

Running Head: VIEWPOINT IN THE VISUAL-SPATIAL MODALITY

Viewpoint in the Visual-Spatial Modality:
The Coordination of Spatial Perspective

Jennie E. Pyers^a

Pamela Perniss^b

Karen Emmorey^c

^aWellesley College, Psychology Department, Wellesley, MA 02481, USA

Tel: 781-283-3736, Fax: 781-283-3730, email: jpyers@wellesley.edu

Corresponding Author

^bUniversity of Brighton, School of Humanities

Checkland Building, BN1 9PH Brighton, UK

p.perniss@brighton.ac.uk

^cSan Diego State University

Laboratory for Language and Cognitive Neuroscience, 6495 Alvarado Road, Suite 200, San

Diego, CA 92120

kemmorey@projects.sdsu.edu

Abstract

Sign languages express viewpoint-dependent spatial relations (e.g., left, right) iconically but must conventionalize from whose viewpoint the spatial relation is being described, the signer's or the perceiver's. In Experiment 1, ASL signers and sign-naïve gesturers expressed viewpoint-dependent relations egocentrically, but only signers successfully interpreted the descriptions non-egocentrically, suggesting that viewpoint convergence in the visual modality emerges with language conventionalization. In Experiment 2, we observed that the cost of adopting a non-egocentric viewpoint was greater for producers than for perceivers, suggesting that sign languages have converged on the most cognitively efficient means of expressing left-right spatial relations. We suggest that non-linguistic cognitive factors such as visual perspective-taking and motor embodiment may constrain viewpoint convergence in the visual-spatial modality.

Keywords: Spatial language; viewpoint; sign language; gesture

Viewpoint in the Visual-Spatial Modality: The Coordination of Spatial Perspective

1. INTRODUCTION

Successful communication about spatial relations depends in large part on the speaker and listener converging on a common viewpoint on a spatial scene. Speakers around the world have conventionalized different ways to clearly convey spatial relations. Some languages use primarily an allocentric frame of reference that describes spatial relations using an absolute coordinate system such as North, South, East, and West, e.g., the tent is west of the picnic table (Levinson, 2003). Speakers of other languages (e.g., English and many Indo-European languages) tend to use a relative frame of reference to talk about spatial relations and to describe the location of an object relative to another object, e.g. the tent is to the left of the picnic table. Speakers of languages that use a relative frame of reference, as opposed to an allocentric frame of reference, face a viewpoint coordination problem—left depends on the location of the person who views the picnic table. To overcome this problem, speakers will often add additional information to clarify the viewpoint, e.g. specifying their own viewpoint, as in my left, or their interlocutor's viewpoint, as in your right. A long history of research on spatial language and cognition supports the primacy of the speaker's egocentric viewpoint (e.g., Clark, 1973; Levelt, 1989; Miller & Johnson-Laird, 1976; Piaget & Inhelder, 1956). An egocentric description is directly available because it represents the speaker's perception of a scene and, as such, requires little cognitive effort to produce. Indeed, several studies show that English speakers preferentially describe spatial relations egocentrically, from their own viewpoint (Levelt, 1989; Miller & Johnson-Laird, 1976).

An egocentric description of spatial relations, however, is by no means the default for speakers. Social, cognitive, communicative, and situational factors may lead speakers to take into

1
2
3 account the viewpoint of their addressee when describing spatial relations and to express spatial
4 relations from their addressee's point of view instead of their own, e.g. saying "on your right"
5 instead of "on my left" (Galati, Michael, Mello, Greenauer, & Avraamides 2013; Herrmann,
6 Bürkle, & Nirmaier 1987; Mainwaring, Tversky, Ohgishi, & Schiano, 2003; Schober, 1993,
7 1995; Tversky & Hard, 2009). Depending on the circumstances, speakers in Schober's (1993)
8 study adopted a non-egocentric perspective anywhere from 29% to 81% of the time. For jointly
9 viewed scenes, speakers were more likely to adopt their addressee's viewpoint when their
10 addressee had a different viewpoint on the scene (i.e., offset by either 90° or 180°). In addition,
11 use of the addressee's viewpoint was even higher when the speaker was asked to record a
12 message for non-present addressees, who potentially bore a greater cognitive load because they
13 would be unable to ask for clarification (Schober, 1993). Similarly, speakers of English as well
14 as speakers of Japanese (considered to be an extremely polite culture) were more likely to use a
15 non-egocentric perspective when the addressee was unfamiliar with a spatial array compared to
16 when the addressee was familiar with the array (Mainwaring et al., 2003), suggesting that
17 perspective use does not simply reflect considerations of politeness (Brown & Levinson, 1987).
18 Thus, speakers seem to be sensitive to the addressee's cognitive effort in understanding spatial
19 descriptions and adapt their utterances accordingly to ensure successful communication.

20
21
22 In addition to the addressee's cognitive effort, the speaker's own cognitive effort may
23 influence perspective choice. For example, Galati and Avraamides (2013) found that speakers
24 who sat at an oblique offset of 135° with respect to their present addressee were more likely to
25 give spatial descriptions from an egocentric perspective. In this situation, representing a non-
26 egocentric perspective involved a more difficult spatial computation. When the cognitive effort
27 was minimized (as when speakers sat at 90° or 180° relative to their addressee), speakers were
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 more likely to give descriptions from their addressee's perspective, confirming Schober's (1993)
4
5 findings.
6

7
8 Overall, the preference for an egocentric or non-egocentric perspective in these situations
9
10 may be explained in terms of the relative cognitive burdens placed on speaker and addressee.
11
12 Mainwaring et al. (2003) suggest that the transformation of a spatial scene into a linguistic
13
14 utterance (scene-to-language) or the transformation of the utterance into a spatial image
15
16 (language-to-scene) dictates the relative cognitive burden of viewpoint alignment for speaker and
17
18 addressee. When comprehending a spoken spatial description, the addressee must mentally map
19
20 the language input onto the spatial scene, and this mapping is more difficult if the addressee must
21
22 interpret the spatial relations from a viewpoint different from their own. Mainwaring et al. (2003)
23
24 further suggest that transforming a spoken utterance into a mental image of a spatial situation is
25
26 more difficult than generating language to describe a known spatial scene, regardless of the
27
28 perspective – egocentric or non-egocentric. When speakers adopt the (non-shared) perspective of
29
30 their addressee during their spatial descriptions, they remove the more difficult perspective-
31
32 taking demands from the addressee's mapping task, presumably leading to more rapid
33
34 comprehension and more successful communication. In general, the choice of non-egocentric
35
36 over egocentric spatial descriptions reflects the need for perspective alignment in successful joint
37
38 communication, and the more salient this need is, the more likely it is that the speaker will adopt
39
40 the perspective of their listener.
41
42
43
44
45
46
47

48 To date, research on viewpoint in spatial descriptions has primarily focused on spoken
49
50 rather than signed languages, leaving unanswered the nature of viewpoint convergence in the
51
52 visual-spatial modality. Crucially, the cognitive requirements that constrain the convergence of
53
54 viewpoint in spatial descriptions may vary as a function of language modality. Sign languages
55
56
57
58
59
60

