The Observer's Perspective Determines Which Cues Are Used When Interpreting Pointing Gestures

Lisa-Marie Krause & Oliver Herbort

Author note

Lisa-Marie Krause, Department of Psychology, Julius-Maximilians-Universität Würzburg, Germany;

Oliver Herbort, Department of Psychology, Julius-Maximilians-Universität Würzburg, Germany

Correspondence to: Lisa-Marie Krause, Department of Psychology, Julius-Maximilians-Universität Würzburg, Röntgenring 11, 97070 Würzburg, Germany, phone +49 931 3181719, email: lisa-marie.krause@uni-wuerzburg.de

Declarations of interest: none

Word count: 10688

Abstract

Though ubiquitous in human communication, pointing gestures are often misunderstood. This paper addresses how the observer's perspective affects pointing perception. More specifically, we test the hypothesis that two different visual cues – namely 1) the vector defined by the pointer's arm or finger and 2) the pointer's index finger position in the observer's visual field – determine pointing perception and that their relative influence depends on the observer's perspective. In three experiments, participants judged the location at which a virtual or real pointer was pointing from different viewpoints. The experiments show that the observer perspective has a considerable effect on pointing perception. The more the observer's gaze direction is aligned with the pointing arm, the more observers rely on the position of the pointing finger in their visual field and the less they rely on its direction.

Keywords

Pointing Perception; Virtual Reality; Cue-Weighting Hypothesis; Gesture Interpretation; Perspective

Public Significance Statement

Pointing gestures are an essential part of human communication, but observers often have difficulties to interpret the target correctly. In this study, we show that two different strategies for the interpretation are used, depending on the observer's perspective on the pointing gesture, and how this affects the perception. When standing close to the pointer, people rely on the same strategy as pointers do – using the pointer's index finger as an indicator for the target location, but when seeing the pointing arm from a sideward viewpoint it is mainly extrapolated.

The Observer's Perspective Determines Which Cues Are Used When Interpreting Pointing Gestures

We commonly point to direct the attention of others to remote locations. In many cases, we use pointing gestures together with verbal descriptions. But verbal descriptions are tedious in many cases, such as describing the position of a star in the night sky. In such situations, we often try to rely exclusively on pointing. However, the accuracy with which these pointing gestures are interpreted may be relatively low. Hence, the usefulness of pointing as a precise means of reference has been questioned (e.g. Butterworth & Itakura, 2000).

One factor that limits the accuracy of the communication with pointing gestures are systematic biases or offsets in pointing perception. Such systematic biases in the perception of pointing gestures to distal referents, which occur when language is not available, have been investigated in a number of experiments. For example, observers typically overestimate the height of a pointed-at location (Bangerter & Oppenheimer, 2006; Herbort & Kunde, 2016; Herbort & Kunde, 2018; Mayer et al., 2020; c.f. Sousa et al., 2019). A comparable bias was reported when a pointer stood on a path of objects on the floor (Wnuczko & Kennedy, 2011). Even though the spacing between objects was 1 m, observers consistently judged the referent to be further away from the pointer than it actually was.

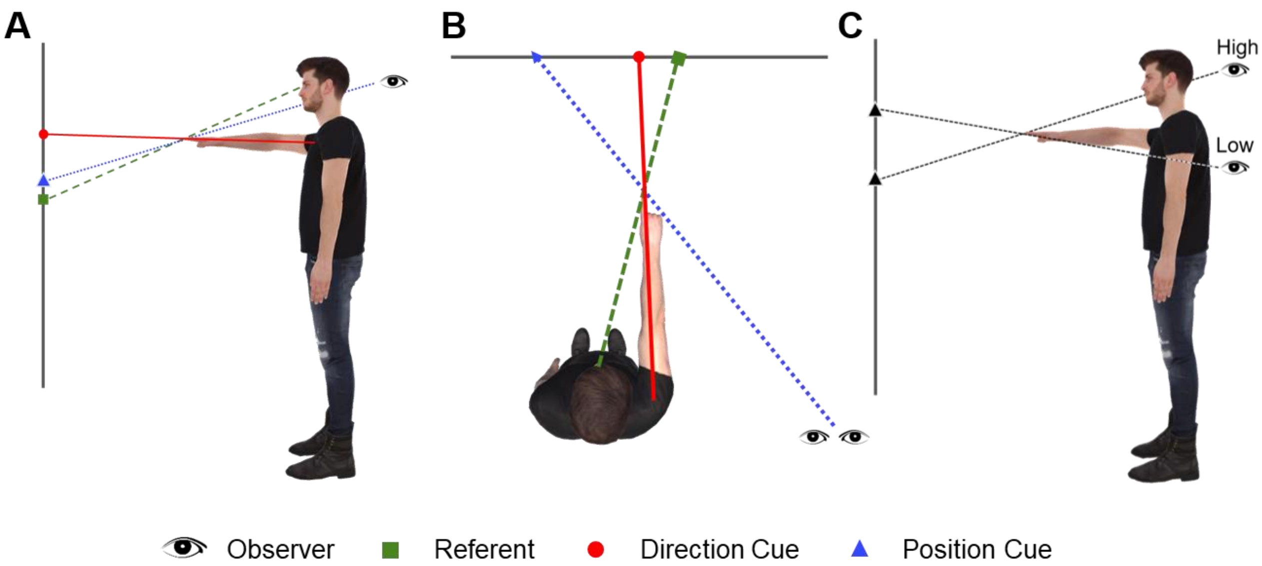
Experiments that focussed on the horizontal component of pointing accuracy yielded mixed results. In a study in which a pointer pointed with her right arm to horizontally arranged objects, observers sitting next to the pointer exhibited a leftward bias (Butterworth & Itakura, 2000). In a comparable setup, Bangerter and Oppenheimer (2006) asked a pointer to indicate numbers on a horizontal number line about 1.5 m in front of her. Observers show biases either to the left or right, depending on whether they sat to the right or left of a pointer (who used the right or left arm, respectively). These biases were larger for contralateral than ipsilateral points, regardless of the used arm. By contrast, Cooney et al. (2018) reported almost perfect accuracy, except for contralateral points, in an experiment, in which a pointer pointed at objects aligned horizontally between pointer and observer.

One explanation for pointer-observer misunderstandings is the usage of two different cues for pointing production and pointing perception: Pointers move their index finger close to the referent in their visual field. Hence, the vector from the pointer's eyes to the index finger points at the actual referent (green, dashed line in Fig. 1ab). However, observers typically try to extend the vector defined by the pointer's arm or index finger (red line in Fig. 1ab), thus misinterpreting the gesture (Herbort & Kunde, 2016; Wnuczko & Kennedy, 2011). While this hypothesis is clearly supported when the height of a referent needed to be deduced from a pointing gesture seen from the side, the model obviously needs elaboration. First, it is not consistent with all the data reported so far (Cooney et al., 2018). Second, it does not account for the finding that pointing gestures may be interpreted differently when seen from different viewpoints as indicated for systematic errors (Herbort et al., 2020; Herbort & Kunde, 2016; Mayer et al., 2020). However, neither a model of the systematic errors of pointing perception nor a detailed explanation of what these distortions are based on could be provided so far that integrates previous findings.

In this article, we aim to elaborate on the above hypothesis. We suggest that pointer-observer misunderstandings can be attributed to the use of the two visual cues and to the different weights pointer and observer put on these two cues in order to infer the pointed-at referent (*cue-weighting hypothesis*), comparable to cue integration of multisensory information in order to derive a robust percept (Ernst & Bülthoff, 2004). Specifically, we suggest that pointers point by bringing the pointing finger on the line between their eyes and the referent (dashed green line, Fig. 1ab). Hence, the vector from the pointer's eyes to the index finger is directed at the actual referent. By contrast, we assume that observers mainly use the extrapolation of the pointing arm (*arm direction cue,* short: direction cue; red line, Fig 1ab) when seeing the pointer from the side (Herbort & Kunde, 2016). Importantly, we assume that the strategy of the observer approaches that of the pointer, the closer the observer gets to the pointer. That is, observers rely the more on the extrapolation of the vector defined by the *observers'* eyes and the *pointer's* fingertip (*fingertip position cue*, short: position cue; dotted blue line in Fig. 1ab) the closer they get to the pointer. Note that the position cue almost always implies different referent locations for pointers and observers because of their typically different eye positions that mark the starting point of the vector of the position cue. However, the vector's second point is always the *pointer's* fingertip. Thus, whereas pointing and interpretation based on the position cue shares the general relationship between eye and pointer's finger, they typically imply different referent locations. Figure 1ab illustrates how this hypothesis accounts for the consistently found upward biases and for leftward biases for right-handed points.

Figure 1

Pictorial Representation of the Fingertip Position Cue and Arm Direction Cue



Note. **A, B.** The cartoons depict a person pointing at a green square on a wall by aligning eye, finger, and referent (dashed green line/square). Observers who extrapolate the arm-finger line (arm direction cue) perceive the referents higher (A) and more leftward (B) than it actually is (red line/dot). Biases of observers who use the position of the pointer's finger in their visual field (finger position cue, dotted blue line/triangle) depend on the observer's viewpoint. **C**. The effect of the observer's view height on the position cue.

Two considerations support the hypothesis. The first one is based on the premise that pointers produce gestures that they themselves would perceive as indicating the referent. This assumption finds support by the kinematics of pointing being strongly affected by perceptual variables (e.g. availability and type of visual information, Herbort & Kunde, 2018; Wnuczko & Kennedy, 2011; Wnuczko et al., 2013). If pointing production, which relies on the fingertip position cue, would be thus controlled by the pointer's perception of his own gesture, it would seem likely that an observer assuming the viewpoint of the pointer would also rely on the position cue – reporting a target location that is optically adjacent to the pointer's fingertip in the observer's visual field. As observers who see the pointer from the side rely on the arm direction cue (Herbort & Kunde, 2016; Wnuczko & Kennedy, 2011), this would imply that both cues are used by observers – depending on their perspective.

An additional consideration that supports the hypothesis pertains to the relative salience of both cues. When the gesture is seen from the side, the straight and clearly visible line from shoulder to fingertip suggests an extrapolation. By contrast, the fingertip position cue is typically implausible when seeing the pointer from the side. However, the further the observer moves away from that side viewpoint while coming closer to the pointer and aligning her view with the pointing arm, the more difficult the arm direction can be derived due to its perspective distortion, which optically foreshortens the arm. At the same time, the position cue gets increasingly plausible since the fingertip moves closer to a reasonable referent in the observer's visual field. Thus, if observers based their interpretations on the cue from which they could derive the needed information about the target location most easily, one would expect that the use of the cues depends on the perspective. Comparable to cue integration theories in sensory perception, we assume that both cues influence the interpretation simultaneously at viewpoints that do not directly favour one over the other cue. Thereby, the exact weight of each cue is perspective depended.

In three experiments, we tested the cue-weighting hypothesis, namely that observers rely increasingly on the position cue and decreasingly on the direction cue the more their perspective approaches that of the pointer. The first two experiments address whether the position cue affects pointing interpretation the more, the closer the observer perspective gets to the pointer perspective in a virtual reality (Experiment 1) and real world experiment (Experiment 2). In Experiment 3, we directly tested how the observer perspective determines the relative influence of both, the position cue and the direction cue.

