesian theory (cf. Knill & Pouget, 2004; Körding & Wolpert, 2004). As the uncertainty associated with the geometric extrapolation increases with distance, referent estimates get increasingly biased toward the prior as distance increases. The Bayesian model can be expressed as follows (for derivation, see the Appendix):

$$\hat{y}_{\text{Bayesian}} = \frac{d^{-2}(1-w)y_{\text{geo}} + wy_0}{d^{-2}(1-w) + w}, \quad (1)$$

where,  $d$  is the horizontal distance between the pole and the pointer's shoulder,  $y_{\text{geo}}$  is the result of geometric extrapolation, and  $y_0$  is the a priori assumed average referent position, which is set to the shoulder height of the pointer. The Bayesian models have one free parameter  $w$ , which relates the variability associated with the linear extrapolations to the variability associated with the prior, both of which are unknown. The parameter  $w$  can assume values between 0 (participants rely exclusively on geometric extrapolation) and 1 (participants rely exclusively on the a priori assumption). The parameter  $w$  was determined individually for each

participant by minimizing the models' trial-wise computed root mean squared error (RMSE). Supplemental Table S1 provides the average value of  $w$  for each experiment reported herein. Additionally, the relative contribution of the prior is shown for the different pointing distances. This table also summarizes  $R^2$ s of the Bayesian model for all experiments as well as  $R^2$ s of purely geometric models for comparison, and shows a clear advantage for the Bayesian model for all gesture interpretation experiments reported in the following.

## Results

**Interpretation task.** Table 1 shows the fits of both models. The Bayesian arm–finger model provided a considerably better fit than the Bayesian eye–finger model,  $t(63) = 45.0$ ,  $p < .001$ ,  $g = 5.63$ . The high fit of the Bayesian arm–finger model suggests that the interpretation of pointing gestures can be adequately described as (nonlinear) extrapolation of the arm and finger.

Figure 2 shows how the arm and head orientation of a pointer affected the interpretation of the gestures. For analysis, the mean referent estimates were submitted to a repeated measures ANOVA with within-subject factors of Distance (1 m, 2 m, 3 m), Arm Orientation ( $-20^\circ$ ,  $-10^\circ$ , . . . ,  $20^\circ$ ), and Head Orientation ( $-20^\circ$ ,  $-10^\circ$ , . . . ,  $20^\circ$ ).<sup>5</sup> Distance affected referent estimates,  $F(2, 126) = 40.7$ ,  $p < .001$ ,  $\eta_p^2 = .393$ . Not surprisingly, arm orientation had a considerable effect on referent estimates,  $F(4, 252) = 2,972.0$ ,  $p < .001$ ,  $\eta_p^2 = .979$ . This effect was modulated by distance,  $F(8, 504) = 593.4$ ,  $p < .001$ ,  $\eta_p^2 = .904$ . By contrast, head orientation affected referent estimates only slightly,  $F(4, 252) = 2.6$ ,  $p = .072$ ,  $\eta_p^2 = .039$ . This effect was modulated by distance,  $F(8, 504) = 3.1$ ,  $p = .006$ ,  $\eta_p^2 = .047$ . There was no significant interaction between arm orientation and head orientation,  $F(16, 1008) = 0.9$ ,  $p = .547$ ,  $\eta_p^2 = .014$ , and no significant three-way interaction,  $F(32, 2016) = 0.8$ ,  $p = .643$ ,  $\eta_p^2 = .013$ .

Even though a significant interaction between distance and head orientation was found, the effect of head orientation was much smaller than the effect of arm orientation. A change in arm orientation of  $10^\circ$  resulted in an average shift of the estimated referent by 16.4 cm, 27.0 cm, and 35.3 cm at distances of 1, 2, and 3 m, respectively. A  $10^\circ$  change in head orientation shifted the estimate of the referent at the same distances by only .1 cm, .1 cm, and .6 cm.

Finally, we tested whether participants' gesture interpretations systematically deviated from linear extrapolations. Repeated-measures ANOVAs with the within-subject factor of Distance (1 m, 2 m, 3 m) were conducted for each arm orientation individually. For arm orientations of  $20^\circ$ ,  $10^\circ$ , and  $-20^\circ$ , contrast analyses revealed a significant quadratic term, all  $F$ s  $\geq 23.9$ , all  $p$ s  $\leq .001$ , all  $\eta_p^2$ s  $\geq .275$ . This shows that referent estimates did not change linearly as a function of distance. The nonlinearity is also reflected in the considerably better fit of the Bayesian model when compared with a model that assumes a geometric extrapolation process (Table S1 of the online supplemental materials).

**Production task.** When the participants were asked to point, they almost fully extended the arm and moved the index finger between the eyes and the referent (see Figure 3), corresponding

<sup>5</sup> We report Greenhouse-Geisser corrected  $p$  values, but uncorrected degrees of freedom, throughout the article.

Table 1  
Mean and SD of the Participant-Wise  $R^2$  by Model and Task

Task	Model	$R^2$	
		Mean	SD
Interpretation	Bayesian eye–finger model	.396	.108
	Bayesian arm–finger model	.910	.100
Production	Eye–finger model	.977	.038
	Arm–finger model	.549	.164

with earlier reports (Bangerter & Oppenheimer, 2006; Taylor & McCloskey, 1988; Wnuczko & Kennedy, 2011). Table 1 shows that, in contrast to the interpretation task, gesture production can thus be better described by the eye–finger model than the arm–finger model,  $t(63) = 18.9, p < .001, g = 2.36$ .

**Relation between interpretation and production.** We tested whether producing pointing gestures before interpretation affects interpretations and their consistency, as well as whether production and interpretation of gestures are correlated on an actor-specific level. First, if the interpretation of pointing gestures involved the motor system, the execution of pointing gestures might affect how, and how consistently, pointing gestures are subsequently interpreted. Hence, we tested whether the average and the standard deviation of referent estimates are affected by the prior production of pointing gestures.