1
2
3 differ from spoken languages in how spatial relations are linguistically marked. Instead of using
4 spatial terms like “right” or “left,” signers of many different and unrelated sign languages
5
6 directly map spatial relations onto the signing space in front of the body. That is, signers create
7
8 an iconic representation in which real-world spatial relationships are mapped onto the relative
9
10 positions of their hands. In essence, the signer’s hands recreate the target spatial relationship in
11
12 signing space. This recreation can be done from the signer’s own perspective of a real or
13
14 imagined scene as illustrated in Figure 1A. To describe a room containing a table on the left, the
15
16 signer can produce a classifier sign for the round table to his left to indicate that the table is on
17
18 the left. As shown in Figure 1B, the signer could also produce the classifier sign for the table on
19
20 his right so that the table is on the left of signing space as viewed by an addressee who is facing
21
22 him (just as you, the reader, are facing the signer in this illustration).
23
24
25
26
27
28

29 (Figure 1 about here)
30
31

32 Regardless of the signer’s choice of viewpoint, the perceiver faces a unique challenge in
33
34 interpreting iconic descriptions of left-right spatial relations. The canonical position of signing
35
36 interlocutors is face-to-face, which means that a signer and addressee view spatial descriptions
37
38 from opposite positions: what a signer articulates on the left side of signing space is perceived by
39
40 the addressee on his right, and vice versa, as illustrated in Figure 1. If signers communicate left-
41
42 right spatial relations from an egocentric viewpoint (e.g., on the left from the signer’s
43
44 perspective) and perceivers interpret them from their own egocentric viewpoint (e.g., on the right
45
46 from the perceiver’s perspective), then the accurate expression and communication of the spatial
47
48 relationship would fail. While the analogical representation of spatial relations seems relatively
49
50 transparent, it is only when viewpoint expression and interpretation are conventionalized that
51
52 signed spatial descriptions can be successful.
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Such coordination of perspectives in sign languages only needs to take place when talking about non-present referents or present referents that are not jointly viewed, i.e. for “non-shared” space (see Figure 2a). For present referents that are visible to both the signer and addressee, the spatial mapping remains the same from any perspective because the hands map directly onto the objects in the shared environment, as illustrated in Figure 2b (Emmorey & Tversky, 2002). In the current study, we are concerned only with spatial descriptions of scenes that are not jointly viewed by the interlocutors, and that thus require perspective coordination. Under these circumstances, successful communication depends on one of the two communication partners taking the other’s perspective: either the signer conveys the left-right spatial relation from the perceiver’s perspective so the perceiver can interpret the description from her own egocentric perspective (Figure 1B), or the signer describes the spatial relation egocentrically and the perceiver must interpret the spatial description non-egocentrically, i.e., from the signer’s perspective (Figures 1A and 2A).

(Figure 2 about here)

Across unrelated sign languages in which spatial relations are depicted iconically by the placement of classifier signs in space (as in Figure 1), signers have been found to produce such descriptions from an egocentric viewpoint (Emmorey, 1996; Emmorey, Klima, & Hickok, 1998; Perniss, 2007; Pyers, Perniss, & Emmorey, 2008). Pyers et al. (2008) asked pairs of signers from eight different sign languages¹ to engage in a picture description task, in which the signers described pictures containing spatial scenes to an addressee, who selected the target picture from an array (the Man and Tree task, Space Stimuli Kit 1.1; Max Planck Institute for Psycholinguistics Field Manual, 1992). Signers across all eight sign languages produced

¹ American Sign Language, Deutsche Gebärdensprache (German Sign Language), Indo-Pakistani Sign Language, Israeli Sign Language, Lingua Italiana dei Segni (Italian Sign Language), Nederlandse Gebarentaal (Sign Language of the Netherlands), Nihon Shuwa (Japanese Sign Language), and Russian Sign Language.

1
2
3 egocentric left-right spatial descriptions (only rarely providing left-right descriptions from their
4 addressee's viewpoint), and addressees overwhelmingly picked the correct picture, interpreting
5 the description from the signer's egocentric viewpoint. These data indicate the conventionalized
6 nature of this viewpoint strategy across many sign languages. The preference for egocentric
7 descriptions across sign languages stands in contrast to speakers' general willingness to adopt a
8 non-egocentric viewpoint to support their addressees' comprehension under similar
9 circumstances. In sign languages, in contrast to spoken languages, the addressee is the one who
10 relinquishes his/her viewpoint.
11
12
13
14
15
16
17
18
19
20
21

22 The consistency observed in viewpoint choice within signers and across distinct sign
23 languages suggests that there are properties of representing spatial relations analogically in the
24 visual-spatial modality that favor representing viewpoint from the signer's perspective, even
25 when the need for perspective alignment is salient. For example, within such an analogical
26 system, the egocentric representation of left-right spatial relations involves a direct mapping of
27 the spatial relation onto the left-right position of the hands, as in Figure 1A, whereas the adoption
28 of the non-egocentric addressee's perspective would require that the signer override his/her
29 mental image of the scene to produce the opposite visual-manual description (e.g., placing the
30 sign for the table on his right in the signing space, even though the table was on his left in the
31 perceived scene, Figure 1B). In contrast, for spoken spatial descriptions there is no potential for a
32 perceptual conflict between the linguistic expression and the observed spatial scene. The signer's
33 task in producing a non-egocentric spatial description may thus be more cognitively demanding
34 than what is required for the non-egocentric production of arbitrary spoken lexical labels.
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

52 Further, the addressee's ability to interpret a spatial description from the signer's perspective
53 may be facilitated by the ability to engage in motor embodiment (Emmorey et al., 1998). That is,
54
55
56
57
58
59
60

1
2
3 sign perception might involve an internal simulation of sign production, as if the sign perceiver
4 were signing herself (Emmorey, 2006). Thus, the relative cognitive burdens placed on the
5
6 producer and perceiver with respect to viewpoint convergence may be differently distributed
7
8
9
10 within an iconic, analog system of spatial representation.
11

12
13 In the current study, we investigated whether the iconically driven property of using
14 space to describe spatial relations constrains viewpoint choice in the visual-spatial modality. In
15 Experiment 1, we directly compared the production of spatial descriptions in ASL by Deaf fluent
16
17 signers to the use of gesture (without speech) by sign-naïve English speakers unaccustomed to
18
19 using the visual-spatial modality as the sole channel of communication. Experiment 1 allowed us
20
21 first to replicate previous findings that ASL signers favor an egocentric expression and non-
22
23 egocentric interpretation of viewpoint-dependent spatial relations. In addition, by directly
24
25 comparing ASL signers to gesturers on a similar task, we could investigate whether it is indeed
26
27 properties of the visual modality that constrain the representation and coordination of spatial
28
29 viewpoint.
30
31
32
33
34
35

36 We hypothesized that the iconic, analogical expression of spatial relations in which the
37
38 relation between objects is recreated by the placement of the hands in space leads to a strong
39
40 preference for an egocentric representation of left-right spatial relations, and as such, both
41
42 signers and gesturers should express spatial relations from their own point of view. We further
43
44 hypothesized that if the pattern of viewpoint alignment observed in sign language is driven by
45
46 properties of the visual-spatial modality, then non-signers should interpret manual expressions of
47
48 spatial relations non-egocentrically. Such a pattern in the gesturers would be quite revealing
49
50 about the effect of modality on viewpoint alignment as these native English speakers would
51
52
53
54
55
56
57
58
59
60