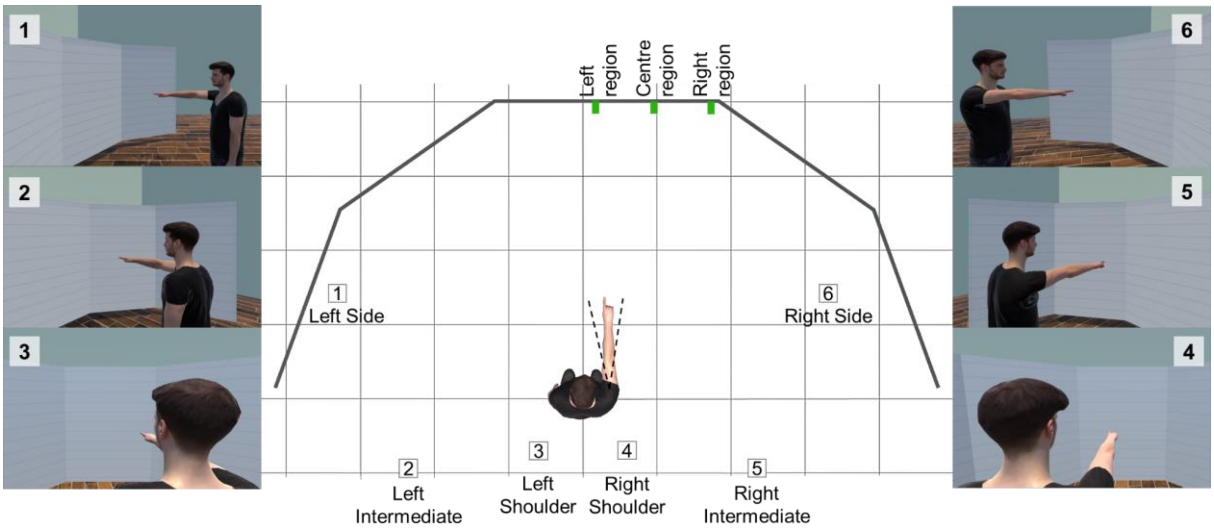
1 Experiment 1

Several experiments already revealed that observers rely on the arm direction cue when seeing the pointer from the side (Herbort & Kunde, 2016; Wnuczko & Kennedy, 2011). The objective of Experiment 1 was to test whether the fingertip position cue increasingly affects pointing perception when the observer viewpoint shifts from a lateral viewpoint to a viewpoint closer to the pointer. Therefore, participants were immersed in virtual reality (VR) and guessed where on a virtual wall a likewise virtual, computer-generated man pointed at. Observers interpreted pointing gestures from different perspectives (a sideways viewpoint, one from over the shoulder and an intermediate one, Fig. 2) to the pointer's left and right. Additionally, participants saw the points at each viewpoint from two different observer heights, while the pointer pointed horizontally to three different regions (left, centre, right) on a curved wall in front of him. Since we focused on the perception and interpretation of pointing gestures in principle, gestures were presented without speech or visible referents.

We assumed that the interpretation of the direction cue does not depend on the viewpoint (due to two fixed points, e.g. pointer's shoulder and index finger). Therefore, identical pointing gestures should be interpreted always in the same way when using the direction cue (Fig. 3a). By contrast, the observer viewpoint is expected to strongly affect the location implied by the position cue, due to the moving observer viewpoint but fixed finger position. To assess the influence of the position cue, we introduced slight variations in the *vertical* observer viewpoint (observer height). On the one hand, these small height variations are expected to have a negligible effect on the relative influence of the position cue. On the other hand, even small height variations considerably affect the *vertical* location implied by the position cue. More specifically, pointing gestures should be judged higher when seen from the lower observer height compared to a slightly higher viewpoint (see Fig. 1c). Hence, the cue-weighting hypothesis predicts a stronger effect of the observer height on *vertical* misunderstandings from a viewpoint close to the pointer than from viewpoints on the side (Hypothesis I). Additionally, based on the considerations laid out in the introduction, we hypothesise that observers judge referents as too high, especially when the pointer is seen from the side (Hypothesis II, see Fig. 1b). Moreover, we expect more leftward horizontal biases from the right shoulder viewpoint than from the left shoulder viewpoint (Hypothesis III).

Figure 2

Layout of Experiment 1



Note. Experimental setup of Experiment 1 with VR model pointing to three target regions on a curved wall. The numbers in the white boxes indicate the different observer viewpoints. Arm directions to the left and right are plotted as black, dashed lines. Different observer heights are not depicted. The surrounding screenshots correspond to the observer perspective on the pointing gesture from the respective viewpoint. The light grey squares are 50 cm x 50 cm.

1.1 A Geometric Model of the Cue-Weighting Hypothesis

However, besides these qualitative predictions, we also derived a geometric model for the cue-weighting hypothesis for three reasons. First, we wanted to assert our qualitative predictions. Second, the geometric model spells out our hypothesis in mathematical form. Third, although the model is certainly an oversimplification (e.g. because it does not take extrapolation non-linearity into account, see Herbort & Kunde, 2016), it allows an assessment of the suitability of our rather geometric approach.

Figure 3a-c gives an example of stimulus processing according to the cue-weighting hypothesis. A detailed mathematical formulation is provided in Supplement Material. In the model, interpretations are quantified as angles relative to the pointing index finger, which makes the model independent of the referent distance. The example is computed for a forward oriented pointing arm (arm azimuth and elevation = 0°). In a first step we computed the extrapolation vectors associated with the arm direction cue and fingertip position cue in the vertical and horizontal direction (Fig. 3a) for the various viewpoints of Experiment 1 (Fig. 2). As the arm has an elevation and azimuth of zero, the direction cue indicates that the referent is in front of the fingertip. The position cue implies that the referent is the more to the left/above the index finger the further to the right/lower the observer is positioned.

In a second step, the relative weight of the direction cue and the position cue were calculated based on the absolute horizontal angle between the vector of the arm and the vector from the observer viewpoint to the pointer's finger (arm view angle) for all viewpoints (Fig. 3b). Here, we assume that the direction cue fully determined pointing interpretations when the arm view angle is 90° and that the position cue increasingly determines interpretations the more the arm view angle deviates from 90° until the position cue fully determines interpretations at an arm view angle of 0° (and 180°). Finally, we computed how a gesture would be interpreted by averaging the angles implied by the position cue and the direction cue with the respective weights for that viewpoint (black lines in Fig. 3c). To compute the vertical and horizontal errors (shaded area), the interpretations were compared to the referent implied by the pointer, which was defined as the vector from the pointer's cyclopean eye to his index finger (green dotted lines in Fig. 3c).

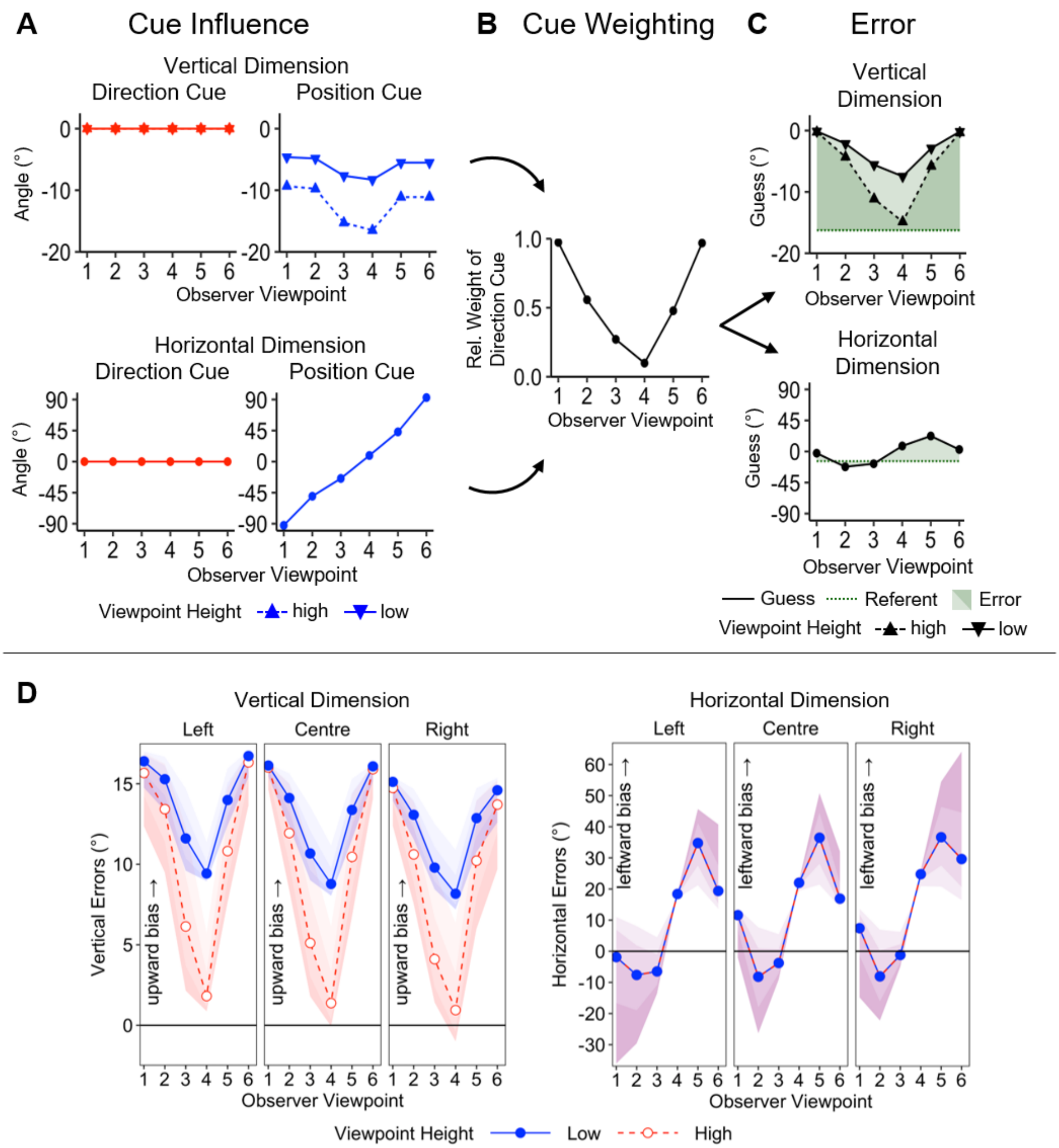
Figure 3d shows the predictions of this model. The lines indicate the model that assumes that the weight of the cues is a linear function of the arm view angle. The shaded areas indicate models that are biased toward the position or direction cue.[1] The model confirmed our qualitative predictions. First (Hypothesis I), the effect of the observer height on vertical errors is highest for shoulder viewpoints (Viewpoint 3 and 4) and smallest for side viewpoints (1, 6). Second (Hypothesis II), vertical errors are biased upward and largest at the side positions (1, 6). Third (Hypothesis III), horizontal errors are more leftward biased from the right shoulder viewpoint (4) than from the left shoulder viewpoint (2). Finally, these patterns are stable across a broad range of mappings between the observer viewpoint and the relative weighting of both cues.

---

[1] As we do not want to make strong predictions on how the arm view angle is mapped on the relative weight of the direction cue, several mappings were computed. In each mapping, the weight of the direction cue was 0.0 when the arm view angle was 0° and 1.0 when the arm view angle was 90°. However, the mappings differed with respect to whether the weights in between were biased toward the direction cue or position cue (for details, see Supplement Fig. 1). For example, the weight of direction cue at an arm view angle of 45° could assume values between .16 and .84.

Figure 3

Predictions for Experiment 1



Note. Graphic representation of our predictions on vertical and horizontal dimension of Experiment 1. **A**. The figure shows how the viewpoint affects the position and direction cue, assuming an arm azimuth/elevation of 0°. Note that observer height should not have any effect on the extrapolation of the direction cue as well as on horizontal plane thus only one line is visible. **B**. The calculated relative weight of the direction cue based on the assumptions of the cue-weighting hypothesis for each observer viewpoint, assuming a linear weighting function. **C**. Calculation of predicted interpretation errors (shaded area between referent and guess location) depending on observer viewpoints and for vertical plane additionally on viewpoint height. **D**. Ensemble functions for each target region when considering different weighting factors for the direction and position cue. Viewpoint numbers indicate the following: 1 = Left Side, 2 = Left Intermediate, 3 = Left Shoulder, 4 = Right Shoulder, 5 = Right Intermediate, 6 = Right Side.