The means and standard deviations of participants’ referent estimates were subjected to split-plot ANOVAs with within-subject factors of Distance, Arm Orientation, and Head Orientation, and the between-subjects factor Task Order (production before interpretation vs. interpretation before production). The ANOVA on the mean of the estimates revealed neither a main effect of task order,  $F(1, 62) = 1.7, p = .201, \eta_p^2 = .026$ , nor an interaction between task order and any of the other factors, all  $F_s \leq 1.9$ , all  $p_s \geq .162$ , all  $\eta_p^2_s \leq .031$ . Likewise, the ANOVA on the average standard deviations of the estimates revealed neither a

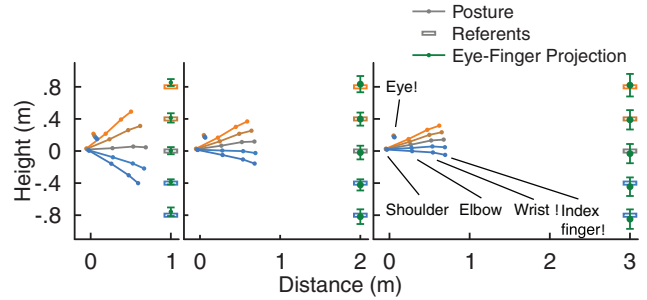


Figure 3. Gesture production. The figure shows the pointing gestures, referents (rectangles), and the extrapolation of the eye–finger line (green points near rectangles) in Experiment 1 as a function of referent position (from orange, highest [.8 m], to blue, lowest [–.8 m]) and distance from a side view. Associated postures and referents are drawn in the same color. Error bars show  $\pm 1$  standard deviation. See the online article for the color version of this figure.

main effect of task order,  $F(1, 62) = 2.5, p = .122, \eta_p^2 = .038$ , nor was any interaction between task order and any of the other factors significant, all  $F_s \leq 1.2$ , all  $p_s \geq .316$ , all  $\eta_p^2_s \leq .019$ . Thus, whether participants produced 150 pointing gestures or not, did not affect how and how consistently pointing gestures were interpreted.

Second, we tested whether participants who leaned toward one model in one task also leaned to the same model in the other task. Participants who leaned to the eye–finger rule as opposed to the arm–finger rule (quantified by the difference of the respective models root mean square errors) during gesture production did not necessarily exhibit the same tendency during interpretation,  $r = -.077, t(62) = -0.6, p = .545$ .

**Discussion**

Experiment 1 revealed several main findings. First, the production and interpretation of pointing gestures directed at distant referents can be described with different geometric rules. When participants point, they move the index finger of the almost fully extended arm between the eyes and the referent (cf. Bangerter & Oppenheimer, 2006; Taylor & McCloskey, 1988; Wnuczko & Kennedy, 2011). By contrast, when participants interpret these gestures, they attempt to extrapolate the vector defined by the pointer’s arm and finger. Moreover, referent estimates changed nonlinearly as a function of distance. The more distant the referent was, the stronger the referent estimate was biased toward a horizontal axis passing through the pointer’s shoulder. A Bayesian model, which construes the interpretation process as the integration of information from linear extrapolation and prior assumptions on likely referent positions, accounts for the nonlinear extrapolation process. Interestingly, head orientation plays at best a marginal role in the interpretations, even though also the gaze might indicate the position of the referent (Bock, Dicke, & Thier, 2008; Butterworth & Itakura, 2000). Finally, no relationship between the production and interpretation of pointing gestures could be established. Implications for theories of action understanding will be discussed in the General Discussion.

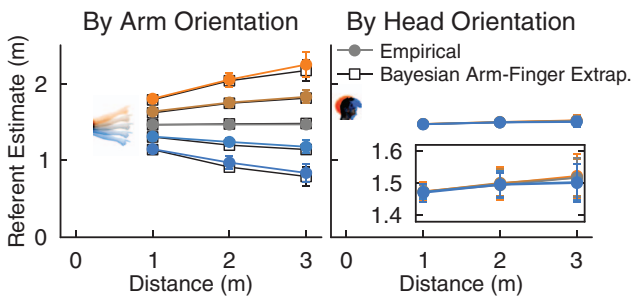


Figure 2. Effect of arm orientation, head orientation, and distance on referent estimates. The figure shows the estimated position of the referents by arm and head orientation (from orange, highest [20°], to blue, lowest [–20°]) and the fit of the Bayesian arm–finger model (left chart) in Experiment 1. The left charts shows data pooled over head orientations, the right chart shows data pooled over arm orientations. The inset shows a fivefold magnification. For clarity, only the arm or head oriented at 20° and –20° are shown, but complete figures were presented to participants. The data were scaled to the metrics of the virtual scenes shown to participants (1 m correspond to 191 pixels). Error bars show  $\pm 1$  standard deviation. See the online article for the color version of this figure.

## Experiment 2a

A number of control experiments follow-up on the conclusion that observers try to extrapolate the arm–finger line to interpret pointing gestures (exemplary stimuli used in Experiments 2a to 2d are provided in Figure S1 of the online supplemental materials). In the interpretation task of Experiment 1, the head orientation of the pointer was varied independent of arm orientation. Given the sometimes conflicting information provided by head and arm orientation, it is possible that participants increasingly relied on a single salient cue (Busemeyer, Myung, & McDaniel, 1993). This might have caused participants to ignore head orientation and might have biased the results toward the arm–finger model. To test whether the presentation of conflicting head and arm orientations in Experiment 1 affected referent estimates, the interpretation task of Experiment 1 was repeated with natural combinations of head and arm orientations.

## Method

Six<sup>6</sup> right-handed (according to the LPI; Coren, 1993) female students (mean age = 20 years) participated. Participants gave informed consent and received payment or course credit. The procedure was identical to the interpretation task of Experiment 1, with the exception that the head orientation of the pointing figure was adjusted to the arm orientation ( $-20^\circ$ ,  $-10^\circ$ , . . . ,  $20^\circ$ ) so that the pointer looked at his index finger. As head orientations were highly correlated with the angle between the eyes and the index finger (mean participant-wise  $r = .901$ ,  $SD = .086$ ) in the gestures produced in Experiment 1, we used a linear regression model to compute head orientations based on the angle between the pointing figure's eyes and the tip of its index finger. All trials were included in the analysis.