1
2
3 this type of communication task. Moreover, English-speaking addressees accustomed to
4
5 interpreting spatial descriptions from their own viewpoint would instead have to take on the
6
7 demands of perspective-taking themselves and interpret the utterance from a non-egocentric
8
9 viewpoint.
10
11

12 In Experiment 2, we assessed the relative cognitive costs of producing and
13
14 comprehending left-right spatial relations in the visual-spatial modality by randomly assigning
15
16 pairs of gesturers to conditions in which they were instructed to produce and comprehend spatial
17
18 descriptions either egocentrically or non-egocentrically (or they were not assigned a specific
19
20 perspective, as in Experiment 1). We hypothesized that there would be an asymmetry in the
21
22 relative cognitive burdens of production and comprehension of gestured viewpoint-dependent
23
24 spatial descriptions such that there would be a greater egocentric bias for production than for
25
26 comprehension in the visual-spatial modality.
27
28
29
30

31 2. EXPERIMENT 1

32 2.1. Method

33
34 2.1.1. Participants. To replicate the previous research that ASL signers prefer an egocentric
35
36 description of spatial relations, we recruited eight Deaf fluent ASL signers (6 females; $M_{\text{age}}=35.2$
37
38 $SD=9.8$) to participate in a communication task. To test our hypothesis that the preference for
39
40 egocentric description exhibited in sign languages is inherent to communication of spatial
41
42 relations in the visual-spatial modality, we also recruited eighteen female undergraduate non-
43
44 signers ($M_{\text{age}}=19.6$, $SD=1.2$) to participate in the same communication task. The Deaf
45
46 participants were all paid for their participation; the non-signers were either paid or received
47
48 course credit for their participation. All participants were tested in pairs, with each randomly
49
50 assigned the role of Producer or Perceiver.
51
52
53
54
55
56
57
58
59
60

1
2
3 2.1.2. Materials and Procedure. Pairs were seated across from each other separated by an
4
5 eleven-inch divider, but always in view of each other. We provided the Producer with a binder of
6
7 16 pictures including 8 pictures that depicted a man standing to the left or right of a tree, and 8
8
9 filler items depicting a man in front of or behind a tree (Figure 3).² The set of pictures we used
10
11 were adapted from The Man and Tree task, (Space Stimuli Kit 1.1; Max Planck Institute for
12
13 Psycholinguistics Field Manual, 1992) by Senghas (2010a). Producers viewed each picture one
14
15 at a time in one of two fixed random orders; each picture was separated by a blank piece of
16
17 paper. The Perceivers' pictures were laid out on the table in a random array. The divider between
18
19 participant pairs prevented the Producer and Perceiver from jointly viewing the spatial scene
20
21 being described (i.e., the stimulus picture). In contrast to previous studies with spoken languages,
22
23 this kind of shielded communication task was necessary because when signers and perceivers
24
25 jointly view a scene, they use shared space in their signed descriptions (see Figure 2b). For
26
27 spoken languages, the opposite is the case: jointly-viewed, real object arrays elicit viewpoint
28
29 marking, but shielded communication tasks using photographs of object arrays do not.

30
31
32
33
34
35
36 (Figure 3 about here)

37
38 ASL signers completed the task in ASL, and non-signers completed the task using only
39
40 two gestures that they were taught, one to represent a man, and one to represent a tree (Figure 4).
41
42 Because our research question focused on the representation of left-right spatial relations,
43
44 participants were neither explicitly instructed to represent the orientation of the man, nor told
45
46 which side of the "man" handshape represented the front or back of the body.³ All Producers
47
48
49
50
51
52

53
54 ² Although front/back spatial relations require perspective alignment, we focus only on left/right relations here. In
55
56 contrast to left/right, front/back specifications are less conventionalized in ASL: they are less consistently marked
57
58 and less consistently associated with an egocentric perspective (Emmorey 1996, 2001).

59
60 ³ The handshape used to convey the man (see Figure 4a) does not have a strongly conventionalized front or back
intrinsic side in ASL. In some cases, the back of the hand can represent the front of the body, as when the handshape

Viewpoint in the Visual-Spatial Modality 12

1
2
3 were instructed to look at the picture in front of them, remember the configuration, turn the page,
4
5 and produce a signed or gestured description of the picture that they viewed. Gesture Producers
6
7 were instructed not to speak during the task. Perceivers were instructed to select the picture that
8
9 best matched the Producer's description, show it to the experimenter (keeping it out of the
10
11 Producer's view), and then return it to the array. They could ask the Producer to repeat their
12
13 spatial description. They were additionally told that they could select the same picture for more
14
15 than one item; this instruction kept Perceivers' selections from being influenced by the process of
16
17 elimination. Because we were interested in the patterns of viewpoint expression and
18
19 comprehension in the visual modality, we allowed participants to take their time with their
20
21 productions and selections in order for them to best convey the information in the picture and to
22
23 make the most accurate selection that matched the gestural configuration. Participants began
24
25 with three practice items depicting two trees or two men next to each other, and then proceeded
26
27 to the test items. The three practice items were symmetrical and did not require viewpoint
28
29 marking. Although both ASL signers and non-signers completed the practice items, the practice
30
31 items were present primarily to offer the non-signers an opportunity to practice producing the
32
33 newly learned gestures. All participants were videotaped for coding offline.

40
41 (Figure 4 about here)

42
43 2.1.3. Coding. The eight items that required viewpoint marking to distinguish left-right spatial
44
45 relations were coded. We coded the accuracy of the Perceivers' selections by scoring as
46
47 successful any selection that correctly mapped the left-right relation. Thus, although Producers
48
49 may have encoded the orientation of the man with respect to the tree by orienting the hand in a
50
51 corresponding way, our coding focused only on the left-right mapping.
52
53

54
55
56 is used in the sign WALK, but in other cases the back of the hand represents the back of the body, as when the
57
58 handshape is used in LIE-DOWN.
59
60

1
2
3 We coded whether Producers expressed the spatial relation from their own viewpoint or
4 from the Perceiver's viewpoint (Figure 5). For example, in viewing the picture of the man
5 standing to the left of the tree, the signer could directly map the locations of the signs for the tree
6 and the man onto the same left-right locations in signing space, matching his/her view of the
7 picture. Such a strategy was classified as the Producer using his/her own "Producer's
8 Perspective." Alternatively, the Producer could reverse the locations of the signs and describe the
9 spatial relation from the Perceiver's perspective, such that the signs would directly map to the
10 pictured objects from the Perceiver's viewpoint. This strategy was coded as "Perceiver's
11 Perspective." Every description produced by the Producer was coded as an individual utterance.
12 For example, the Producer may have repeated a description of a single picture more than once if
13 the Perceiver asked for clarification. If the signer repeated a description and adopted a different
14 viewpoint strategy, one of their descriptions would be coded as "Producer's Perspective" and the
15 second as "Perceiver's Perspective." Thus, for each Producer, the percentage of spatial
16 descriptions produced using each viewpoint was computed.

17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36 Similarly, we coded whether Perceivers interpreted the description of the spatial relation
37 from their own viewpoint or from the Producer's viewpoint. For example, the Producer could
38 express a spatial relation by placing the sign for 'man' on his/her left side of signing space, and
39 the sign for 'tree' on his/her right side of signing space. If the Perceiver interpreted this
40 relationship as the man standing to the left of the tree, as evidenced by the Perceiver's picture
41 selection, the interpretation was coded as being from the "Producer's Perspective." If, on the
42 other hand, the Perceiver interpreted the description as displaying a man standing to the right of
43 the tree, it was coded as "Perceiver's Perspective", as this mapped onto the Perceiver's own view
44 of the signs.