1.2 Methods

1.2.1 Participants

24 right-handed volunteers (17 females) between 20 and 62 years (M = 33.0) recruited online from the participant pool of the Department of Psychology of the University of Würzburg signed written, informed consent. For participation, they received course credit or payment. All had normal or corrected to normal vision. To estimate the power, we bootstrapped samples (1000 data sets each for various sample sizes) from a pilot experiment and conducted the ANOVA on vertical errors. A sample size of n = 9 allows detection of the critical interaction between viewpoint and height after a Bonferroni-correction with a power of $1-\beta = .96$ ($\alpha = .05$). To be on the save side and to get a better estimate of other effects, we settled on a sample size of n = 24. This and the following experiments were approved by the local ethics committee (GZEK 2019-20).

1.2.2 Stimuli and Apparatus

Participants sat in front of a desk and put on a HTC Vive Pro or HTC Vive Pro Eye headset. Both Head Mounted Displays (HMD) differ insofar as the HTC Vive Pro Eye is equipped with an integrated eye tracker, which was not used in the reported experiments. The field of view (110°), the screens (3.5”, resolution: 1440 x 1600 pixels per eye) and other technical specifications are the same. Up to two participants were independently tested in each session. Participants used a wireless mouse with their dominant hand. Figure 2 shows the layout of the virtual scene. A male virtual, computer-generated pointer (height: 175 cm, shoulder-to-fingertip-distance: 68 cm) stood in an empty room facing a 445 cm wide and 200 cm high curved wall. To strengthen depth perception, horizontal lines were painted on the wall. The rotations of the lateral segments were 35° and 70°. Each wall was 128 cm wide (centre segment:

152 cm) and stood 2 m in front of the pointer and 20 cm above the floor. A red translucent disc served as a cursor, which could be moved over the wall by using the mouse.

Participants always assumed the observer role, while the pointing gestures were always presented by a human-like virtual person, which was used in each trial and for each participant (see Fig. 2). The pointer was placed in the origin of the virtual coordinate system. In behind conditions, the observer stood 30 cm behind the pointer and 30 cm to each side from the mid-line. For rightward viewpoints, the observer was positioned at an angle of 90° (side) or 45° (intermediate) to the extended arm of the pointer. For leftward viewpoints, the coordinates of the rightward viewpoints were mirrored to the left side of the pointer. Independently of the participant's actual height, all volunteers saw the pointing gestures from two different height levels – approx. at pointer's forehead (high – 176 cm) or his upper lip (low – 161 cm) – during the experiment. These specific height levels were chosen for two reasons: First, the pointer's fingertip had to be visible for each pointing gesture even from shoulder viewpoints. Therefore, the visibility at left behind viewpoints set the lower limit. Second, the height difference between pointer and observer should be plausible and also occur in daily life. With a height difference of 15 cm, this criterion was met.

The camera in the virtual world was rotated for each observer viewpoint so that participants always saw the pointer in front of them with no need to turn their heads. This minimised the risk for simulator sickness (e.g. headaches, nausea, vertigo, etc.) and indisposition. When pointing, the pointer extended his right arm and index finger in such a way that the pointer's cyclopean eye, index finger, and referent were aligned (Herbort & Kunde, 2016; O'Madagain et al., 2019; Taylor & McCloskey, 1988). The pointing arm was always completely stretched out in all trials (see Fig. 2, Screenshot 1). For generating the pointing gestures, the arm azimuth was set to either -10° (left), 0° (straight ahead), or 10° (right). The arm elevation was set to either -2.3° (slightly downward), 0° (horizontal), or 2.3° (slightly upward). That is, for an elevation and azimuth of 0°, the arm was perpendicular to the centre wall segment. For a more naturalistic scene, the pointer oriented his head toward the referent. Note that head orientation influences interpretations only minimal (Herbort & Kunde, 2016).

1.2.3 Procedure

At each trial onset, the virtual pointer pointed at an invisible referent location on the wall. Participants moved the cursor to the location they believed the pointer pointed at. By clicking the left mouse button, participants registered the cursor position. Afterwards, the pointer lowered his arm and looked straight ahead for 500ms, before the next trial started.

Before the first block, participants completed 12 trials of training, which we excluded from the analysis. Blocks differed with respect to the observer viewpoint (behind the shoulder, intermediate, side; each on right and left side of the pointer) and viewpoint height (176 cm, 161 cm). In each block, all nine different referents (3 arm elevations x 3 arm azimuths) were presented once in a pseudo-random order. Each block type was presented five times. Participants went through all blocks in randomised order. In summary, the experiment involved 540 trials split in 60 blocks with self-paced breaks and took approximately 22 minutes.

1.2.4 Data Reduction and Analysis

The hypothetical referent locations were computed by intersecting the line through the pointer's cyclopean eye and fingertip with the wall (Mayer et al., 2018). We defined the angular horizontal and vertical error as the difference between the referent and the participant's guess in the respective dimension. As misunderstandings can be expected to be practically zero when the finger touches the referent and as both cues originate from the fingertip, the pointer's fingertip served as vertex for the computation of angular errors. A positive value indicates that the guess was to the left or above the actual referent location. In the current experiment, one degree of error corresponded to approximately 2.3 cm on the wall or 0.7° of visual angle from the pointer's perspective. Notice, that in similar experiments other vertices were used, for example the shoulder of the pointer (Bangerter & Oppenheimer, 2006). Trials in which the guess deviated horizontally or vertically by more than 50 cm from the participant's mean in the respective condition (data split by region, viewpoint, height) were excluded from analysis (1.6% of trials).

1.3 Results and Discussion

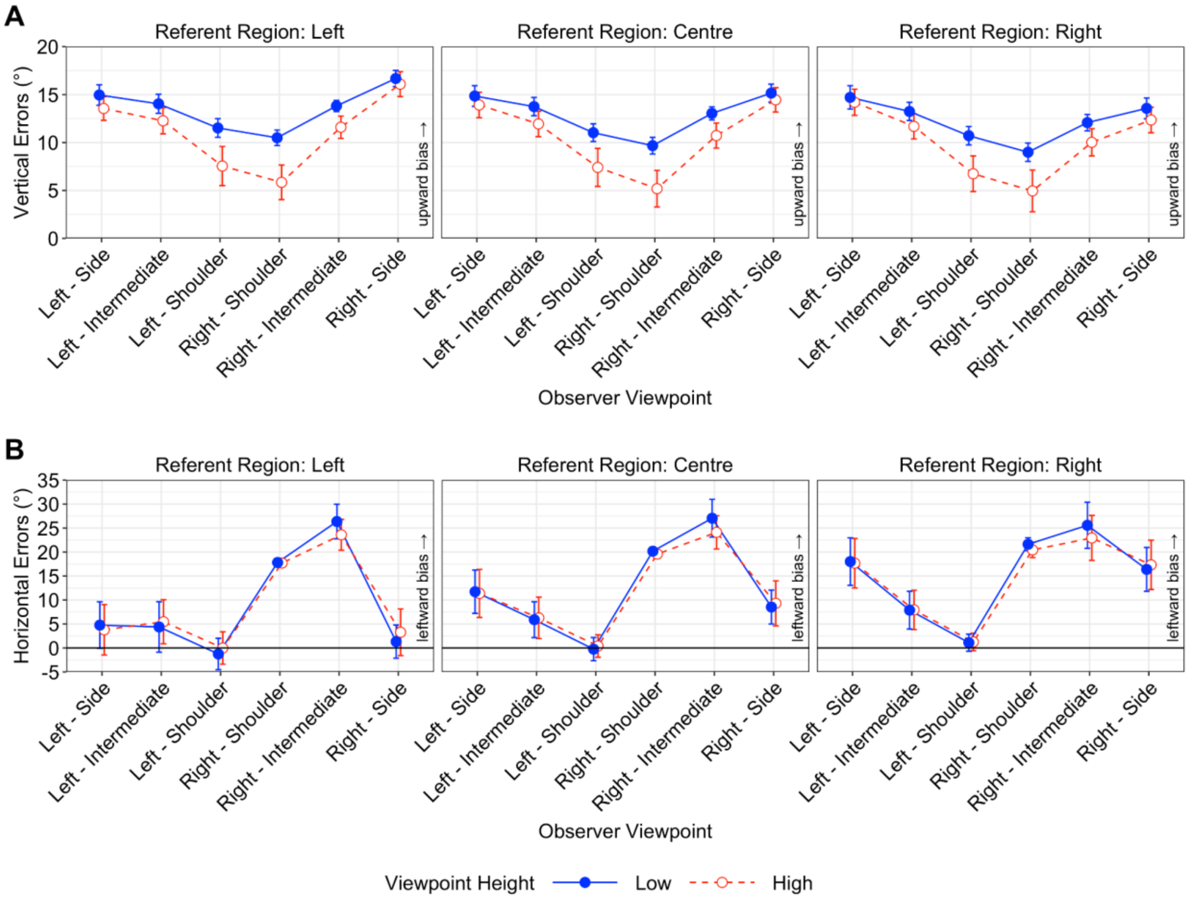
The influence of perspective on the interpretations of pointing gestures has been analysed with repeated measures ANOVA with within-subject factors viewpoint (left side, left intermediate, behind left shoulder, behind right shoulder, right intermediate, right side), height (high, low) and region (left, centre, right) separated for both dimensions. Figure 4 shows the vertical and horizontal errors. Table 1 reports the results of the ANOVAs. We report the Greenhouse-Geisser corrected *p*-values to correct for sphericity violations of repeated measures ANOVA[2]. We first report the statistics pertaining to the predicted effects (Hypothesis I-III) and then report and discuss additional significant effects.

---

[2] Note that we however report – for this and all following experiments – the uncorrected *df*s and additionally the Greenhouse-Geisser-ε.

Figure 4

Mean Angular Errors of Experiment 1



Note. The figure shows mean angular errors in the vertical (**A**) and horizontal (**B**) dimension for each of three target regions. Error bars show 95% CI. Positive values indicate guesses too far to the left or above. One degree corresponds to approximately 2.3 cm.

Table 1

Results of ANOVA for Vertical and Horizontal Errors

Effect	Vertical					Horizontal				
	F	df	p	η_p^2	ε	F	df	p	η_p^2	ε
Height	**56.28**	**(1,23)**	**< .001**	**.71**	**-**	0.86	(1,23)	.363	.04	-
Viewpoint	**151.76**	**(5,115)**	**<.001**	**.87**	**.46**	**26.76**	**(5,115)**	**<.001**	**.54**	**.29**
Region	**241.71**	**(2,46)**	**<.001**	**.91**	**.89**	**19.86**	**(2,46)**	**<.001**	**.46**	**.53**
Height x Viewpoint	**35.93**	**(5,115)**	**<.001**	**.61**	**.49**	**5.66**	**(5,115)**	**.002**	**.20**	**.56**
Height x Region	**3.95**	**(2,46)**	**.027**	**.15**	**.99**	1.50	(2,46)	.236	.06	.94
Viewpoint x Region	**36.11**	**(10,230)**	**<.001**	**.61**	**.40**	**27.34**	**(10,230)**	**<.001**	**.54**	**.33**
Height x Viewpoint x Region	**3.87**	**(10,230)**	**.002**	**.14**	**.57**	0.64	(10,230)	.678	.03	.53

Note. Significant main effects and interactions are printed in bold.

1.3.1 Predicted Effects

Hypothesis I: As expected, the interaction between height and viewpoint for vertical guesses reached significance in the predicted direction. The influence of view height increased continuously when the pointer-observer distance decreased. The effect of height (averaged over regions) increased from side to intermediate viewpoints and from intermediate viewpoints to shoulder viewpoints for both the left and right side, as indicated by repeated contrasts, all $F(1,23) \geq 11.38$, all $p \leq .003$, all $\eta_p^2 \geq 0.33$. Thus, the central hypothesis of Experiment 1 was confirmed.

Hypothesis II: As expected, participants guessed the referent higher than intended, which can be confirmed by comparing the 95% confidence intervals (CI) in Figure 4a with zero.