## Results and Discussion

The gesture interpretations in Experiment 2a can also be accurately described by the Bayesian arm–finger rule (mean participant-wise  $R^2 = .949$ ).  $R^2$ s did not differ significantly between Experiment 1 and 2a,  $t(68) = -0.9$ ,  $p = .347$ ,  $g = -0.404$ . Figure 4a shows that the interpretation of natural pointing gestures resembled those of Experiment 1. Referent estimates in both experiments were also numerically very similar. The average absolute difference between corresponding data points of Experiment 2a and Experiment 1 (resulting from averaging over head orientations for each arm orientation and distance) was 1.1 cm or 2 pixels. Thus, the independent variation of head and arm orientation in the interpretation task of Experiment 1 did not affect how participants interpreted the pointing gestures. As arm orientations were identical in both experiments but head orientations differed, Experiment 2a supports the finding of Experiment 1 that head orientation has little effect on the interpretation of pointing gestures.

## Experiment 2b

Experiments 1 and 2a suggest that participants rely on the arm to infer the pointers referent. If this was correct, referent estimates based on vision of the arm, but not the rest of the pointer, should resemble referent estimates based on vision of the entire pointer.

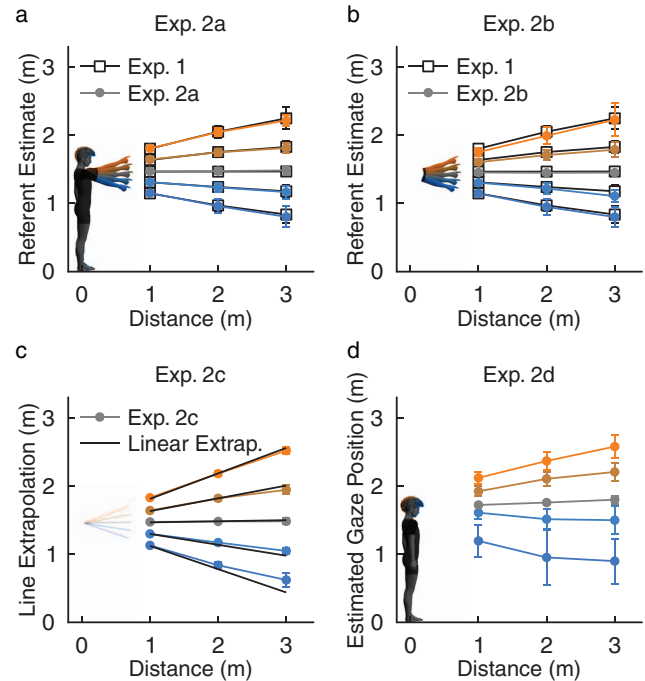


Figure 4. Results of Experiments 2a to 2d. Estimated position of referent of pointing gestures (a, b), line extrapolations (c), and gaze (d) by orientation and distance for Experiment 2a to 2d. The line colors refer to different arm (a, b), line (c), and head (d) orientations (from orange, highest  $[20^\circ]$ , to blue, lowest  $[-20^\circ]$ ). (a, b) The black lines show the responses of Experiment 1 for comparison. (c) The black lines show linear extrapolations of the stimulus line. The responses are scaled to the metrics of the virtual scene (1 m corresponds to 191 pixels). Error bars reflect  $\pm 1$  standard deviation. See the online article for the color version of this figure.

## Method

Six students (three women; mean age = 20 years; five right-handed, one left-handed, according to the LPI; Coren, 1993) participated. Participants gave informed consent and received payment or course credit. Experiment 2b was identical to Experiment 2a, with the exception that only the arm of the pointing figure was visible. Participants were asked to estimate where the arm was pointing. All trials were included in the analysis.

## Results and Discussion

If only the arm was shown, interpretations were also in line with the Bayesian arm–finger rule (mean participant-wise  $R^2 = .925$ ) and closely resembled those in which a full person was shown. Figure 4b shows that the results of Experiment 2b and Experiment 1 experiments were almost identical. The average absolute differences between corresponding data points in the current experiment

<sup>6</sup> Because of the highly consistent responses in the interpretation part of Experiment 1, only six participants were recruited for each of Experiments 2a to 2d. This sample size resulted in a power of  $\beta = .9996$  and  $\beta = .7850$  to detect an effect that correspond to 50% and 25%, respectively, of the effect size for the difference between the models in the interpretation task of Experiment 1.

and Experiment 1 and 2a were 3.0 cm (or 6 pixels) and 2.7 cm (or 5 pixels), respectively. As removing everything but the pointer's arm has little effect on participants' referent estimates, interpretations of pointing gestures seem to be mainly based on the pointer's arm.

### Experiment 2c

The Bayesian arm–finger model construes the interpretation of pointing gestures as the observer's attempt to linearly extrapolate the vector defined by the pointer's arm and finger. To test to which extent the interpretation of pointing gestures resembles line extrapolations, we asked participants to indicate the intersection of the extrapolation of a black line with the vertical pole in the stimuli. If the interpretation of pointing gestures could be construed as extrapolations of the arm–finger line, the Bayesian arm–finger model should also account for line extrapolations. Furthermore, nonlinear biases in the extrapolation could be expected. However, as the extrapolation of the line can be expected to be associated with less uncertainty than the extrapolation of the arm, line extrapolations might deviate less from linear extrapolation than the extrapolation of the arm–finger line.

### Method

Six right-handed participants (according to the LPI; Coren, 1993; five women, mean age = 20 years) gave informed consent and received payment or course credit. Experiment 2c was identical to Experiment 2a, with the exception that the pointer was replaced by a black line from the shoulder to the index finger. All trials were included in the analysis.