(Figure 5 about here)

2.2. Results

All four of the ASL Producers used their own viewpoint 100% of the time, and the ASL Perceivers readily adopted the Producer's Perspective 91% of the time (See Table 1); this viewpoint compatibility corresponded to successful selections on 91% of trials. The small amount of variability in the Perceivers' performance reflected the fact that three of the four Perceivers each selected a picture representing the Perceiver's Perspective once.

(Insert Table 1 here)

Like signers, gesture Producers conveyed the spatial relations predominantly from their own point of view, adopting an egocentric perspective at a rate that did not differ significantly from the ASL signers ($t(11)=1.25$, $p=.24$; see Table 1). Gesturers were quite variable as a group: five Producers expressed egocentric viewpoints exclusively, two did so more than 75% of the time, while the remaining two participants adopted the non-egocentric Producer's Perspective more than 75% of the time, although neither did so exclusively. Thus, while there was no statistical difference between the ASL and gesture groups, there was a qualitative difference in consistency, albeit with a strong preference for the egocentric viewpoint. Unlike the ASL Perceivers, however, the gesture Perceivers rarely interpreted the Producers' gestures as reflecting the Producers' perspective; only four of the nine participants adopted the non-egocentric perspective, and each did so for only one trial. Gesture Perceivers chose their own egocentric viewpoint for interpretation on the majority of trials and significantly more often than what was observed for the ASL signers ($t(11)=2.47$, $p<.0001$; see Table 1). This pattern in the gesturers resulted in a viewpoint incompatibility – with neither participant relinquishing their egocentric perspective – in the majority of trials, leading to only 17 correct matches out of 72 in

1
2
3 the entire sample for the eight target items. The viewpoint compatibility necessary for correct
4
5 matches and successful communication was thus only infrequently achieved. The majority of
6
7 successful matches occurred when Producers communicated the spatial relation from the
8
9 Perceiver's Perspective, taking a non-egocentric perspective, while the Perceiver maintained an
10
11 egocentric perspective in interpretation (see Figure 6).
12

13
14
15 (Insert Figure 6 about here)
16

17 18 2.3. Discussion

19
20 In Experiment 1, even with a small sample size, we robustly replicated previous work on
21
22 sign languages demonstrating that signers prefer to express and interpret viewpoint-dependent
23
24 spatial relations from the signer's egocentric perspective (Emmorey & Tversky, 2002; Pyers et
25
26 al., 2008). In addition, we found that sign-naïve gesturers expressed viewpoint-dependent spatial
27
28 relations from an egocentric perspective, indicating the immediate availability of this
29
30 perspective. In contrast to English speakers, who under a similar face-to-face paired
31
32 communication task readily adopt a non-egocentric perspective to convey viewpoint-dependent
33
34 spatial relations for their addressees (Schober, 1993), both ASL signers and sign-naïve gesturers
35
36 almost exclusively expressed spatial relations from their own viewpoint, rarely adopting the
37
38 perspective of their addressee. Further, this strategy was understood by ASL Perceivers (but not
39
40 by gesture Perceivers), who predominantly adopted a non-egocentric perspective to correctly
41
42 interpret the descriptions of viewpoint-dependent spatial relations. Although this paradigm did
43
44 not elicit non-egocentric productions on the part of signers, there may be other discourse
45
46 situations, as has been observed with speakers, that might elicit non-egocentric productions. In
47
48 this face-to-face communication task, however, we clearly see the conventionalization of
49
50 viewpoint in ASL spatial descriptions.
51
52
53
54
55
56
57
58
59
60

Viewpoint in the Visual-Spatial Modality 16

1
2
3 We hypothesized that the consistent pattern of viewpoint representation and interpretation
4 that has been observed across multiple sign languages may emerge because of the iconic, analog
5 properties of the visual-spatial modality. We found that although sign-naïve gesturers
6 egocentrically expressed left-right spatial relations in the visual-spatial modality, their
7 communication partners did not consistently interpret the spatial expression as from the
8 Producer's point of view. That the gesturers did not exhibit the same degree of viewpoint
9 alignment observed for signers suggests that viewpoint conventionalization is a necessary
10 process for visual-spatial languages, and one that must emerge over time.
11
12
13
14
15
16
17
18
19
20
21

22 Evidence from an emerging sign language, Nicaraguan Sign Language (NSL), provides
23 support for this idea. The current form of NSL does not yet exhibit a systematic means of
24 representing and interpreting spatial viewpoint (Pyers, Shusterman, Senghas, Spelke, &
25 Emmorey, 2010; Senghas, 2010b). NSL signers sometimes convey the viewpoint from their own
26 perspective and sometimes from their addressee's perspective. The variable performance of the
27 Nicaraguan signers shows that conventionalization of viewpoint does not emerge early in the
28 development of a sign language. However, one reason for the late emergence of viewpoint
29 conventionalization may be related to the cognitive abilities of this population. The same signers
30 who inconsistently represent viewpoint also show delays on theory-of-mind tasks that require
31 understanding that other people have different perspectives (Pyers & Senghas, 2009). Thus, the
32 lack of conventionalization observed in the language may be related to a poor understanding of
33 what other people know (and don't know) and of the kinds of information that are relevant for
34 successful communication. These pragmatic features related to theory-of-mind development
35 could drive conventionalization in a language. The gesturers in our study, on the other hand,
36 were typically developing undergraduates with presumably mature theory-of-mind abilities, but
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 with limited experience communicating solely in the visual-spatial modality. As such, the
4
5
6 gesturers were not experienced enough with the visual-spatial modality to understand the need
7
8 for perspective convergence.
9

10 That sign languages need to conventionalize perspective for the accurate communication
11
12 of viewpoint-dependent spatial relations does not resolve the question of why sign languages
13
14 seem to converge on a pattern of expressing and interpreting viewpoint from the signer's rather
15
16 than the addressee's perspective, as is observed in spoken languages. In Experiment 2, we tested
17
18 our hypothesis that the pattern of viewpoint convergence in sign languages is a direct result of
19
20 the asymmetry in cognitive cost for the signer and the addressee when adopting a non-egocentric
21
22 perspective.
23
24
25

26
27 For spoken language, adopting a non-egocentric perspective incurs a cognitive cost for
28
29 the speaker and likely also for the addressee. It is unknown, however, whether this cognitive cost
30
31 is greater for the speaker compared to the addressee. We hypothesized that in the visual modality
32
33 the cognitive cost for the signer is greater than for the addressee because of the iconic nature of
34
35 spatial language. As noted in the introduction, for the signer to adopt the addressee's perspective,
36
37 he or she must generate a spatial description that is in direct perceptual (visual and
38
39 somatosensory) conflict with the spatial scene. We suggest that this perceptual conflict presents a
40
41 relatively greater cognitive load for the signer compared with the addressee, whose cognitive
42
43 load may be reduced by the ability to internally simulate the signer's manual productions (i.e.,
44
45 motor embodiment; see Pickering and Garrod, 2013). Thus, in Experiment 2, we predicted that
46
47 the cognitive cost to adopt a non-egocentric perspective would be greater for the gesture
48
49 Producer than for the Perceiver.
50
51
52
53
54

54
55
56
57
58
59
60

were typically developing undergraduates with presumably mature theory-of-mind abilities, but

3. EXPERIMENT 2

3.1. Method

3.1.1. Participants. We randomly assigned 94 female sign-naïve undergraduates ($M_{age}=19.6$ years; range, 18-23 years) to one of three conditions: Unassigned Perspective ($N=22$), Producer's Perspective ($N=34$), or Perceiver's Perspective ($N=38$).⁴ All participants were tested in pairs, and within each pair, participants were randomly assigned the role of Producer or of Perceiver.