Hypothesis III: Judgements from the left shoulder viewpoint were considerably further to the right than judgements from the right shoulder viewpoint, as a contrast shows, $F(1,23) = 184.35$, $p < .001$, $\eta_p^2 = 0.89$.

Our model suggests that both cues are integrated on the level of individual trial. However, the data could also result from some participants generally preferring the arm direction and others preferring the fingertip position cue for the interpretation or because interpretations were based on the direction cue in some trials and the position cue in others. However, Dip-Tests for bi-/multimodality and visual inspections of histograms ruled out this possibility (see for details the Supplement Material).

1.3.2 Vertical Errors

Moreover, several additional effects reached significance. In line with our geometric model, holding the lower viewpoint led to higher judgements. Furthermore, the effect of observer viewpoint reached significance. Observers overestimated the referent location more when standing at the side than at the intermediate viewpoints and more when standing at the intermediate viewpoints than on the shoulder ones, as indicated by repeated contrasts for each side, all $F(1,23) \geq 37.03$, all $p < .001$, all $\eta_p^2 \geq 0.62$. As the observer viewpoint was always above the pointer's shoulder, this is in line with our hypothesis (and was also predicted by the geometric model). Additionally, there was a main effect of the region, which was modulated by the observer viewpoint. Descriptively, the effect of region was largest at the right side viewpoint and decreased the more, the more the observer viewpoint moved toward the left. This might by rooted in the greater distance between observer and fingertip, which results in a smaller effect of the region on the location indicated by the position cue, at right side viewpoints compared to left ones. Observer height affected the interpretation significantly more at the left than at the right target region, as a contrast shows, $F(1,23) = 7.86$, $p = .010$, $\eta_p^2 = 0.26$. Finally, the three-way interaction was also significant, likely because the vertical bias decreased from left to right region independently of observer height and viewpoint except for the higher right side viewpoint, where this pattern reversed.

1.3.3 Horizontal Errors

The effect of target region reached significance. Leftward biases were smallest for contralateral pointing gestures and increases to ipsilateral points, as indicated by repeated contrasts between adjacent regions, $F_{\text{Left/Centre}}(1,23) = 20.47$, $p < .001$, $\eta_p^2 = 0.47$ and $F_{\text{Centre/Right}}(1,23) = 17.19$, $p < .001$, $\eta_p^2 = 0.43$. Moreover, region and viewpoint interacted because the effect of region was descriptively larger at the sideward viewpoints than at the other viewpoints. That is, leftward biases were highest for points to the right and relatively low for leftward points. One possible explanation is that the perception of the respective azimuth is difficult to perceive and this ambiguity fosters central tendencies (Herbort et al., 2020). Depending on the target region, the central tendency is differently impacting error sizes of interpretations, while the absolute interpretation error depends on the viewpoint and the general bias direction on only using the right arm for pointing gestures. Finally, the significant interaction between height and viewpoint most likely emerged because height affected horizontal judgements at the right intermediate viewpoint but did not have a considerable effect at the remaining viewpoints.

1.3.4 Conditions of Maximal Accuracy

Our hypothesis implies that accuracy should be highest when pointer and observer share their viewpoint. Under this condition, the position cue receives most weight and is the least biased. In Experiment 1, the conditions in which observers were most accurate were indeed those in which the pointer's and observer's view were aligned the most. For vertical interpretations, the vertical angle between the eye-finger lines of pointer and observer determines the bias of the position cue. This angle is smallest at the right shoulder viewpoint. Vertical interpretations were indeed most accurate at the right shoulder viewpoint, as indicated by paired-sample *t*-tests of average absolute vertical errors at the right shoulder viewpoint compared to all other viewpoints, all $t(23) \leq -7.42$, all $p < .001$, all $d \leq -1.51$.

For the horizontal interpretation, the horizontal angle between the eye-finger lines is minimal at the left shoulder viewpoint. At this viewpoint the absolute average errors were smallest compared to all other viewpoints, all $t(23) \leq -2.88$, all $p \leq .008$, all $d \leq -0.59$.

1.3.5 Summary

In summary, Experiment 1 showed that the observer's viewpoint has a considerable effect on pointer-observer misunderstandings. Most important, it supported the cue-weighting hypothesis. Although both cues were potentially available in all conditions, the position cue had a larger effect – as indicated by the effect of height – when the observers aligned their view with that of the pointer. Contrary, it had almost no effect when observers watched the pointer from the side. Finally, the general pattern of empirical results resembled the predictions of our geometric model remarkably well, lending additional credibility to the cue-weighting hypothesis.

2 Experiment 2

We conducted Experiment 1 in VR because this enables precise control over the pointer's posture and the observer viewpoint. However, it raises the question whether similar results would be found in a real-life setting. On the one hand, comparisons between pointing perception in artificial and real-life settings found no fundamental differences (Herbort & Kunde, 2016). On the other hand, VR clearly differs from reality in numerous ways that could have affected the results. For example, participants often underestimate the egocentric distance of objects in VR settings (El Jamiy & Marsh, 2019; Ng et al., 2016; Renner et al., 2013; Wong & Gutwin, 2010). If such underestimations affect the pointing arm and the wall to different degrees (e.g. due to differences in the graphics quality) pointer-observer biases could be distorted – especially from more oblique viewpoints. Likewise, the necessarily limited resolution of VR and our 3D models could prevent participants from using more nuanced information that is readily available in reality. Additionally, the generalisability could be limited by differences in the execution of the pointing gesture between virtual and real pointer, which would also slightly affect the interpretation. To examine whether the previous results generalise to the real world, we reconstructed Experiment 1 in a real world setting in Experiment 2. Note that Experiment 2 should not be considered a direct replication of Experiment 1, as a number of alterations had to be made. Most importantly, the number of viewpoints was reduced, the observer height was now manipulated by moving pointer and referents to a higher or lower position relative to observers, and observers noted the coordinates of their guess on a clipboard. Nevertheless, the critical aspects of Experiment 1 were adopted in Experiment 2. That is, we kept the spatial dimension (distance to and size of the wall, general distances between pointer and observers, same difference in heights) and the pointing gestures and corresponding referent locations. Furthermore, we kept the shoulder and side viewpoints, which we expected to differ most with respect to the relative contribution of each cue. We expect the data from VR environment to be generalizable and therefore, a data pattern similar to Experiment 1 should emerge. More precisely, if the cue-weighting hypothesis also holds in the real world, we again expect a larger effect of the (relative) observer height on vertical judgements for the shoulder viewpoints (Hypothesis I) as well as the corollary expectations that we also had for Experiment 1 (Hypotheses II-III).

2.1 Method

2.1.1 Participants

Thirty-four new volunteers (27 females, 33 right handed) between 20 and 67 years (M = 32.9) participated after signing written, conformed consent and received course credit or payment. To estimate the power, we bootstrapped samples (1000 data sets for each of various sample sizes) from the data of Experiment 1 (excluding the intermediate condition) and conducted the ANOVA on vertical errors. A sample size of n = 12 allows detection of the critical interaction between viewpoint and height after a Bonferroni-correction with a power of $1-\beta = .97$ (α =

.05). We decided to scheduled nine four-participant sessions to be able to collect valid data of at least 24 participants but that allowed for a total of 36 sign ups.

2.1.2 Stimuli and Apparatus

We transferred the experimental design of Experiment 1 as follows (Fig. 5). A wooden scaffold with posters served as the curved wall. To simplify the referent identification for observers, roman numerals (I-V) marked the wall segments. Posters were printed with white and grey squares, each 4 cm x 4 cm, labelled horizontally with numbers 1-50 and vertically with letters A-Z, followed by AA-AF (centre wall: A-AL; see Fig. 5).

During the experiment, all participants (observers) sat down on chairs. The assistant (pointer) sat on a height-adjustable stool. All chairs were precisely aligned in advance, marking each observer viewpoint and the exact pointer position. We had to modify the shoulder viewpoints, after we noticed that the pointing arm and finger were not always completely visible for all participants. Consequently, we moved them further behind (81 cm) and apart (85 cm) from the pointer.[3] Side viewpoints were at the pointer's shoulder height and 165 cm away. The intermediate observer viewpoints of Experiment 1 were not used this time. In this experiment, we manipulated the height of the pointer by adjusting the stool-height instead of manipulating the height of the observers (as in Exp. 1). Therefore, the pointer's seat height was either 38 cm or 53 cm above the floor. To keep the analysis comparable to the previous experiment, we coded the factor (relative) viewpoint height from the observer perspective. That is, a high viewpoint implies that stool and pointer held the lower position and vice versa.

A female assistant (173 cm tall, shoulder-to-fingertip distance: 75 cm) always held the pointer role and pointed in each trial by extending her right arm in a straight line and by placing

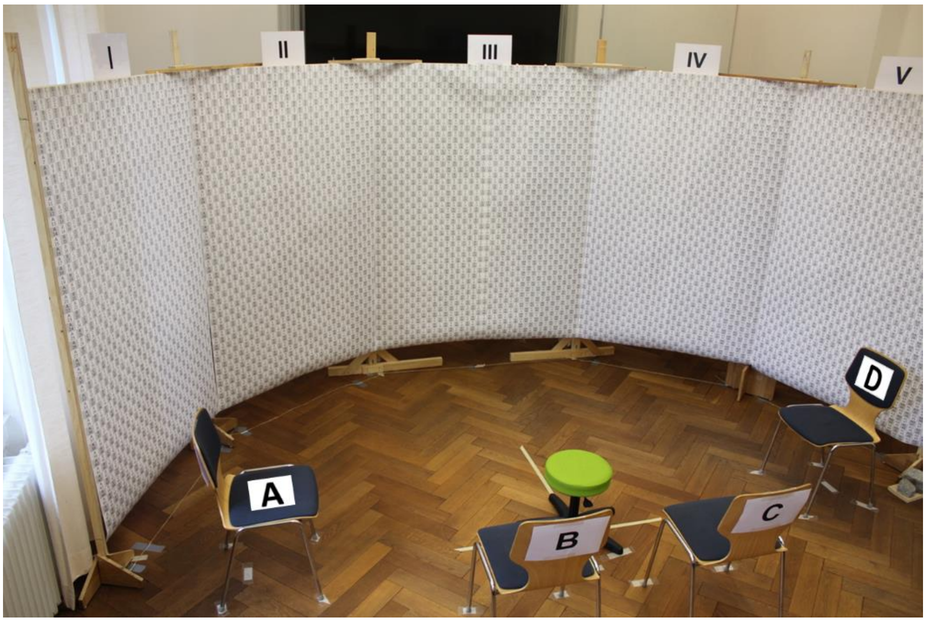
---

[3] Like in Experiment 1, the interpretation errors should be lowest where the angle between the eye-finger lines of pointer and observer is smallest – at both shoulder viewpoints. Despite the distance modifications, this is still true for the left shoulder viewpoint for horizontal errors and for the right shoulder viewpoint for vertical errors.

the fingertip in the centre of the relevant square in her visual field. We determined the referents in advance so that the same arm azimuths and elevations as in Experiment 1 were adopted by the pointer. The pointer used two different sets of referents to enable identical arm postures when pointing from the low and high viewpoint. An experimenter stood in the back of the room to coordinate the setup of the various blocks and instructed the pointer.

Figure 5

Layout of Experiment 2



Note. Observer viewpoints side (chairs A, D) and behind (chairs B, C). The assistant sat on the green height-adjustable stool. All target regions/referents were on the centred wall III.

2.1.3 Procedure

Up to four volunteers participated in each session and were instructed to not talk during the experimental trials. At each trial onset, participants closed their eyes while the experimenter showed a referent code (e.g. III N32) to the pointer. Then, the pointer pointed at the referent and observers wrote down their answer on a clipboard and said “yes”. Once all participants completed the trial, the pointer lowered her arm and the next trial started.