### Results and Discussion

The extrapolation of the line is captured by the Bayesian arm–finger rule (mean participant-wise,  $R^2 = .972$ , defining shoulder and index finger position by the line end points). Figure 4c shows that the instructed line extrapolations deviated from geometric linear extrapolation. This was statistically tested with repeated-measures ANOVAs with the within-subject factor Distance (1 m, 2 m, 3 m), which were conducted individually for each line orientation. A significant quadratic contrast was found for the  $-20^\circ$  line orientation,  $F(1, 5) = 10.7$ ,  $p = .022$ ,  $\eta_p^2 = .682$ , and a marginally significant quadratic contrast was found for the  $10^\circ$  line orientation,  $F(1, 5) = 5.6$ ,  $p = .02$ ,  $\eta_p^2 = .528$ . Thus, participants asked explicitly to extrapolate a line responded comparably with participants asked to judge the referents of pointing gestures. However, deviations from linearity were smaller in the current experiment than in the gesture extrapolation experiments. This is signified by a lower value of the model parameter  $w$  in the current experiment than in Experiments 1,  $t(68) = 3.6$ ,  $p = .001$ ,  $g = 1.530$ , and Experiment 2a,  $t(10) = 5.9$ ,  $p > .001$ ,  $g = 3.425$ . The lesser deviation from linearity is in line with the Bayesian framework, because the line most likely provided clearer directional information than the pointing figure.

### Experiment 2d

In Experiment 1, the pointer's head orientation played a negligible role in the interpretation of pointing gestures. This is sur-

prising because participants might have considered that pointers fixate the referent and might thus have also relied on head orientation to identify the referent (Todorović, 2006). Moreover, gaze following has been shown to be accurate in general (Bock et al., 2008; Butterworth & Itakura, 2000). In Experiment 2d, we asserted that the tiny effect of head orientation on participants' interpretation of pointing gestures did not result because participants were unable to differentiate the pointer's head orientations in our stimuli. Hence, the interpretation part of Experiment 1 was repeated, but participants were asked to indicate where on the vertical pole the pointer was looking. Only when participants were able to use the pointer's head orientation to infer gaze direction should participants' estimates depend on head direction.

### Method

Six right-handed (according to the LPI; Coren, 1993) students (five women, mean age = 23 years) participated. Participants gave informed consent and received payment or course credit. The procedure was identical to Experiment 2a, with the exception that the head orientations used in Experiment 1 were presented ( $-20^\circ$ ,  $-10^\circ$ , ...,  $20^\circ$ ). The arms of the figure were always beside the torso. Participants were asked to infer the position on the pole at which the person was looking. All trials were included in the analysis.

### Results and Discussion

Figure 4d shows that participants could differentiate the different head orientations. A repeated-measures ANOVA with within-subject factors of Head Orientation ( $-20^\circ$ ,  $-10^\circ$ ,  $0^\circ$ ,  $10^\circ$ ,  $20^\circ$ ) and Distance (1 m, 2 m, 3 m) revealed that head orientation strongly affected the estimated position of the gaze,  $F(4, 20) = 68.6$ ,  $p < .001$ ,  $\eta_p^2 = .932$ , and that this effect increased with distance  $F(8, 4) = 35.8$ ,  $p < .001$ ,  $\eta_p^2 = .878$ . Additionally, the gaze location was estimated to be higher for more distant positions,  $F(2, 10) = 10.9$ ,  $p = .014$ ,  $\eta_p^2 = .685$ . Thus, the negligible influence of head orientation in Experiment 1 did not result because participants could not differentiate between the different head orientations or could not extract directional cues from the head. Rather, gaze direction was not incorporated into the referent estimate.

### Experiment 3

Referent estimates depend on the observer's point of view (Bangerter & Oppenheimer, 2006). Experiment 3 tested whether the interpretation of pointing gestures could be described as extrapolation of the arm–finger line for perspectives other than the side view used so far. Participants were shown scenes of a pointer and a vertical pole comparable with those in Experiment 2a but from different perspectives. As explicitly instructed, linear extrapolation might also be view-dependent scenes in which the pointer was replaced by a rod with the same orientation as the pointer's arm were presented to a second group of participants. These participants were instructed to indicate where the extrapolation of the rod intersected with the pole. If participants extrapolated the arm–finger line also when seeing the pointer from other perspectives, the Bayesian arm–finger model should provide good fits independent of the observer's point of view. Additionally, the



interpretation of pointing gestures should be comparable with the extrapolation of a rod for different points of view.

## Method

**Participants.** Eleven women and one man from the Würzburg area (11 right-handed, one left-handed, according to the LPI; Coren, 1993; mean age = 28 years) participated for course credit or payment after giving informed consent. The sample size in each group was chosen to correspond to that of Experiments 2a to 2d.

**Stimuli and procedure.** For Experiment 3, the stimuli of Experiment 2a were adapted for 3D presentation with the anaglyph method and presented on CRT monitors (1280 × 1024 pixel, 75 Hz). Each scene was rendered from three points of views (side view, intermediate view, shoulder view; see Figure S1 of the online supplemental materials), which were selected to provide a good view on the pointer and the pole. The camera offset for the left-eye and right-eye image was 6.5 cm.

A trial began with the presentation of a beige screen for 500 ms, followed by a 3D scene. Participants used the mouse to position a cursor, which was perceived by the participant as horizontal line moving on the surface of the pole. For analysis, participants' responses were converted into metrics in the virtual world.

In the experiment, four factors were varied. Three different views were presented. Arm orientation could be 20° (up), 0°, or -20°. The (virtual) distance between pointer and pole could be 1 m, 2 m, or 3 m. One group of participants saw an image of a pointer and was instructed to estimate his referent (gesture interpretation task). Another group of participants only saw a black-and-white rod at the position of the pointer's arm and was asked to intersect the extrapolation of the rod with the pole (rod extrapolation task).

Before data collection, 10 training trials were administered. Then, eight blocks of 54 trials followed (two repetitions of each combination of three views, three arm orientations, and three distances). In half of the blocks, the initial cursor position was at the foot of the pole; in the other, it was at the top of the pole. Altogether, 432 trials (excluding training trials) were administered, separated by self-paced breaks between the blocks. All trials were included in the analysis. Block order, trial order, and assignment to the groups were random.