3.1.2. Materials and Procedures. We followed the same procedure as in Experiment 1, but the stimuli included only objects that had no intrinsic front/back features (a cup, a piece of paper, a chopstick, and a ball). The 12 test items depicted two of the four objects next to each other (Figure 7). The pictures were presented to the Producer in a binder in one of three random orders. The Perceiver's pictures were laid out in a fixed array that was consistent across all subjects (Figure 7), and the Producer viewed the Perceiver's array of pictures to better understand the Perceiver's possible choices. Producers and Perceivers were taught one-handed gestures for the objects (Figure 8), and they practiced producing and perceiving the gestures during four practice items that depicted the individual objects.

(Figure 7 about here)

(Figure 8 about here)

In the Unassigned Perspective condition, Producers and Perceivers received the same instructions as participants in Experiment 1. In the Producer's Perspective condition, participants were explicitly told to gesture and interpret the spatial relation from the Producer's perspective; with these instructions, the Producer was required to use an egocentric perspective while the

⁴ Due to experimenter error, one pair that had been pre-assigned to the Producer's Perspective condition actually received instructions for the Perceiver's Perspective condition. Thus, instead of having an equivalent number in each

1
2
3 Perceiver adopted the non-egocentric perspective of the Producer. In the Perceiver's Perspective
4
5 condition, participants were told to gesture and to interpret the spatial relation from the
6
7 perspective of the Perceiver. Here, the Producer relinquished her own perspective and produced
8
9 a gesture configuration that was the opposite of what she saw, but that aligned with the
10
11 Perceiver's view of the picture. Participants in the assigned perspective conditions were
12
13 instructed to produce the gestures and to make their picture selections as fast as they could. All
14
15 participants were videotaped, and the videotapes were digitized at 30fps for coding offline.

16
17
18
19
20 3.1.3. Coding. We coded the expression and interpretation of viewpoint in the same way as
21
22 Experiment 1, and we determined the accuracy of the viewpoint expression for the two
23
24 experimental conditions by identifying whether the participant correctly produced or interpreted
25
26 the gesture using the assigned viewpoint. For the experimental conditions where perspectives
27
28 were assigned, we also recorded how quickly the Producer made the correct gesture; we
29
30 measured the duration from the time that they turned the stimulus page to the onset of the correct
31
32 gestural configuration. We recorded how long it took the Perceiver to select the correct picture
33
34 from the array by recording the duration of time that spanned from the onset of the Producer's
35
36 gestural configuration to when the Perceiver touched the correct picture in the array. We used the
37
38 time of gesture onset rather than offset because the gestures representing the objects and their
39
40 relative positions could be readily perceived well before the offset of the gestural configuration.
41
42 In addition, we observed some perceivers making their selections before the offset of the gesture.
43
44 We excluded long reaction times by trimming the timing data to $\pm 2SD$ from the mean time to
45
46 produce the gesture or to $\pm 2SD$ from the mean time to select the corresponding picture observed
47
48 across all three conditions. This trimming reduced the number of correct items included in the
49
50 Producer's timing analysis by 19 leaving a total of 370 correct items produced by 34 participants
51
52 condition, there were 17 pairs in the Perceiver's Perspective condition and 19 pairs in the Producer's Perspective
53
54 condition.

1
2
3 across both experimental conditions for the Producer's timing analysis. For the Perceivers, 8
4
5 correct items were trimmed, leaving a total of 351 correct items selected by 34 participants
6
7
8 across both experimental conditions for the Perceiver's timing analysis. All items were included
9
10 in analyses of viewpoint accuracy, and all items where the Producer conveyed the correct
11
12 viewpoint were included in our measure of the Perceiver's picture selection accuracy.
13

14 15 3.2. Results 16

17
18 In the Unassigned Perspective condition we found an egocentric bias on the part of both
19
20 Producers and Perceivers, just as we observed in Experiment 1. The Producers conveyed the
21
22 spatial relations predominantly from their own point of view, with eight of the eleven Producers
23
24 doing so for more than 75% of the trials (see Table 1). The Perceivers interpreted the gestures
25
26 from their own egocentric viewpoint for the majority of trials, leading to only 57 out of 132
27
28 successful picture selections (see Table 1). Here, the majority of the successful picture matches
29
30 (n=33) were made by three Perceivers (out of 11) who spontaneously adopted the (non-
31
32 egocentric) perspective of the Producer for more than 75% of the trials.
33
34

35
36 We then compared performance in the Producer's Perspective condition to that in the
37
38 Perceiver's Perspective condition. As shown in Table 2, Producers were significantly more
39
40 accurate in their viewpoint production when they communicated the spatial relation from their
41
42 own perspective than from the Perceiver's perspective. One Producer failed to adopt the
43
44 Perceiver's perspective on all trials in that condition, but no such global perspective failures were
45
46 observed in the egocentric Producer's perspective condition. The difference in the Producers'
47
48
49
50
51
52
53
54
55

57
58 of freedom.
59
60

1
2
3 accuracy between conditions remained significant even when we eliminated the one Producer
4 who failed to adopt the Perceiver's perspective when instructed, $t(17.39)=2.21$, $p=.04$.⁵
5
6
7

8 The advantage in the Producer's Perspective condition was also reflected in the timing of
9 correct gesture production in the first trial, but not in the average gesture time across all trials
10 (see Table 2): Producers who were required to gesture from a non-egocentric perspective were
11 initially slower to produce their gestures than those allowed to communicate from an egocentric
12 perspective, but they eventually adapted to the task such that the cost was no longer reflected in
13 the aggregate timing across all trials.
14
15
16
17
18
19
20
21

22 For Perceivers, there was no significant difference in viewpoint accuracy between
23 perspective conditions (see Table 2), and one Perceiver in each condition globally failed to adopt
24 the correct viewpoint for all trials. There was no significant difference between groups in the
25 number of correct picture selections made by Perceivers when the Producer conveyed the correct
26 perspective (Producer's Perspective: $M=9.58$, $SD=4.30$; Perceiver's Perspective: $M=8.63$,
27 $SD=4.89$), $t(34)=.62$, $p=.53$. We also observed no significant difference in how quickly
28 Perceivers made a selection when they were required to comprehend descriptions from the
29 Producer's perspective compared to their own perspective either on the first trial alone or across
30 all 12 trials (see Table 2).
31
32
33
34
35
36
37
38
39
40
41
42

43 (Insert Table 2 here)

44 4. GENERAL DISCUSSION

45
46
47
48 Piaget and Inhelder (1956) were among the first to argue that perspective taking is
49 challenging, and that a hallmark of mature cognition is the ability to take the perspective of
50 another person. Yet several studies have shown that even neurotypical adults sometimes struggle
51
52
53
54
55
56
57
58
59
60

⁵ When a Levene's test detects unequal sample variances, we report the corrected t-test, including fractional degrees

of freedom.