Three initial trials of training were not included in the analysis. The remainder of the experiment was divided into eight blocks (four viewpoints x two pointer heights), each consisted of 27 (nine different referents repeated three times) trials in pseudorandom order. Block order was pseudo-randomised under the constraint that participants within one session always occupied different observer viewpoints. Between blocks, the experimenter asked the observers to change seats and told the pointer to adjust the stool height (low, high) if necessary. In summary, the experiment involved 270 trials and took approximately 60 minutes in total.

2.1.4 Data Reduction and Analysis

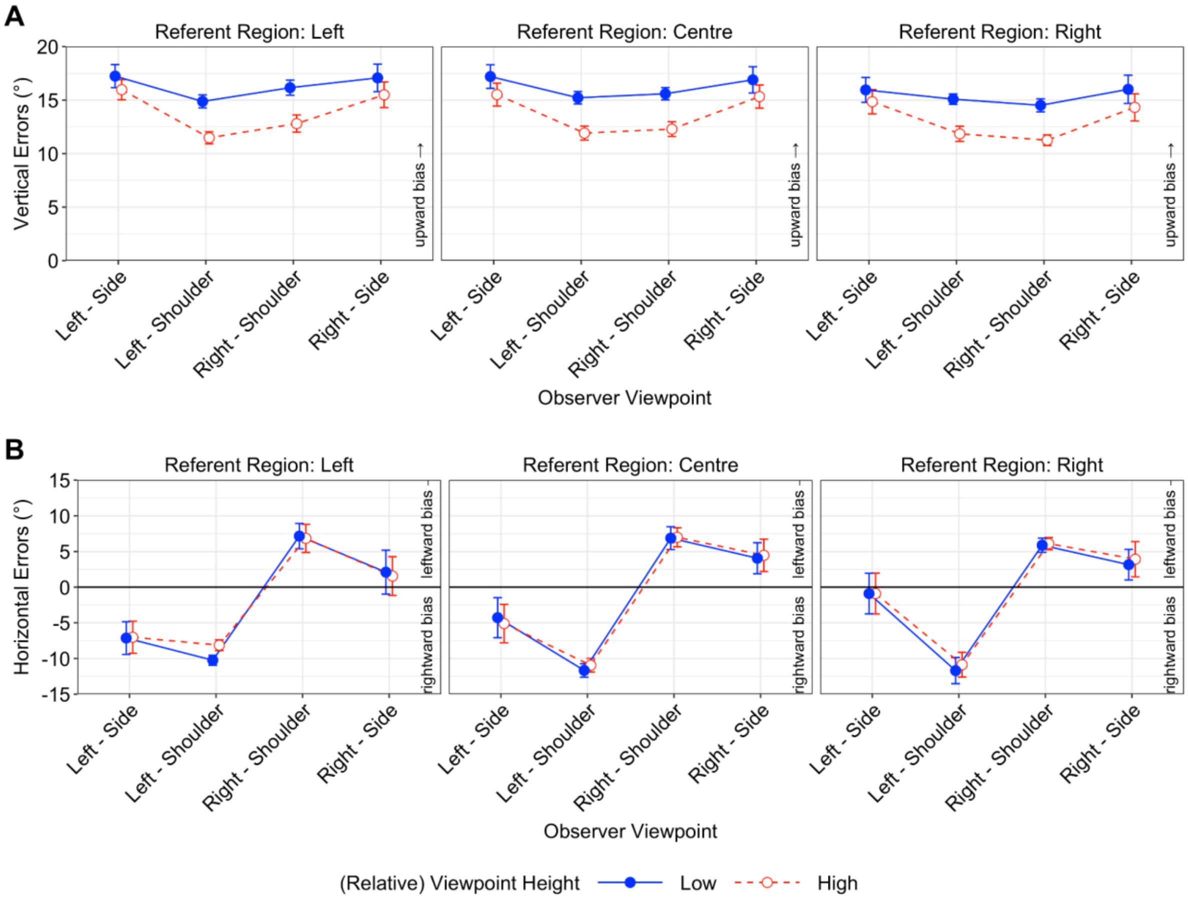
Observers' judgements were transformed into 3D locations based on the centres of the relevant squares' coordinates. On that basis, the angular deviation was calculated identically to the previous experiment. Note that the results are relatively insensitive to the exact location of the pointer's finger. Changing the pointer's finger position by 20 cm along each of the cardinal axes would not have resulted in a different pattern of significant and non-significant ANOVA effects. Using the previously used outlier criterion, 0.8% of the trials were excluded from analysis.

2.2 Results and Discussion

The influence of perspective on spatial interpretation of pointing gestures has been analysed with repeated measures ANOVA with within-subject factors viewpoint (left side, behind left shoulder, behind right shoulder, right side), relative viewpoint height of the observer (high, low) and region (left, centre, right) separated for both dimensions. Figure 6 shows vertical and horizontal errors. Table 2 reports the results of the ANOVAs, including Greenhouse-Geisser corrected *p*-values.

Figure 6

Mean Angular Errors of Experiment 2



Note. The figure shows mean angular errors in the vertical (**A**) and horizontal (**B**) dimension for each of three target regions. Error bars show 95% CI. Positive values indicate guesses too far to the left or above. One degree corresponds to approximately 2.3 cm.

Table 2

Results of ANOVA for Vertical and Horizontal Errors

Effect	Vertical					Horizontal				
	F	df	p	η_p^2	ε	F	df	p	η_p^2	ε
Height	**226.33**	**(1,33)**	**< .001**	**.87**	**-**	1.81	(1,33)	.187	.05	-
Viewpoint	**32.31**	**(3,99)**	**<.001**	**.50**	**.52**	**73.17**	**(3,99)**	**<.001**	**.69**	**.45**
Region	**66.91**	**(2,66)**	**<.001**	**.67**	**.90**	1.91	(2,66)	.175	.06	.55
Height x Viewpoint	**17.64**	**(3,99)**	**<.001**	**.35**	**.85**	1.49	(3,99)	.229	.04	.77
Height x Region	0.62	(2,66)	.528	.02	.93	0.56	(2,66)	.549	.02	.87
Viewpoint x Region	**11.28**	**(6,198)**	**<.001**	**.26**	**.73**	**21.81**	**(6,198)**	**<.001**	**.40**	**.54**
Height x Viewpoint x Region	0.77	(6,198)	.562	.02	.77	1.81	(6,198)	.152	.05	.49

Note. Significant main effects and interactions are printed in bold.

2.2.1 Predicted Effects

Hypothesis I: According to our main hypothesis, the effect of height on vertical judgements depended on the observer viewpoint, as indicated by the significant height x viewpoint interaction. Contrasting the averaged vertical errors revealed that the influence of relative observer height was significantly lowered when holding the side viewpoints compared to shoulder viewpoints, $F(1,33) = 48.76$, $p < .001$, $\eta_p^2 = 0.60$.

Hypothesis II: Participants overestimated the height of the referent in all conditions systematically, which can be confirmed by comparing the 95% CI in Figure 6a with zero.

Hypothesis III: Sitting on the pointers left led to rightward biases and vice versa. This effect was found at all viewpoints but was most pronounced at shoulder viewpoints (Hypothesis III), $F_{\text{Side}}(1,33) = 15.04$, $p < .001$, $\eta_p^2 = 0.31$ and $F_{\text{Shoulder}}(1,33) = 320.02$, $p < .001$, $\eta_p^2 = 0.91$.

2.2.2 Vertical Errors

The referent was guessed higher at side than at shoulder viewpoints, which is in line with a stronger influence of the direction cue for the side viewpoints, $F(1,33) = 41.79$, $p < .001$, $\eta_p^2 = 0.56$. As we expected a pattern comparable to Experiment 1, we checked for differences between corresponding viewpoints on the pointer's left and right with further contrasts. We found a significant asymmetry in estimations from shoulder viewpoints, $F_{\text{Shoulder}}(1,33) = 4.30$, $p = .046$, $\eta_p^2 = 0.12$, but not from side viewpoints, $F_{\text{Side}}(1,33) = 1.26$, $p = .270$, $\eta_p^2 = 0.04$.

Lower observer viewpoints resulted in higher judgements. While the difference between left and centre region was not significant, vertical errors declined from centre to right target region, indicated by a repeated contrast, $F(1,33) = 84.90$, $p < .001$, $\eta_p^2 = 0.72$. Viewpoint and region interacted. When the observer was on the left, points to the left were interpreted as higher than points to the right. When the observer sat on the right, this pattern reversed.

2.2.3 Horizontal Errors

Target region had no effect on horizontal errors but further modulated the significant main effect of viewpoint. Like in Experiment 1, the region descriptively affected errors at side viewpoints more than at shoulder viewpoints. The region had only a mild effect on the shoulder viewpoints, but a considerable effect on the side viewpoints, in which rightward points were interpreted as further to the left. Finally, note that descriptively the general leftward bias of Experiment 1 did not reappear in Experiment 2. The latter finding is in line with the findings of Bangerter and Oppenheimer (2006).

2.2.4 Effect of VR

Concerning the comparability of VR and real world setting, the results of this experiment show qualitatively the same structure as in Experiment 1 and confirm our core hypotheses I-III in a real life setting. Nevertheless, since this experiment was not a replication of Experiment 1, it was to be expected that the data would match those from the first experiment only partially. For example, the magnitude of vertical biases as well as the effect of the (relative) observer height were similar in both experiments. By contrast, the general leftward bias found in Experiment 1 was not present in Experiment 2, although the effect of viewpoint on horizontal judgements was comparable between experiments. We can only speculate why the leftward bias was present in Experiment 1 and not in Experiment 2. As already mentioned in the introduction, the execution of the pointing gesture could be one possibility. That is, the virtual pointer fully aligned arm and finger but the assistant aligned the azimuth of the index finger more to her line of sight, thus reducing the discrepancy between the direction cue and the vector from the fingertip to the actual referent.

Unlike in Experiment 1, pointing errors were not lowest in the conditions in which the pointer and observer perspective were most similar. This can be attributed to the modified distances between pointer and observer at shoulder viewpoints. While very small deviations in the interpretation can only be achieved by keeping pointer and observer as close to each other as possible, even relatively small increases in the distance, as were necessary in Experiment 2, lead to a significant increase in the interpretation errors. Furthermore, while Experiments 1 and 2 show pointing interpretations follow similar general rules in VR and real world settings, they hardly allow a more fine-grained, quantitative comparisons of both settings. However, it is possible and rather likely that even with an exact replica of the design, significant differences between VR and reality would be found in the size of the errors, as suggested by Mayer et al. (2018). Although the main focus there was on pointer performance, the authors were able to identify significant differences in the pointing accuracy between VR and reality. Considering the reduced field of view on HMDs and limited depth criteria, such as accommodation, deviations seem plausible also for the interpretation of pointing gestures as they also restrict the observer’s perception. Nevertheless, Experiments 1 and 2 show that VR is a valid and promising method for examining pointing perception in the real world. Moreover, recent research has suggested a variety of measures that could further close the gap between reality and virtual reality. For example, people underestimate distances in VR by approximately 15%, whereas they have little difficulty in analogue situations (Ng et al., 2016). More depth cues in the form of structured materials like wallpapers, floor covering, and detailed realistic and complex surroundings (e.g. furniture, doors, windows) could offer an easy to realise remedy here (Renner et al., 2013). In addition, recreating the actual lab in VR and giving participants an avatar could further improve the depth perception by conveying further reference (Renner et al., 2013).

In summary, the accordance of findings in the qualitative structure as well as the similarity in effect sizes and absolute errors with Experiment 1 are remarkable even when not all depth cue-providing possibilities had been exhausted, and allow insights gained in VR to be transferred to reality. In conclusion, Experiment 2 confirms in a real life setting that observers increasingly rely on the position cue when aligning their view with that of the pointer.