## Results and Discussion

Figure 5 shows how participants interpreted identical pointing gestures viewed from different perspectives. The Bayesian version of the arm-finger model provided a good fit for all views. In the gesture interpretation task, the average participant-wise  $R^2$ s were .987 for the side view, .989 for the intermediate view, and .947 for the shoulder view. In the rod extrapolation task, the average  $R^2$ s were .981 for the side view, .969 for the intermediate view, and .935 for the shoulder view.

To test for effects of the perspective and the task, a split-plot ANOVA with within-subject factors of View (side, intermediate, shoulder), Arm/Rod Orientation (-20°, 0°, 20°), and Distance (1 m, 2 m, 3 m), and a between-subjects factor of Task (gesture interpretation vs. rod extrapolation) was conducted. Estimates tended to be about 3 cm lower in the shoulder view than in the side view condition,  $F(2, 20) = 4.2, p = .055, \eta_p^2 = .297$ . A significant

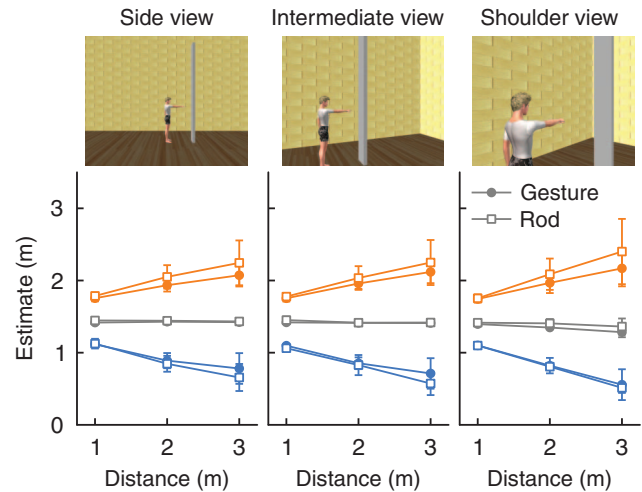


Figure 5. Gesture interpretation and rod extrapolations from different viewpoints. The figure shows the estimated position of the referent (gesture interpretation task) and the estimated intersection of the extrapolated rod with the pole (rod extrapolation task) for different views and distances in Experiment 3. The colors refer arm to orientations of 20° (orange, highest), 0° (gray, middle), -20° (blue, lowest). The images show the different perspectives used. Error bars reflect  $\pm 1$  standard deviation. See the online article for the color version of this figure.

effect of arm/rod orientation,  $F(2, 20) = 282.6, p < .001, \eta_p^2 = .966$ , and a significant interaction between distance and arm/rod orientation was found,  $F(4, 40) = 64.1, p < .001, \eta_p^2 = .865$ . Arm/rod orientation interacted with view,  $F(4, 40) = 28.4, p < .001, \eta_p^2 = .740$ . This effect was further modulated by the factor Distance, as signified by the interaction between arm/rod orientation, distance, and view  $F(8, 80) = 25.9, p < .001, \eta_p^2 = .721$ . Most importantly, neither the interaction including the factor Task nor the main effect of task reached significance, all  $F$ s  $\leq 2.4$ , all  $p$ s  $\geq .133$ , all  $\eta_p^2$ s  $\leq .193$ .

In sum, not surprisingly, the view slightly affected estimates. However, no significant differences between the two tasks could be found. For all views and both tasks, the Bayesian arm-finger model provides a close fit. This suggests that the interpretation of pointing gestures can be described as the observer's attempt to extrapolate the pointer's arm-finger line for a variety of perspectives.

## Experiment 4

In the previous sections, gesture production and interpretation were examined in isolation. Next, we tested whether the models identified for gesture production and interpretation capture the discrepancy of the referents pointed at by one participant (the pointer) and the estimate of that referent by another participant (the observer) in an in situ dyadic task. If pointers aligned the index finger with the target, but observers extrapolated the arm-finger line, observers' estimates of the pointer's referents should be systematically too high. Additionally, if the interpretation and production models accounted for the misunderstanding of pointing gestures, a combination of both models should be able to predict observers' referent estimates based on the referent provided to the pointer.

## Method

**Participants.** Six pointer–observer dyads (one woman; mean age = 20 years; 11 right-handed, one left-handed, according to the LPI; Coren, 1993) participated for course credit or payment after giving informed consent. The number of dyads was chosen to correspond to that of Experiments 2a to 2d and 3.

**Stimuli and procedure.** The pole with the numbered referents used in Experiment 1 was positioned in front of the pointer. The observer was seated 2 m to the right of the pointer, behind a small desk. A computer monitor between desk and pointer was used to display the referent number the pointer was supposed to point at. The observer entered her estimate in a laptop on the desk.

A trial began when a referent number from 27 to 52 (corresponding to heights between 180 cm and 80 cm from the floor, in steps of 4 cm) was displayed on the pointer’s screen. Then the pointer pointed at the referent. Once the observer entered an estimate, a beep sounded and the pointer was reminded on his screen to lower the arm. Two seconds later, the next trial began. The experiment consisted of five warm-up trials followed by three blocks of 26 trials each. In each block, the pointer had to point at all referents in pseudorandom order. The distance between pointer and pole differed between blocks (1 m, 2 m, and 3 m). The order of pointer-pole distances was counterbalanced over dyads and randomly assigned. The participants were assigned to the role of pointer and observer by toss of a coin. After the experiment, the shoulder height, arm length, and eye position relative to the shoulder of the pointer was measured. To keep the situation as natural as possible, we did not record the pointers movements with a motion tracker. The data of all trials (except warm-up trials) were included in the analysis.