1
2
3 with the additional mental computation required to take another perspective into account, making
4 their online use of perspective-taking sometimes unreliable (Keysar, Barr, Balin, & Brauner,
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

with the additional mental computation required to take another perspective into account, making their online use of perspective-taking sometimes unreliable (Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003). Successful communication and comprehension of viewpoint-dependent spatial relations, however, requires that one communication partner take on the additional cognitive load of adopting a non-egocentric perspective.

Viewpoint alignment does not occur spontaneously when communicating in the visual-spatial modality. In Experiment 1 and in the Unassigned Perspective condition of Experiment 2, gesturers who had limited experience with communicating in the visual-spatial modality did not automatically align their perspective to successfully describe viewpoint-dependent spatial relations via gesture. Although speakers have experience communicating in the visual spatial-modality using co-speech gestures, they may not be driven to perspective convergence in this modality because they can rely on information transmitted through speech. Little is known about viewpoint encoding and interpretation in co-speech gesture; however, Gullberg and Kita (2009: 258) mention in a footnote that addressees interpret directional information present in co-speech gesture egocentrically, just as we have observed in our current study. The degree to which speakers align viewpoint for spatial information communicated in co-speech gesture is an open question. When gesture is not accompanied by speech, however, viewpoint alignment is not automatic.

The lack of viewpoint alignment observed in the sign-naïve gesturers indicates that over time, viewpoint alignment may need to emerge in visual-spatial communication systems such as sign languages. However, it remains unclear what shapes the pattern of viewpoint alignment in sign languages. For example, only signers of NSL who exhibit a mature understanding of others' perspectives seem to have converged on a systematic way of expressing left-right relations

1
2
3 (Pyers et al., 2010). The variability in the gesturers' performance in Experiment 1 and the
4
5 Unassigned Perspective condition of Experiment 2 suggest that there may be individual
6
7 differences in perspective-taking skill. In contrast, the homogeneity of the signing participants
8
9 suggests that individual variability does not affect processes that have become conventionalized
10
11 in the language, although see Quinto-Pozos, Singleton, Hauser, Levine, Garberoglio, and Hou
12
13 (2013) for a discussion of how cognitive impairments in perspective-taking can negatively
14
15 impact viewpoint understanding in signers. The current study cannot speak to the amount and
16
17 kind of communicative experience in the visual-spatial modality that cognitively mature adults
18
19 need in order to converge on a system for successful communication about viewpoint-dependent
20
21 spatial relations. Perhaps if sign-naïve gesturers were allowed to communicate during the task,
22
23 we might see rapid alignment because such interactions could highlight what aspects of the task
24
25 are ambiguous.
26
27
28
29
30

31
32 The results from the ASL signers in Experiment 1 confirm what has been observed across
33
34 many sign languages: the preferred pattern of viewpoint representation and interpretation is for
35
36 the signer to maintain his/her own perspective and the interlocutor to adopt the signer's
37
38 perspective. The consistency in viewpoint representation and interpretation suggests that this
39
40 pattern may be driven by the iconic properties of the visual-spatial modality. Whereas English
41
42 speakers may readily adopt another viewpoint on a spatial scene through a shift in possessive
43
44 pronoun and change in spatial locative (my left to your right), such a viewpoint shift may be
45
46 significantly more difficult in the visual modality. The results from Experiment 2 show that
47
48 Producers provide more accurate spatial descriptions when instructed to convey the relation from
49
50 their own point of view rather than from their interlocutor's point of view. Perceivers interpreted
51
52 the viewpoint equally well when instructed to maintain an egocentric point of view or to adopt a
53
54
55
56
57
58
59
60

Viewpoint in the Visual-Spatial Modality 24

1
2
3 non-egocentric point of view. Thus, the cost of adopting a non-egocentric viewpoint seems to be
4
5 greater for Producers than for Perceivers, and as such sign languages appear to have converged
6
7 on the most cognitively efficient means of expressing left-right spatial relations. At first,
8
9 Producers were slower to initiate the gestured description from the Perceiver's perspective
10
11 relative to their own perspective, but over multiple trials, this difference in response time
12
13 between the two perspective conditions disappeared. The lack of overall difference in response
14
15 time may have occurred because we used an offline task and gross measurements of time.
16
17 Assessing processing time with more precise measures may provide clearer evidence for the on-
18
19 line cognitive costs associated with adopting a non-egocentric versus an egocentric viewpoint in
20
21 the visual-spatial modality.
22
23
24
25
26

27 The current study cannot identify the precise nature of the cognitive challenge, but we
28
29 speculate that for Producers, taking a non-egocentric perspective in the visual-spatial modality
30
31 may make additional demands on the executive function system, specifically conflict control,
32
33 that do not arise for spoken language users. Conflict control involves inhibiting a prepotent
34
35 response and performing an opposite response (e.g. saying “left” when looking at an arrow
36
37 pointing right, or “night” when looking at a picture of the sun). Although children start to exhibit
38
39 this ability at the end of the preschool years (Carlson, Mandell, & Williams, 2004), adults still
40
41 find more complex versions of conflict control tasks cognitively taxing, as evidenced, for
42
43 example, in the Stroop effect (Stroop, 1935) and, related to location, in the spatial Stroop and
44
45 Simon effects (e.g. Lu & Proctor, 1995; MacLeod, 1991 for reviews). We speculate that to take
46
47 on the Perceiver's perspective, signers face a conflict inhibitory control task that requires them to
48
49 inhibit their own view or mental image of the left-right relation in a scene and produce the
50
51 opposite, but nevertheless iconic, relation in signing space—a task that could be made even more
52
53
54
55
56
57
58
59
60

1
2
3 difficult by the visual and somatosensory feedback they receive from their own signing. Because
4
5 spoken languages do not represent spatial relations in an iconic, analog way, speakers do not face
6
7 the same conflict between their image of a spatial scene and the vocal linguistic utterance that
8
9 they produce. Critically, in contrast to signers, there is no conflicting visual or relevant
10
11 somatosensory feedback. Without a direct conflict between the spatial image, perceived or
12
13 imagined, and the linguistic utterance, speakers may find it much easier than signers to adopt the
14
15 addressee's perspective, as is evidenced in the relatively higher rates of non-egocentric
16
17 perspective use by English speakers in similar face-to-face communication tasks (e.g. Schober,
18
19 1993). Indeed, if sign languages relied primarily on lexical, non-analogical linguistic
20
21 representations of space, we might predict a pattern of viewpoint alignment that more closely
22
23 resembles what has been observed with English speakers, with Producers taking on the
24
25 responsibility of representing spatial relations non-egocentrically. In terms of the relative
26
27 cognitive burdens placed on communication partners, Mainwaring et al.'s (2003) idea that
28
29 transforming a spatial scene into language is easier than transforming language into a spatial
30
31 scene may hold primarily for spoken language viewpoint-dependent spatial descriptions. In the
32
33 visual modality, transforming a perceived spatial scene into language for a communication
34
35 partner may incur an additional cognitive cost that is not present for the spoken modality.
36
37
38
39
40
41
42