3 Experiment 3

Previous experiments in which observers watched the pointer from the side revealed that observers primarily rely on the arm direction cue (Herbort & Kunde, 2016; Wnuczko & Kennedy, 2011). In Experiments 1 and 2, we demonstrated that in addition, the fingertip position cue is used during pointing interpretation and becomes increasingly influential, when the observer shifts from a sideward viewpoint next to the pointer position. However, as the arm elevation was kept relatively constant in these experiments, the data do not allow conclusions about the effect of the observer viewpoint on the influence of the direction cue and its relation to the position cue. Therefore, Experiment 3 directly examined the relative weighting of both cues in the interpretation of pointing gestures for various observer perspectives. Thus, we orthogonally manipulated the target locations implied by the fingertip position (position cue) and the arm direction (direction cue).

In VR, participants looked at a pointing arm from various observer viewpoints while standing inside a hollow cylinder (Fig. 7). The fingertip was either located somewhat above or below the observer's point of view. Using the position cue, namely extrapolating the line between the observer's eyes and the pointing finger, thus resulted in two unique heights (H and L). Additionally, the arm was pointing either upward or downward, so that it was oriented in such a way that the direction cue also indicated either height H or height L. To implement these manipulations, solely a disembodied forearm with a realistic hand was presented for pointing gestures. While resulting in a less naturalistic setting, this approach guaranteed full visibility from all viewpoints.[4] Note that pointing perception relies almost exclusively on the perception of the arm, hand and fingertip, and the observer perspective, whereas factors such as gaze direction (Cooney et al., 2018), head direction (Herbort & Kunde, 2016) or visibility of the pointer's body (Bangerter & Oppenheimer, 2006; Herbort & Kunde, 2016) hardly affect pointing perception. That is, if participants exclusively relied on the direction cue, the arm direction but not the finger position should affect vertical judgements. Conversely, if participants only relied on the position cue the finger position but not the arm direction should affect vertical judgements. If both cues were involved we expect participants to select intermediate locations from which indications of the relative contribution of both cues to the interpretation can be derived. Note that unlike in Experiments 1 and 2, the core dependent variable is the position indicated by the observer and not the misunderstanding between pointer and observer.

Based on the previous findings and the before-mentioned assumptions, we expect the following pattern: a decrease of the effect of the arm direction and an increase of the effect of the finger position on vertical judgements when moving from a sideward viewpoint to a perspective that is more aligned with the arm. Although the horizontal judgements are not critical for Experiment 3, our assumptions also have implications for that dimension. We again expect

---

[4] This had the additional advantage that we were able to ensure that the participants' interpretations were not influenced by differences in body postures, which would have been necessary to realize the different arm-hand configurations in a naturalistic way.

that interpretations from a viewpoint behind the arm but to the left should fall further to the right than interpretations made from behind the arm but to the right. Furthermore, this effect should decrease the further the perspective approaches a side view.

3.1 Methods

3.1.1 Participants

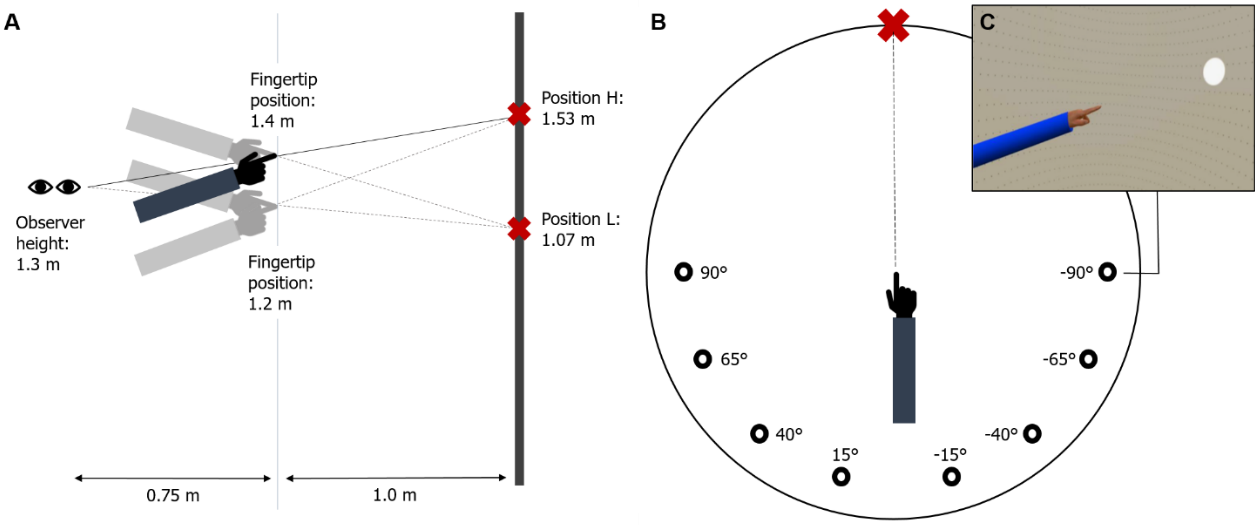
After signing written, informed consent, 36 new volunteers (28 women) between 18 and 58 years (*M* = 23.89) participated in the experiment. All participants were right-handed and received payment or course credit. As the previous experiments revealed large effects of the perspective that could be easily detected with sample sizes between 9 and 12 participants, we expected similar conditions for Experiment 3. Nevertheless, to get a better estimate for other effects and as a safety margin, we collected data of 36 participants.

3.1.2 Stimuli and Apparatus

As in the previous experiments, participants always held the observer role. Comparable to Experiment 1, they sat in front of a desk and wore a HTC Vive Pro or HTC Vive Pro Eye headset. Within the virtual scene, participants were placed inside a 2.6 m high hollow cylinder with a diameter of 2 m. Small, dark dots on the cylinder wall should strengthen the depth impression as this visually reinforces the parallax that results from the different distances between participant and pointing arm or cylinder wall while moving. Participants hold eight different viewpoints (±15°, ±40°, ±65°, ±90° relative to the pointing arm vector; see Fig. 7), at a distance of 75 cm from the cylinders' centre and 1.30 m above ground level. Participants saw always the same virtual forearm and hand (shoulder-to-fingertip-distance: 80 cm, size of the hand comparable to Experiment 1) without body. A left hand was displayed when standing on the right and a right hand was displayed when standing on the left to guarantee an unblocked view of the index finger. The fingertip was always in the cylinder's centre with an arm azimuth of constantly 0°. The arm was oriented in such a way, that both, the extrapolation of the vector from the ‘shoulder’ to the fingertip (direction cue) and the extrapolation of the vector between the observers’ eye and pointer’s finger (position cue), intersected the cylinder at a height of either 1.07 m or 1.53 m. That is, the fingertip was always either 1.20 m or 1.40 m above ground level. The arm’s elevations were 7.60° or -18.43° when the finger occupied the 1.40 m position and 18.43° or -7.60° when the finger occupied the 1.20 m position.

Figure 7

Layout of Experiment 3



Note. Schematic layout of the experimental setup. **A.** Profile of the setting. An arm in four different arm postures (all possible combinations of upward/downward arm direction and high/low index finger position) pointed either at two referent heights H or L (red crosses) on the wall straight ahead. Observer height was fix at 1.3 m, thus the eye-finger line indicated either height H or L. **B**. Plan view of the setting. Circles show different observer viewpoints. The arm always pointed straight ahead on the inner wall of the cylinder. **C.** The screenshot of the experiment exemplary depicts the observer's perspective from the -90° viewpoint on the forearm and hand. The white disc served as the cursor with which the pointed-at location was marked by the observers.

3.1.3 Procedure

Up to two volunteers participated independently in one session. Four initial trials of training were excluded from final analysis. At each trial-onset, the arm was presented and observers where placed at one of eight viewpoints. After moving a cursor with the mouse to the assumed referent location and clicking the left mouse button, the screen went grey for 500 ms. Then, the next trial started. Each of eight blocks contained all 32 possible combinations of pointing gesture and observer viewpoint twice. The experiment took on average 27 minutes.

3.1.4 Data Reduction and Analysis

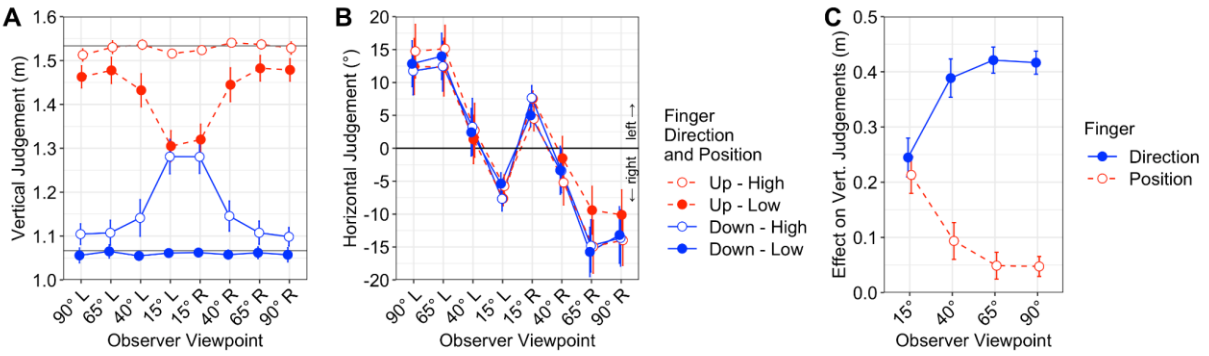
Unlike Experiments 1 and 2, we analyse the absolute location of the guesses in the current experiment. As the distance from the fingertip to the wall was always 1.0 m, the vertical guess location was measured as distance from the ground (and not as angle as in the other experiments). As horizontal location, the angular position of the cursor relative to the 0° position was used. Positive values indicating guesses to the left. The outlier threshold was identical to Experiment 1, resulting in the exclusion of 1.7% of trials from analysis.

3.2 Results and Discussion

The mean vertical and horizontal interpretations were analysed with repeated measures ANOVA with within factors fingertip position (high, low), arm direction (upward, downward), observer viewpoint (15°, 40°, 65°, 90°) and observer side (left, right) for both dimensions. Table 3 reports the results of the ANOVAs with Greenhouse-Geisser corrected *p*-values. *P*-values that do not survive the Bonferroni-Holm correction are in parenthesis and are not interpreted. Results are shown in Figure 8.

Figure 8

Mean Errors of Experiment 3



Note. **A**. Vertical guess location depending on observer viewpoint, finger direction and fingertip position. The grey lines indicate the y-coordinates where the eye-finger-lines and the extrapolation of the arm intersect the wall. **B**. Horizontal angular guess location depending on observer viewpoint, finger direction and fingertip position. Black line at 0° indicates the direction in which the arm was oriented. Positive values donate leftward biases. **C**. The isolated effect of both finger direction and finger position on participant's estimates from different viewpoints. Error bars show 95% CI.

Table 3

Results of ANOVA with Vertical and Horizontal Errors.