## Results and Discussion

For analysis, we pooled the signed difference between the observers’ estimates and pointers’ referents over referent positions (80 cm to 180 cm) for each distance (1 m, 2 m, 3 m). Figure 6 shows that observers’ estimates were average to high (all  $t_s \geq 4.6$ ;  $p_s \leq .006$ , all  $g_s \geq 1.86$ ). A repeated-measures ANOVA with the factor Distance revealed that overestimations of the referents increased with distance,  $F(2, 10) = 105.9$ ,  $p < .001$ ,  $\eta_p^2 = .955$ , as

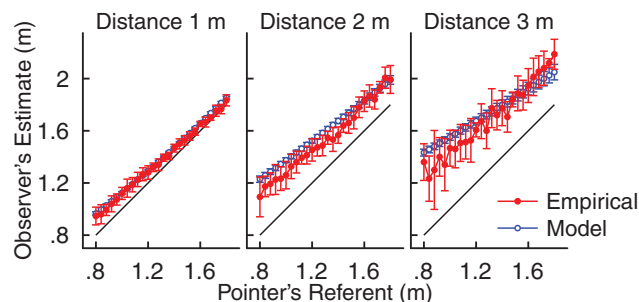


Figure 6. Observer’s referent estimates by pointer’s actual referent in Experiment 4. The figure shows the observers’ referent estimates (red) and model fit (blue) by the position the pointers were pointing at. Perfect judgments of the observer would fall on the black line. Error bars reflect  $\pm 1$  standard deviation. See the online article for the color version of this figure.

could be expected if the production of pointing gestures follows the eye–finger rule but interpretation was based on the arm–finger line.

To quantitatively assess to which extent the rules account for misinterpretation of pointing gestures, we modeled the observer’s responses with the models identified for gesture production and interpretation. Based on the eye–finger rule, the referent provided to the pointer, and the pointers body geometry, the postures during pointing were reconstructed. That is, pointing postures were not recorded but inferred from the eye–finger rule. The Bayesian arm–finger model was then used to predict the observer’s referent estimates from the reconstructed postures.<sup>7</sup> This reliably reproduced the observers’ referent estimates based on the pointers’ referents (mean participant-wise  $R^2 = .885$ ). Thus, the misunderstanding between pointers and observers in real-life dyadic situations can be attributed to the differences in gesture production and interpretation.

## General Discussion

When pointing at distal locations, our participants almost fully extended their arm, thereby moving the index finger between their eyes and the referent. The interpretation of these pointing gestures can be construed as the observer’s attempt to extrapolate a vector defined by the extended arm and finger. Similar to line extrapolation, the attempt to extrapolate the arm–finger line resulted in nonlinear pattern of referent estimates. The further the referents were away, the more the estimates were biased toward a horizontal axis passing through the pointer’s shoulder. The nonlinearity of the extrapolation process is well described as Bayesian-optimal integration of perceptual information and prior assumptions on likely referent positions. Finally, as pointers move the finger of the extended arm between their eyes and the referent, observers’ systematically judged the target of the pointing gesture as too high (see Video S1 of the online supplemental materials for another demonstration of this bias).

A Bayesian model was used to describe participants’ referent estimates. The use of the model was motivated by the finding that human vector extrapolations systematically deviated from linear extrapolations (Bouma & Andriessen, 1968). The present experiments suggest that the interpretation of pointing gestures is subject to such biases as well. This becomes evident by comparing the fit of the Bayesian model with purely geometric models (Table S1 of the online supplemental materials). Even though a purely geometric model can account for 67% of the variance of referent estimates (average over all participants that interpreted pointing gestures), the explained variance can be raised to 91% with the Bayesian extension of the geometric model. The advantage of the Bayesian approach becomes mostly evident when the distance between

<sup>7</sup> All computations were restricted to the pointer’s sagittal plane. The pointer’s shoulder’s y coordinate was set to the measured height; the x coordinate was set to the distance between pointer and pole (100 cm, 200 cm, or 300 cm). The coordinates of the eyes were assumed to be at a fixed position relative to the shoulder. First, the eye–finger rule was applied to reconstruct the pointer’s arm orientation, assuming that pointers fully extended their arms. Second, the Bayesian arm–finger model was used to predict the observer’s referent. The parameter  $w$  was fitted to actual referent estimates individually for each dyad, the value of  $y_0$  was set to the pointer’s shoulder height.

pointer and referent is large. When pointing gestures toward referents at the 3 m distance were interpreted, the relative weight of the prior was on average .35. By contrast, it was only about .06 when the distance was 1 m (Table S1 of the online supplemental materials). Finally, beside an improved fit, the Bayesian approach enables to capture several aspects of the data, such as the effect of the task/stimulus on the deviation from nonlinearity (cf. Text 1 of the online supplemental materials).

The presented models for gesture production and interpretation, which were exclusively based on the static position of eye, shoulder, and finger during pointing, accounted for the judgments of observers in natural dyadic pointing tasks. This shows that the static posture assumed during pointing conveys at least a large part of the information used to comprehend pointing gestures. In turn, this suggests that other potential sources of information—such as the dynamics of the gesture or the gaze direction of the pointer—play, at best, a minor role.

Our results constrain current theories on action understanding. It has been proposed that understanding motor actions of others involves corresponding own motor processes (Calvo-Merino et al., 2005; Fogassi et al., 2005; Gallese & Goldman, 1998; Gallese & Sinigaglia, 2011; Wolpert et al., 2003). However, we found no evidence that processes for pointing gesture production affect spatial aspects of gesture interpretation. First, we identified a general discrepancy between production and interpretation. Second, there was no relationship between interpretation and production of pointing gestures on an actor-specific level. Third, although motor experience has been shown to affect action perception in other domains (Casile & Giese, 2006), gesture interpretation was independent of whether it was preceded by the production of over 150 pointing gestures. This suggests that the spatial understanding of pointing gestures is not biased by processes involved in gesture production. Of course, the absent effects of a person's own pointing style or prior experience on the spatial interpretation of pointing gestures does not imply that the motor system is not involved at all. For example, the execution of pointing movements prior to interpretation might have had no effect, because pointing was highly overlearned in all participants. Moreover, it can be speculated that the motor system might be involved in aspects of the task other than the identification of the referent. For example, it might play a role in identifying another person's posture as a pointing gesture in the first place (Ping, Goldin-Meadow, & Beilock, 2014).