43 The Perceivers in Experiment 2 showed no greater cost when adopting the perspective of
44
45 the Producer than when they were allowed to maintain their own egocentric viewpoint. We
46
47 speculate that the Perceiver may recruit non-linguistic strategies to help them adopt the signer's
48
49 perspective. Visual perspective-taking (i.e., putting yourself in someone else's shoes) is an easier
50
51 task than mentally rotating objects in space (Amorim, Isableu, & Jarraya, 2006; Wraga, Creem,
52
53 & Proffitt, 1999). Motor embodiment – the ability to imagine our bodies in another location in
54
55
56
57
58
59
60

1
2
3 order to envision the view from that vantage point – seems to underpin spatial perspective-taking
4
5 in non-linguistic tasks (Kessler & Thomson, 2010). Further, humans seem to rely on mentally
6
7 transforming their own bodies to interpret information about another body (see Wilson, 2001, for
8
9 discussion). In the case of sign languages, the perceivers' perspective-taking is likely facilitated
10
11 by the fact that they can use motor simulation to mentally imitate the signer's manual
12
13 productions, an argument that has been similarly made with respect to gesture comprehension
14
15 (Alibali & Hostetter, 2010). By mapping their own body onto that of the signer, addressees could
16
17 generate the correct viewpoint to accurately interpret the spatial relationship. However, this kind
18
19 of embodiment is not available to sign producers when expressing a non-egocentric perspective;
20
21 both looking at the spatial relation represented in the photograph from one's own perspective or
22
23 imagining how a communication partner would perceive the same relation in the corresponding
24
25 photograph would lead to the production of identical linguistic forms. Thus, internal simulations
26
27 may only benefit sign language perceivers.
28
29
30
31
32
33

34 In essence, we suggest that analogical mapping of spatial relations in sign languages
35
36 places different cognitive challenges on signers and addressees compared to the cognitive
37
38 challenges faced by users of a spoken language where viewpoint-dependent spatial relations are
39
40 primarily expressed with arbitrary words. The perceptual conflict encountered by the signer in
41
42 producing a non-egocentric spatial description presents a clear cognitive load compared to the
43
44 immediately available and non-conflicting egocentric perspective. There appears to be a
45
46 discernible benefit to the signer in maintaining an egocentric perspective in spatial description.
47
48 The cognitive burden on the signer could alone explain the cross-linguistic similarity in
49
50 viewpoint alignment observed across different sign languages. For the perceiver, there seems to
51
52 be no significant benefit to maintaining an egocentric perspective in interpretation. Therefore, the
53
54
55
56
57
58
59
60

1
2
3 perceiver may more easily relinquish his/her egocentric perspective and can perhaps leverage
4
5 non-linguistic cognitive tools such as motor embodiment to adopt a non-egocentric perspective.
6
7

8 This asymmetry in cognitive load for signer and addressee has implications for models of
9
10 language processing with respect to whether and how audience design fits into the processing
11
12 model. One suggestion is that the addressee's perspective and knowledge are already taken into
13
14 account during the stage of message formulation, following a "principle of optimal design" from
15
16 the outset (Clark, 1992; Clark & Murphy, 1982; Horton & Keysar, 1996). Another proposal
17
18 holds instead that message planning is egocentric from the outset, and that the addressee's
19
20 perspective is taken into account as needed, with monitoring for the needs of the addressee
21
22 happening at later stages of production (Dell & Brown, 1991). An egocentric default has also
23
24 been posited for dual-process models of perspective taking, which assume that both production
25
26 and comprehension are initially egocentric, even when interlocutors are aware that successful
27
28 communication requires taking the other's perspective into consideration (Epley, Keysar, van
29
30 Boven, & Gilovich, 2004; Gilbert & Gill, 2000). Overcoming the egocentric bias to adjust to
31
32 another perspective is a subsequent processing step that requires effort and practice (Keysar,
33
34 Barr, & Horton, 1998). The difference that we observe between spoken and signed language
35
36 modalities with respect to optimal viewpoint conventionalization (with spoken languages often
37
38 adopting an addressee-viewpoint strategy and signed languages adopting an egocentric viewpoint
39
40 strategy) suggests that the extent to which audience design can be observed in spatial
41
42 perspective-taking tasks is dependent on the properties of the language modality in which the
43
44 spatial descriptions are given. To better understand where and how audience design affects
45
46 message planning requires measures that account for the relative cognitive burden of production
47
48 and comprehension during communication, as well as the relative costs of unsuccessful
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 communication (Keysar, Barr, Balin, & Brauner, 2000). The results of our study suggest that
4
5 properties of the modality may modulate the distribution of these costs.
6
7

8 In sum, viewpoint alignment in the visual-spatial modality is necessary, but it does not
9
10 come for free and is not automatic. Our results indicate a greater cost to accuracy when sign
11
12 language producers relinquish their viewpoint, and the results support the hypothesis that
13
14 viewpoint alignment must emerge with communicative experience over time. The nature of the
15
16 visual-spatial modality likely constrains the pattern of viewpoint convergence, allowing signers,
17
18 but not addressees, to remain egocentric.
19
20

21 22 ACKNOWLEDGEMENTS 23

24 We thank Alexandra del Solar, Jessica Gonzalez, Annemarie Kocab, Jenny Lu, Rachel Magid,
25
26 Cindy O'Grady, Jennifer Petrich, Andrea Takahesu Tabori, Sarah Wylly, and Kate Zmich for
27
28 research assistance. We also thank Ann Senghas for the stimuli used in Experiment 1. This
29
30 research was supported by a postdoctoral fellowship funded by NIDCD Training Grant 5 T32
31
32 DC00041, a Wellesley College Provost's Office Grant, a fellowship from the Radcliffe Institute
33
34 for Advanced Study (J.E.P), NIH/NIDCD grant R01 DC010997 (K.E.), and by a Marie Curie
35
36 Fellowship MEIF-CT-2006-041886 (P.P.).
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

- 1
2
3
4
5
6 Alibali, M.W., & Hostetter, A.B. (2010). Mimicry and simulation in gesture comprehension.
7
8 Behavioral and Brain Sciences, 33, 433-434.
9
- 10 Amorim, M.A., Isableu, B., & Jarraya, M. (2006). Embodied spatial transformations: “Body
11
12 analogy” for the mental rotation of objects. Journal of Experimental Psychology, 13, 527-
13
14 547.
15
16
- 17 Brown, P., & Levinson, S.C. (1987). Politeness: Some universals in language use. Cambridge,
18
19 UK: Cambridge University Press.
20
21
- 22 Carlson, S.M., Mandell, D.J., & Williams, L. (2004). Executive function and theory of mind:
23
24 Stability and prediction from ages 2 to 3. Developmental Psychology, 40, 1105-1122.
25
26
- 27 Clark, H.H. (1973). Space, time, semantics, and the child. In Moore, T.E. (Ed.), Cognitive
28
29 development and acquisition of language (pp. 27-63). New York: Academic Press.
30
31
- 32 Clark, H.H. (1992). Arenas of language use. Chicago: University of Chicago Press.
33
- 34 Clark, H.H., & Murphy, G.L. (1982). Audience design in meaning and reference. In J.F. Leny,
35
36 W. Kintsch (Eds.), Language and comprehension (pp. 287-299). Amsterdam: North-
37
38 Holland.
39
40
- 41 Dell, G.S., & Brown, P.M. (1991). Mechanisms for listener-adaptation in language production:
42
43 Limiting the role of the “model of the listener.” In D. Napoli & J. Kegl (Eds.), Bridges
44
45 between psychology and linguistics (pp. 105-129). New York: Academic Press.
46
47
- 48 Emmorey, K. (1996). The confluence of space and language in signed languages. In P. Bloom,
49
50 M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), Language and space (pp. 171-209).
51
52 Cambridge, MA: MIT Press.
53
54