Effect	Vertical					Horizontal				
	F	*df*	*p*	η_p^2	*ε*	*F*	*df*	*p*	η_p^2	*ε*
Arm Direction	**933.8**	**(1,35)**	**<.001**	**.96**	**-**	3.9	(1,35)	.056	.10	-
Fingertip Position	**62.7**	**(1,35)**	**<.001**	**.64**	**-**	5.8	(1,35)	(.021)	.14	-
Viewpoint	2.0	(3,105)	.156	.05	.54	0.6	(3,105)	.563	.02	.74
Observer Side	1.8	(1,35)	.183	.05	-	**17.1**	**(1,35)**	**<.001**	**.33**	**-**
Direction x Position	0.3	(1,35)	.610	.01	-	4.6	(1,35)	(.039)	.12	-
Direction x Viewpoint	**82.4**	**(3,105)**	**<.001**	**.70**	**.48**	**7.9**	**(3,105)**	**<.001**	**.18**	**.88**
Direction x Side	4.6	(1,35)	(.039)	.12	-	1.5	(1,35)	.227	.04	-
Position x Viewpoint	**158.8**	**(3,105)**	**<.001**	**.82**	**.54**	1.0	(3,105)	.372	.03	.84
Position x Side	0.6	(1,35)	.461	.02	-	**13.6**	**(1,35)**	**.001**	**.28**	**-**
Viewpoint x Side	0.4	(3,105)	.676	.01	.68	**144.0**	**(3,105)**	**<.001**	**.80**	**.69**
Direction x Position x Viewpoint	2.5	(3,105)	.084	.07	.70	1.2	(3,105)	.315	.03	.84
Direction x Position x Side	0.2	(1,35)	.685	.01	-	**28.0**	**(1,35)**	**<.001**	**.45**	**-**
Direction x Viewpoint x Side	1.1	(3,105)	.359	.03	.90	2.6	(3,105)	.068	.07	.86
Position x Viewpoint x Side	0.6	(3,105)	.600	.02	.90	3.8	(3,105)	(.016)	.10	.89
Direction x Position x Viewpoint x Side	0.7	(3,105)	.568	.02	.92	2.7	(3,105)	.066	.07	.79

Note. *p*-Values that do not survive the Bonferroni-Holm correction are in parentheses. Significant main effects and interactions are printed in bold.

3.2.1 Predicted Effects

The results confirmed our hypotheses. First, the expected interaction between viewpoint and arm direction on the vertical dimension was significant. Figure 8c shows that the effect of arm direction on vertical judgements was smallest when the view was aligned with the arm (15°) and increased for more sideward viewpoints (90°). Likewise, the expected interaction between viewpoint and fingertip position was significant. Figure 8c shows that the influence of the finger position on vertical judgements was greatest in the 15° condition and decreased the more it deviated from the 15° viewpoint. Both effects are also significant if only the viewpoint levels 15° and 90° are included in the analysis; viewpoint x direction: $F(1,35) = 89.26$, $p < .001$,

η_p^2 = .72, position x viewpoint: $F(1,35) = 183.72$, $p < .001$, η_p^2 = .84. Descriptively, the arm orientation and finger position contributed about equal to pointing perception in the 15° viewpoint. However, the arm orientation dominated interpretations at all other observer viewpoints.

Second, observing the arm from the left behind viewpoint (15°) resulted in more rightward estimates than observing the arm from right behind viewpoint (-15°), as indicated by paired *t*-test averaged over arm direction and finger position, $t(35) = 8.71$, $p < .001$, $d = 1.45$. This corresponds to the results of Experiments 1, 2 and also to the results of Bangerter and Oppenheimer (2006). This relationship decreased and eventually flipped for the more lateral observer viewpoints. A similar trend was also found in Experiments 1 and 2.

3.2.2 Vertical Errors

In the following, we report and discuss additional effects that are significant. Not surprisingly, guesses were higher when the arm pointed upward or when the fingertip held the high position, as indicated by the main effects arm direction and fingertip position. No other effects reached significance. Moreover, it is worth to note that vertical judgements were always very close to the lower (1.07 m) or higher (1.53 m) position, when the heights implied by both cues coincided, regardless of viewpoint and side. This has two important implications. First, it suggests that the vertical processing of both cues is well described by an extrapolation of the eye-finger line or the arm direction in Experiment 3. Second, although other factors than the finger position and arm direction may be involved in pointing perception, such factors have most likely only a small effect.

3.2.3 Horizontal Errors

The (vertical) arm direction descriptively influenced horizontal interpretations stronger for the two most lateral viewpoints (90°/65°), in which upward points resulted in more leftward judgements. The effect of fingertip position was stronger at the rightward observer side, whereby the higher viewpoint led to a greater bias towards the observer than the low viewpoint. The three-way interaction between fingertip position, arm direction, and observer side reached significance. Higher fingertip positions were interpreted as more leftward from the right and more rightward from the left but only for upward points.

3.2.4 Summary

Experiment 3 elaborated on the previous experiments by independently manipulating the heights implied by direction and position cue and thus allowed for a direct comparison of the influence of both cues. The results confirmed the cue-weighting hypothesis. From a sideward viewpoint, mainly the direction of the pointing gesture determined pointing perception. As the observer perspective gets aligned with the pointing arm, the influence of the position cue increases and that of the direction cue decreases. Moreover, position cue and direction cue appear to be the primary determinants of pointing perception.

4 Post-hoc analysis: Effect of View on the Arm on Processing the Fingertip Position

In Experiments 1 and 2, both the observer's perspective and the target region affected pointing perception. Both variables affect the angle under which the observer perceives the pointer's arm. To test whether this angle might be the primary determinant of how pointing gestures are interpreted – as suggested by Exp. 3 and implied by the geometric model –, we reanalysed the data of all experiments. For each experiment and each experimental condition, we computed the absolute angle between the vector defined by the pointer's arm and the vector from the observer's eyes (HMDs in Experiments 1 and 3; estimated head positions in Exp. 2) to the pointer's index finger in the horizontal plane. If the observer is perfectly looking down the pointing arm, this angle is 0°, if the pointer perceives the arm from the side, it is 90°. To quantify the effect of different vertical observer viewpoints independent of the differences in the distances between observer and the pointer's finger, we computed the *relative effect of the viewpoint height on vertical judgements* (see Appendix for details). This variable expresses by how many degrees of angle (measured from the pointer's index finger) the vertical judgements changed when the elevation at which the observer perceived the pointer's index finger in her visual field was increased by 1°. Positive values mean that lower observer viewpoints result in higher judgements. A value of zero would indicate that the vertical judgement is independent of where the pointer's finger is located in the observer's visual field. An absolute value of one would result when observers based their judgement entirely on the index finger position as perceived in their visual fields.

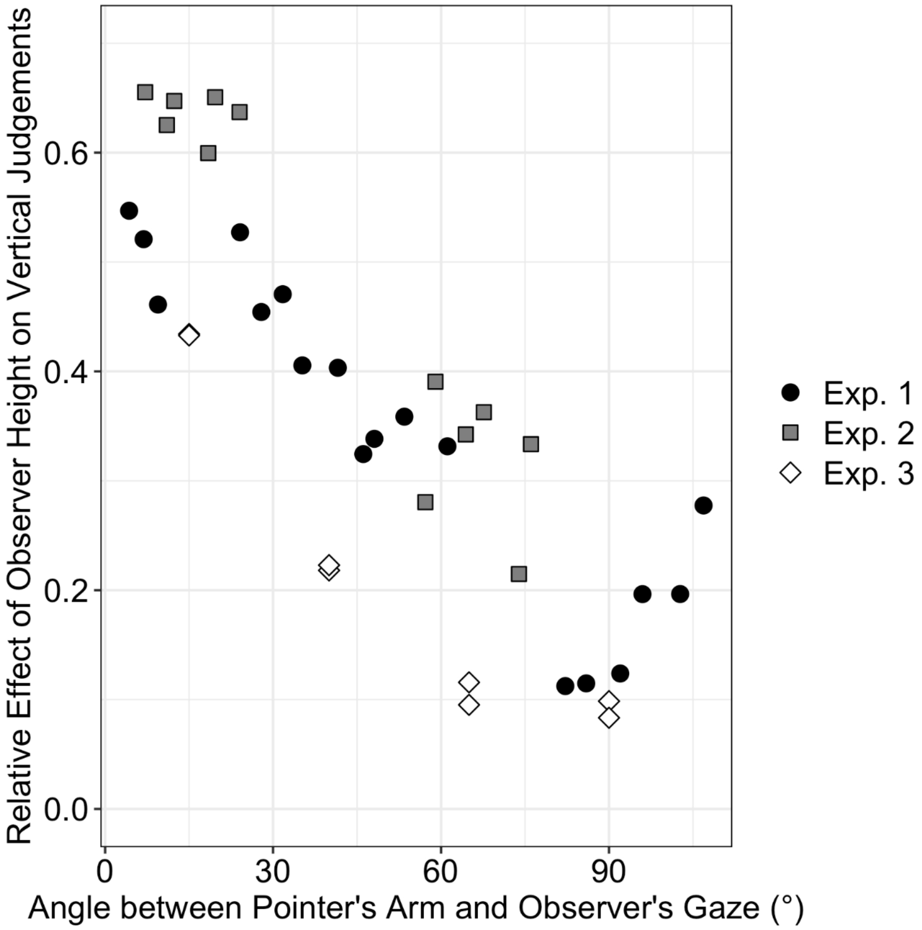
Figure 9 depicts the relationship between both quantities and reveals three important findings. First, the angle under which the observer perceives the pointer's arm appears to be a key factor for determining the relative contribution of the finger position to the interpretation of pointing gestures. Correlations were strong for each experiment (Exp. 1: -.904, Exp. 2: -.955, Exp. 3: -.931) and for the entire data set (r = -.836).

Second, despite their differences, the results of all experiments reveal a fairly consistent pattern. This indicates that the reported effect of the viewpoint height relative to the finger position is not bound to the specific layout (e.g. wall shape, pointed-at locations) or mode of presentation (real vs. VR; realistic virtual pointer vs. arm only).

Third, in Experiments 1 and 2 (but not Experiment 3) the distances between observer and fingertip were typically smallest when the observer was located behind the pointer. This could have inflated the effect of the observer height on vertical judgements at these viewpoints. The above analysis accounted for this by normalising the effect of the observer height on vertical judgements. Nevertheless, it revealed clear relationship between the observer perspective and the effect of the observer height on visual judgements.

Figure 9

Correlational Post-hoc Analysis



Note. The figure plots the relative effect of observer height against the arm view angle. Each symbol marks a combination of observer viewpoint and target region for one of the experiments. A value of zero means no influence of observer height on vertical judgements, while an absolute value of one indicates that interpretations rely completely on the index finger position in the visual field. Positive values indicate that lower viewpoints lead to higher estimates.

5 General Discussion

We tested the cue-weighting hypothesis in three experiments. According to this hypothesis, the position cue (extrapolation of the vector from the observer's eyes to the pointer's fingertip) predominantly determines guesses from viewpoints close to that of the pointer but has little impact on sideward viewpoints. By contrast, the direction cue (extrapolation of the pointer's arm or finger) determines guesses from side viewpoints but has only a minor effect when standing close to the pointer.

It has previously been shown that observers extrapolate the pointer's arm when seeing the pointer from the side and that pointing perception depends on the observer's perspective (Herbort et al., 2020; Mayer et al., 2020). We extended these studies by showing that the cue-weighting hypothesis accounts for these findings and provided a detailed explanation for the specific effects of the observer perspective on systematic errors in the interpretation of pointing gestures. More specifically, Experiment 1 showed that the influence of the position cue increases the more, the closer the observer moves to the pointer in a VR setup. Experiment 2, confirmed this finding in a real world setup. Whereas Experiments 1 and 2 exclusively focussed on the influence of the position cue, the position cue and direction cue were manipulated concurrently in Experiment 3. Besides providing additional support for the cue-weighting hypothesis, this experiment also indicated that pointing interpretation is largely determined by these two cues. Finally, a post-hoc analysis showed that the observer's perspective on the arm affected the influence of the position cue rather consistently in all three experiments despite their differences in presentation format and experimental conditions.

In the following, we discuss the effect of perspective as well as the systematic biases in and the geometry of pointing perception for an accurate understanding of pointing. We relate the current findings to previous studies and, moreover, expand on the limitations of the studies.

5.1 Effect of Observer Perspective in Previous Experiments

The present experiments suggest that differences in the perspectives used in previous studies may be an important reason for those inconsistencies. When our observers watched the pointer from the side, their interpretations resembled those from previous experiments which employed a similar perspective (Herbort & Kunde, 2016; Sousa et al., 2019; Wnuczko & Kennedy, 2011). When the same gestures were observed from viewpoints next to the pointer, which were comparable to those used by Bangerter and Oppenheimer (2006), also the biases resembled those reported by these authors. As vertical biases shrank the more the observer approached the pointer's viewpoint, it can be speculated that they might practically vanish once the observer assumes the pointer perspective, as has been reported by Akkil and Isokoski (2016). Furthermore, the data are descriptively resembling those of Mayer et al. (2020) who also found an perspective dependency of pointing interpretation, e.g. a rather leftward bias especially from rightward viewpoints as well as an overall upward bias.