Nevertheless, the current experiments suggest that both gesture production and interpretation are mostly determined by the salient perceptual features, which differ for pointers and observers. Whereas pointers use their index finger akin to a cursor in their visual field, observers rely on the arm and finger as the most salient feature of the pointer's posture. This perceptual view on pointing is also supported by an experiment in which participants were pointing while they saw themselves in a mirror (i.e., from a third-person perspective; Wnuczko & Kennedy, 2011). In this case, pointers tended to align the arm–finger line with the referent, corresponding to the rules that usually describe the *interpretation* of pointing gestures. Likewise, arm orientation was found to be lower if pointers closed their eyes during pointing (Taylor & McCloskey, 1988; Wnuczko & Kennedy, 2011).

The present results may improve human–technology interaction, which increasingly relies on pointing gestures to control robots or computers (Breuer et al., 2012; Nickel & Stiefelhagen, 2007) or

guide user's attention by pointing humanoid companions (Noma, Zhao, & Badler, 2000). These systems need to incorporate different models for human pointing gesture production and interpretation to effectively interact with humans. On the one hand, gesture recognition systems need to be based on human models for gesture production. Likewise, embodied agents in virtual environments that aim for realism should be based on gesture production models (Rickel, 2001). On the other hand, pointing gestures that are aimed at easy understandability (e.g., in tutoring scenarios; Noma et al., 2000) should embrace models for human gesture interpretation to allow for the generation of effective gestures, even though they are unrealistic.

Finally, knowing how the rules underlying pointing and its interpretations deviate from each other can improve everyday communication. The recommendations derived from the present findings are simple: First, your guess of someone else's referent of pointing is most likely too high. Second, to improve communication by pointing, point a little bit lower than you would normally do.

## References

- Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, *15*, 415–419. <http://dx.doi.org/10.1111/j.0956-7976.2004.00694.x>
- Bangerter, A., & Oppenheimer, D. (2006). Accuracy in detecting referents of pointing gestures unaided by language. *Gesture*, *6*, 85–102. <http://dx.doi.org/10.1075/gest.6.1.05ban>
- Bock, S. W., Dicke, P., & Thier, P. (2008). How precise is gaze following in humans? *Vision Research*, *48*, 946–957. <http://dx.doi.org/10.1016/j.visres.2008.01.011>
- Bouma, H., & Andriessen, J. J. (1968). Perceived orientation of isolated line segments. *Vision Research*, *8*, 493–507. [http://dx.doi.org/10.1016/0042-6989\(68\)90091-6](http://dx.doi.org/10.1016/0042-6989(68)90091-6)
- Breuer, T., Giordana Macedo, G. R., Hartanto, R., Hochgeschwender, N., Holz, D., Hegger, F., . . . Kraetzschmar, G. K. (2012). Johnny: An autonomous service robot for domestic environments. *Journal of Intelligent & Robotic Systems*, *66*, 245–272. <http://dx.doi.org/10.1007/s10846-011-9608-y>
- Busemeyer, J. R., Myung, I. A., & McDaniel, M. A. (1993). Cue competition effects: Empirical tests of adaptive network learning models. *Psychological Science*, *4*, 190–195. <http://dx.doi.org/10.1111/j.1467-9280.1993.tb00486.x>
- Butterworth, G. (2003). Pointing is the royal road to language for babies. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 9–33). Mahwah, NJ: Erlbaum.
- Butterworth, G., & Itakura, S. (2000). How the eyes and hand serve definite reference. *British Journal of Developmental Psychology*, *18*, 25–50. <http://dx.doi.org/10.1348/026151000165553>
- Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex*, *15*, 1243–1249. <http://dx.doi.org/10.1093/cercor/bhi007>
- Casile, A., & Giese, M. A. (2006). Nonvisual motor training influences biological motion perception. *Current Biology*, *16*, 69–74. <http://dx.doi.org/10.1016/j.cub.2005.10.071>
- Cheng, K., Shettleworth, S. J., Huttenlocher, J., & Rieser, J. J. (2007). Bayesian integration of spatial information. *Psychological Bulletin*, *133*, 625–637. <http://dx.doi.org/10.1037/0033-2909.133.4.625>
- Coren, S. (1993). The lateral preference inventory for measurement of handedness, footedness, eyedness, and earedness: Norms for young adults. *Bulletin of the Psychonomic Society*, *31*, 1–3. <http://dx.doi.org/10.3758/BF03334122>

- Deneve, S., & Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links. *Journal of Physiology, Paris*, 98, 249–258. <http://dx.doi.org/10.1016/j.jphysparis.2004.03.011>
- Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science*, 308, 662–667. <http://dx.doi.org/10.1126/science.1106138>
- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2, 493–501. [http://dx.doi.org/10.1016/S1364-6613\(98\)01262-5](http://dx.doi.org/10.1016/S1364-6613(98)01262-5)
- Gallese, V., & Sinigaglia, C. (2011). What is so special about embodied simulation? *Trends in Cognitive Sciences*, 15, 512–519. <http://dx.doi.org/10.1016/j.tics.2011.09.003>
- Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27, 712–719. <http://dx.doi.org/10.1016/j.tins.2004.10.007>
- Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, 427, 244–247. <http://dx.doi.org/10.1038/nature02169>
- Liebal, K., Behne, T., Carpenter, M., & Tomasello, M. (2009). Infants use shared experience to interpret pointing gestures. *Developmental Science*, 12, 264–271. <http://dx.doi.org/10.1111/j.1467-7687.2008.00758.x>
- Nickel, K., & Stiefelwagen, R. (2007). Visual recognition of pointing gestures for human–robot interaction. *Image and Vision Computing*, 25, 1875–1884. <http://dx.doi.org/10.1016/j.imavis.2005.12.020>
- Noma, T., Zhao, L., & Badler, N. I. (2000). Design of a virtual human presenter. *IEEE Computer Graphics and Applications*, 20, 79–85. <http://dx.doi.org/10.1109/38.851755>
- Pavel, M., Cunningham, H., & Stone, V. (1992). Extrapolation of linear motion. *Vision Research*, 32, 2177–2186. [http://dx.doi.org/10.1016/0042-6989\(92\)90078-W](http://dx.doi.org/10.1016/0042-6989(92)90078-W)
- Pechmann, T., & Deutsch, W. (1982). The development of verbal and nonverbal devices for reference. *Journal of Experimental Child Psychology*, 34, 330–341. [http://dx.doi.org/10.1016/0022-0965\(82\)90050-9](http://dx.doi.org/10.1016/0022-0965(82)90050-9)
- Ping, R. M., Goldin-Meadow, S., & Beilock, S. L. (2014). Understanding gesture: Is the listener's motor system involved? *Journal of Experimental Psychology: General*, 143, 195–204. <http://dx.doi.org/10.1037/a0032246>
- Rickel, J. (2001). Intelligent virtual agents for education and training: Opportunities and challenges. In A. de Antonio, R. Aylett, & D. Ballin (Eds.), *Intelligent virtual agents* (pp. 15–22). Heidelberg, Germany: Springer. [http://dx.doi.org/10.1007/3-540-44812-8\\_2](http://dx.doi.org/10.1007/3-540-44812-8_2)
- Salomon, A. D. (1947). Visual field factors in the perception of direction. *The American Journal of Psychology*, 60, 68–88. <http://dx.doi.org/10.2307/1417328>
- Saxe, R. (2005). Against simulation: The argument from error. *Trends in Cognitive Sciences*, 9, 174–179. <http://dx.doi.org/10.1016/j.tics.2005.01.012>
- Schmidt, L. (1999). Adult understanding of spontaneous attention-directing events. *Ecological Psychology*, 11, 139–174. [http://dx.doi.org/10.1207/s15326969eco1102\\_2](http://dx.doi.org/10.1207/s15326969eco1102_2)
- Singh, M., & Fulvio, J. M. (2005). Visual extrapolation of contour geometry. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 102, 939–944. <http://dx.doi.org/10.1073/pnas.0408444102>
- Taylor, J. L., & McCloskey, D. I. (1988). Pointing. *Behavioural Brain Research*, 29, 1–5. [http://dx.doi.org/10.1016/0166-4328\(88\)90046-0](http://dx.doi.org/10.1016/0166-4328(88)90046-0)
- Todorović, D. (2006). Geometrical basis of perception of gaze direction. *Vision Research*, 46, 3549–3562. <http://dx.doi.org/10.1016/j.visres.2006.04.011>
- Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, 78, 705–722. <http://dx.doi.org/10.1111/j.1467-8624.2007.01025.x>
- Wittgenstein, L. (1953). *Philosophical investigations*. Oxford, UK: Blackwell.
- Wnuczko, M., & Kennedy, J. M. (2011). Pivots for pointing: Visually-monitored pointing has higher arm elevations than pointing blindfolded. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 1485–1491. <http://dx.doi.org/10.1037/a0024232>
- Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, 358, 593–602. <http://dx.doi.org/10.1098/rstb.2002.1238>

(Appendix follows)

## Appendix

### Derivation of the Bayesian Extrapolation Model

The following equation was used to model human attempts of linear extrapolation, which systematically deviate from geometric, linear extrapolation:

$$\hat{y}_{Bayesian} = \frac{d^{-2}(1-w)y_{geo} + wy_0}{d^{-2}(1-w) + w}. \quad (1)$$

In the following, Equation 1 is derived. Two information sources with Gaussian noise can be optimally fused by calculating a weighted mean, in which the values provided by each source are weighted with the inverse of the variance associated to the respective source (e.g., Knill & Pouget, 2004; Körding & Wolpert, 2004). In our case, geometric extrapolation and a priori assumptions about referent positions can thus be optimally fused as follows:

$$\hat{y}_{Bayesian} = \frac{\sigma_{geo,d}^{-2}y_{geo} + \sigma_0^{-2}y_0}{\sigma_{geo,d}^{-2} + \sigma_0^{-2}}, \quad (A.1)$$

where  $y_{geo}$  and  $\sigma_{geo,d}$  are the result of the geometric extrapolation and the standard deviation of the associated noise at distance  $d$ , respectively, and  $y_0$  and  $\sigma_0$  are the center and the standard deviation of the a priori assumption about the referent distribution, respectively. As the standard deviations of extrapolations are approximately proportional to the length of the extrapolation (Pavel et al., 1992; Salomon, 1947), we define the standard deviation associated with extrapolations by a distance  $d$  as

$$\sigma_{geo,d} = d \sigma_{geo,1m} \quad (A.2)$$

where  $\sigma_{geo,1m}$  denotes the standard deviations of the Gaussian noise associated with geometric extrapolations at a distance of 1 m. Inserting Equation A.2 in Equation A.1 results in

$$\hat{y}_{Bayesian} = \frac{(d \sigma_{geo,1m})^{-2}y_{geo} + \sigma_0^{-2}y_0}{(d \sigma_{geo,1m})^{-2} + \sigma_0^{-2}}. \quad (A.3)$$

By expanding, Equation A.3 can be rewritten as

$$\hat{y}_{Bayesian} = \frac{d^{-2} \frac{\sigma_{geo,1m}^{-2}}{\sigma_{geo,1m}^{-2} + \sigma_0^{-2}} y_{geo} + \frac{\sigma_0^{-2}}{\sigma_{geo,1m}^{-2} + \sigma_0^{-2}} y_0}{d^{-2} \frac{\sigma_{geo,1m}^{-2}}{\sigma_{geo,1m}^{-2} + \sigma_0^{-2}} + \frac{\sigma_0^{-2}}{\sigma_{geo,1m}^{-2} + \sigma_0^{-2}}}. \quad (A.4)$$

Only the relative size of  $\sigma_{geo}$  and  $\sigma_0$  are relevant. Thus, we take the following replacement in the fraction of Equation A.4, where  $w$  will later be estimated based on the empirical data:

$$w := \frac{\sigma_0^{-2}}{\sigma_{geo,1m}^{-2} + \sigma_0^{-2}} \quad (A.5)$$

With the replacement in Equation A.5, Equation A.4 can be rewritten as Equation 1:

$$\hat{y}_{Bayesian} = \frac{d^{-2}(1-w)y_{geo} + wy_0}{d^{-2}(1-w) + w} \quad (1)$$

Table S1 of the online supplemental materials provides the average value of  $w$  for each experiment reported in the present article.

Received April 17, 2015

Revision received July 14, 2015

Accepted July 16, 2015 ■