5.2 Systematic Biases in Pointing Perception

In our experiments, we focussed on systematic biases in pointing perception. This raises the question how such biases compare to unsystematic biases and the question how accurate pointing perception is. Our results suggest that such systematic misunderstandings are a fundamental aspect of pointer-observer misunderstandings. As can be seen by the error bars in Figures 4 and 6, observers significantly overestimated the pointed-at location in each of the 60 different conditions of Experiments 1 and 2. In 52 conditions, the average bias exceeded 10°. Likewise, significant horizontal biases can be found in 47 of the 60 conditions. Absolute horizontal biases were larger than 5° in 40 conditions. These systematic errors set a clear limit on the maximal acuity of pointing perception. Moreover, our experiments show that pointing perception results from the integration of different cues, neither of which coincides with the actual referent (except one exactly assumes the pointer's perspective). Thus, cases in which pointing perception is unbiased should be rather considered as situations, in which competing biases cancel each other out instead of situations, in which observers genuinely process pointing gestures very accurately.

5.3 The Geometry of Pointing Perception

The rationale for our studies was based on a rather geometric notion of the position and direction cue, as becomes evident by the formal model laid out for Experiment 1. Despite the good qualitative fit of the model, we acknowledge that pointing perception cannot be described by such simple geometrical rules alone. For example, observers even fail to linearly extrapolate the vector defined by the arm in a simple two-dimensional scene (Herbort & Kunde, 2016). Much stronger biases can be expected in everyday-life or the present experiments, in which relevant information (e.g. horizontal orientation of the arm, egocentric depth of arms and potential referents) are more difficult to estimate. The 90° condition of Experiment 3 appears to illustrate this case. Here, vertical judgements seem almost entirely determined by the direction cue but horizontal judgements are considerably affected by the observer viewpoint. Moreover, the direction of horizontal biases does not correspond with what would be geometrically predicted based on the position cue. Thus, finger direction and its position should be rather considered as inputs to perceptual processes that are never perfectly accurate. Likewise, how exactly the provided information is integrated and which underlying processes additionally influence the interpretation remains to be examined. Possible further determinants of pointing interpretation are discussed in the next section.

5.4 Limitations and Future Work

We conducted two out of three experiments in VR, thus trading in external validity for control over the pointer, the environment, and partially over the observer's viewpoint, as well as for practical reasons such as a higher rate of data collection. Although perception in VR and reality certainly differs and although our virtual pointing gestures could only approximate real pointing, we are convinced that our main conclusions generalise beyond VR. One reason for this conviction is that Experiment 2 was conducted in a real life setting and replicated the critical pattern of the VR experiment. Furthermore, the results were relative consistent in all three experiments despite their rather different formats, ranging from a real world setup to a highly abstracted VR setup. Hence, our conclusions are fairly independent of the presentation format. Finally, the VR experiments replicated pattern of results comparable to previous experiments that employed human pointers (e.g. Bangerter & Oppenheimer, 2006). That said, minor aspects of pointing interpretation certainly root in our peculiar methods. For example, pointing production, which is strongly guided by the perception of the own pointing gesture (Herbort et al., 2020), differs between VR and the real world (Mayer et al., 2018).

Furthermore, future studies might address possibly variables that lie outside the scope of the presented hypothesis. Although Experiment 3 indicated that pointing interpretation can be mostly explained by the use of the direction and the position cue, observers might base their estimation on additional cues in other situations. Their identification and the weight of their contribution from different viewpoints – even when expected to be rather small – might be addressed in future work.

Moreover, although we formalised the fingertip position and arm direction cue in a specific way, other operationalisations of the cues might be more accurate. For example, although we defined the position cue technically as the extrapolation of the cyclopean eye-finger vector, using the dominant eye as starting point might lead to more accurate, but probably not fundamental different models of pointing interpretation. However, we doubt that the described experiments allow to discern such minor variations in the operationalisation of the visual cues and doing so was not in the scope of the present article.

As discussed above, the processing of especially the direction cue is likely subject to a number of influences. First, the perception of the arm and finger orientation may exhibit central tendencies, as suggested by Experiments 1 and 2 (c.f. Herbort et al., 2020). Second, the extrapolation hinges on the perceived egocentric distance of the finger and the pointed-at regions. These may be especially biased in VR setups (Ng et al., 2016). Third, factors such as the salience (Lücking et al., 2015) or density (Mayer et al., 2020) of potential referents could be expected to affect pointing interpretation. As all of these factors most likely have a nuanced but intricate effect on pointing interpretations, future research needs to combine the manipulation of the factors in question with formal models of the geometry of pointing.

Unlike in our experiments, pointing is typically accompanied by speech. Thus, many pointer-observer misunderstandings could, eventually, be resolved orally. Nevertheless, we argue that pointing perception is an aspect that deserves to be studied in its own right. First, pointing often appears to be an easy and quick way to communicate the identity of a referent and thus might – at some stage of an interaction – be the primary channel of communication. Second, in many situations, one cannot easily resolve pointer-observer misunderstandings verbally, for example when communicating with pre-verbal infants, when the interlocutors do not share a common language or the necessary vocabulary, or when observers are not expected to directly address the pointer. In such situations, knowing the biases of pointing perception might considerably facilitate communication. Finally, a further understanding of biases in pointing perception informs attempts to reduce biases, for example in collaborative virtual environments (Mayer et al., 2020; Sousa et al., 2019).

5.5 Conclusion

To conclude, the current experiments have shown that pointing perception and consequently pointer-observer misunderstandings are based on at least two different aspects of the pointing gesture: the direction of the pointing arm and finger as well as the position of the pointer's fingertip in the observer's visual field. The relative weight of both aspects depends on the angle under which observers look at the pointing arm. The more acute this angle is, the larger is the effect of the finger position and the smaller is the effect of the arm's and finger's direction. This implies that pointing perception depends strongly on the observer perspective. Typically, this results in considerable vertical biases from side perspectives and horizontal biases for perspectives close to that of the pointer. From a more general perspective, our experiments show that the apparently mundane activity of interpreting a pointing gesture, is indeed an intricate – albeit sometimes misguided – perceptual process that integrates information from a variety of sources.

Acknowledgements

We thank Wilfried Kunde for helpful comments and proof-reading, Rainer Scheuchenpflug for statistical support as well as Nina Fuhrmann, Leonhard Höhnel, and Linnéa Hölterhoff for assistance with the data collection. This research was funded by a grant of the German Research Foundation DFG He6710/5-1.

Open Practice Statement

The datasets generated during and/or analysed during the current experiments are available in the OSF repository https://osf.io/bsxwa/.

6 References

Akkil, D., & Isokoski, P. (2016). Accuracy of interpreting pointing gestures in egocentric view. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany*, 262–273. https://doi.org/10.1145/2971648.2971687

Bangerter, A., & Oppenheimer, D. M. (2006). Accuracy in detecting referents of pointing gestures unaccompanied by language. *Gesture, 6*(1), 85-102. https://doi.org/10.1075/gest.6.1.05ban

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

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

El Jamiy, F., & Marsh, R. (2019). Distance Estimation In Virtual Reality And Augmented Reality: A Survey. *Proceedings of the 2019 IEEE International Conference on Electro Information Technology (EIT), Brookings, SD, USA*, 063-068. https://doi.org/10.1109/EIT.2019.8834182

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in cognitive sciences, 8*(4), 162-169. https://doi.org/10.1016/j.tics.2004.02.002

Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *Annals of statistics, 13*(1), 70-84. https://doi.org/10.1214/aos/1176346577

Herbort, O., Krause, L.-M., & Kunde, W. (2020). Perspective determines the production and interpretation of pointing gestures. *Psychonomic Bulletin & Review*, 1-8. https://doi.org/https://doi.org/10.3758/s13423-020-01823-7

Herbort, O., & Kunde, W. (2016). Spatial (mis-)interpretation of pointing gestures to distal referents. *Journal for Experimental Psychology: Human Perception and Performance, 42*(1), 78-89. https://doi.org/10.1037/xhp0000126

Herbort, O., & Kunde, W. (2018). How to point and to interpret pointing gestures? Instructions can reduce pointer–observer misunderstandings. *Psychological Research, 82*(2), 395-406. https://doi.org/10.1007/s00426-016-0824-8

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

Maechler, M. (2016). Diptest: Hartigan's dip test Statistic for unimodality - corrected. *R package version 0.75-7.* https://cran.r-project.org/package=diptest

Mayer, S., Reinhardt, J., Schweigert, R., Jelke, B., Schwind, V., Wolf, K., & Henze, N. (2020). Improving Humans' Ability to Interpret Deictic Gestures in Virtual Reality. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honululu, USA, 26*, 1-14. https://doi.org/10.1145/3313831.3376340

Mayer, S., Schwind, V., Schweigert, R., & Henze, N. (2018). The Effect of Offset Correction and Cursor on Mid-Air Pointing in Real and Virtual Environments. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 24*, 1-13. https://doi.org/10.1145/3173574.3174227

Ng, A. K., Chan, L. K., & Lau, H. Y. (2016). Depth perception in virtual environment: The effects of immersive system and freedom of movement. *Proceedings of the 2016 International Conference on Virtual, Augmented and Mixed Reality, Toronto, Canada*, 173-183. https://doi.org/10.1007/978-3-319-39907-2_17

O'Madagain, C., Kachel, G., & Strickland, B. (2019). The origin of pointing: Evidence for the touch hypothesis. *Science Advances, 5*(7), eaav2558. https://doi.org/10.1126/sciadv.aav2558

Renner, R. S., Velichkovsky, B. M., & Helmert, J. R. (2013). The Perception of Egocentric Distances in Virtual Environments - A Review. *ACM Computing Surveys, 46*(2), 1-40. https://doi.org/10.1145/2543581.2543590

Sousa, M., dos Anjos, R. K., Mendes, D., Billinghurst, M., & Jorge, J. (2019). Warping Deixis: Distorting Gestures to Enhance Collaboration. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK*, 1-12. https://doi.org/10.1145/3290605.3300838

Taylor, J. L., & McCloskey, D. I. (1988). Pointing. *Behavioural Brain Research, 29*(1-2), 1-5. https://doi.org/10.1016/0166-4328(88)90046-0

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*(5), 1485-1491. https://doi.org/10.1037/a0024232

Wnuczko, M., Kennedy, J. M., Niemeier, M., & Singh, K. (2013). Mirror image arm used in monocular, binocular, and blindfolded pointing. *Psychonomic Bulletin & Review, 20*(1), 95-100. https://doi.org/10.3758/s13423-012-0318-7

Wong, N., & Gutwin, C. (2010). Where are you pointing? The accuracy of deictic pointing in CVEs. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, Georgia, USA*, 1029–1038. https://doi.org/10.1145/1753326.1753480

7 Appendix

The *relative effect of the observer height on vertical judgements* was computed as follows for each observer viewpoint in Experiments 1, 2, and 3. First, we computed the effect of the (rel.) observer height on the elevations of the observer's judgements. This was done by computing the difference between angular vertical errors obtained from the lower and higher position of a specific observer viewpoint. Second, we computed the effect of the (rel.) observer heights on the elevations of the index finger from the observer's viewpoint. By dividing the first quantity by the second, we obtained the *relative effect of the observer height on vertical judgements*. For Experiment 3, this approach was not feasible because the relative position of the index finger in the observer's visual field was not manipulated independently of the presented arm postures. Hence, we computed a linear regression of the elevation of vertical judgements on the arm orientation and the elevation of the finger in the observer's visual field for each observer viewpoint. Here, the *relative effect of the observer height on vertical judgements* was defined as the beta value related to the elevation of the finger from the observer's viewpoint.