- 1
2
3 Emmorey, K. (2001). Space on hand: The exploitation of signing space to illustrate abstract
4
5 thought. In M. Gattis (Ed.), *Spatial schemas and abstract thought*, (pp. 147 – 174).
6
7 Cambridge, MA: MIT Press:
8
9
- 10 Emmorey, K. (2006). The signer as an embodied mirror neuron system: Neural mechanisms
11
12 underlying sign language and action. In M.A. Arbib (Ed.), *Action to language via the*
13
14 *mirror neuron system*, (pp. 103-135). Cambridge UK: Cambridge University Press.
15
16
- 17 Emmorey, K., Klima, E., & Hickok, G. (1998). Mental rotation within linguistic and non-
18
19 linguistic domains in users of American Sign Language. *Cognition*, 68, 221-246.
20
21
- 22 Emmorey, K., & Tversky, B. (2002). Spatial perspective choice in ASL. *Sign Language and*
23
24 *Linguistics*, 5, 3-25.
25
26
- 27 Epley, N., Keysar, B., van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric
28
29 anchoring and adjustment. *Journal of Personality and Social Psychology*, 87, 327-339.
30
31
- 32 Galati, A., & Avraamides, M.N. (2013). Flexible spatial perspective-taking: conversational
33
34 partners weigh multiple cues in collaborative tasks. *Frontiers in Human Neuroscience*
35
36 7(618), 1-16. doi: 10.3389/fnhum.2013.00618.
37
38
- 39 Galati, A., Michael, C., Mello, C., Greenauer, N.M., & Avraamides, M.N. (2013). The
40
41 conversational partner's perspective affects spatial memory and descriptions. *Journal of*
42
43 *Memory and Language*, 68, 140-159.
44
45
- 46 Gilbert, D.T., & Gill, M.J. (2000). The momentary realist. *Psychological Science*, 11, 394-398.
47
48
- 49 Goldin-Meadow, S., McNeill, D., & Singleton, J. (1996). Silence is liberating: Removing the
50
51 handcuffs on grammatical expression in the manual modality. *Psychological Review*,
52
53 103, 34-55.

- 1
2
3 Herrmann, T., Bürkle, B., & Nirmaier, H. (1987). Zur hörerbefugten Raumreferenz:
4 Hörerposition und Lokalisationsaufwand. *Sprache & Kognition*, 6(3), 126-137.
5
6
7
8 Horton, W.S., & Keysar, B. (1996). When do speakers take into account common ground?
9
10 *Cognition*, 59, 91-117.
11
12
13 Kessler, K., & Thomson, L. A. (2010). The embodied nature of spatial perspective taking:
14 embodied transformation versus sensorimotor interference. *Cognition*, 114, 72–88.
15
16
17 Keysar, B., Barr, D., & Horton, W.S. (1998). The egocentric basis of language use: Insights from
18 a processing approach. *Current Directions in Psychological Science*, 7, 46-50.
19
20
21
22 Keysar, B., Barr, D., Balin, J., & Brauner, J. (2000). Taking perspective in conversation: The
23 role of mutual knowledge in comprehension. *Psychological Science*, 11, 32-38.
24
25
26
27 Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, 89,
28 25-41.
29
30
31
32 Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT
33 Press.
34
35
36
37 Levinson, S. C. (2003). *Space in language and cognition: explorations in cognitive diversity*.
38 Cambridge UK: Cambridge University Press.
39
40
41
42 Lu, C.-H., & Proctor, R.W. (1995). The influence of irrelevant location information on
43 performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin &*
44 *Review*, 2, 174-207.
45
46
47
48
49 MacLeod, C.M. (1991). Half a century of research on the Stroop effect: An integrative review.
50 *Psychological Bulletin*, 109, 163-203.
51
52
53
54
55
56
57
58
59
60

1
2
3 Miller, G. A., & Johnson-Laird, P. N. (1976). Language and perception. Cambridge, MA:
4
5 Harvard University Press.

6
7
8 MPI for Psycholinguistics Field Manual (1992). In Stephen C. Levinson (ed.), Space stimuli kit
9
10 1.2: November 1992, 51. Nijmegen: Max Planck Institute for Psycholinguistics.

11
12
13 Perniss, P. M. (2007). Achieving spatial coherence in German Sign Language narratives: The use
14
15 of classifiers and perspective. *Lingua*, 117, 1315-1338.

16
17
18 Perniss, P., & Özyürek, A. (under revision). Channels of expression differentially modulate
19
20 iconicity and embodiment in co-speech gesture and sign language.

21
22
23 Piaget, J., & Inhelder, B. (1956). The child's conception of space. London: Routledge & Kegan
24
25 Paul.

26
27
28 Pickering, M., & Garrod, S. (2013). An integrated theory of language production and
29
30 comprehension. *Journal of Behavior and Brain Sciences*, 36, 375-392.

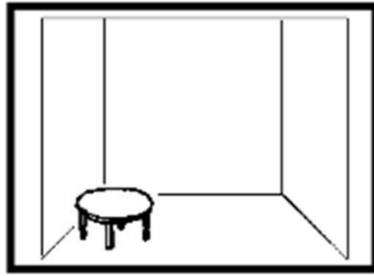
31
32
33 Pyers, J., Perniss, P., & Emmorey, K. (2008, February). Viewpoint in the visual-spatial modality.
34
35 Paper presented at the 30th annual convention of the German Society of Linguistics
36
37 workshop on Gestures: A comparison of signed and spoken languages, Bamberg,
38
39 Germany.

40
41
42 Pyers, J., & Senghas, A. (2009). Language promotes false-belief understanding: Evidence from
43
44 learners of a new sign language. *Psychological Science*, 20, 805-812.

45
46
47 Pyers, J., Shusterman, A., Senghas, A., Spelke, E., & Emmorey, K. (2010). Evidence from an
48
49 emerging sign language reveals that language supports spatial cognition. *Proceedings of*
50
51 *the National Academy of Science*, 107, 12116–12120.

- 1
2
3 Quinto-Pozos, D., Singleton, J.L., Hauser, P.C., Levine, S.C., Garberoglio, C.L., & Hou, L.
4
5
6 (2013). Atypical signed language development: A case study of challenges with visual-
7
8 spatial processing. *Cognitive Neuropsychology*, 30, 332-359.
9
10 Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition*, 47, 1-24.
11
12 Schober, M.F. (1995). Speakers, addressees, and frames of reference: Whose effort is minimized
13
14 in conversations about location? *Discourse Process*, 20, 219-247.
15
16
17 Senghas, A. (2010a). From gestures to grammar – How learners created Nicaraguan Sign
18
19 Language. Keynote address. Delivered at the Child Language Seminar, City University,
20
21 London, June 25, 2010.
22
23
24 Senghas, A. (2010b). The emergence of two functions for spatial devices in Nicaraguan Sign
25
26 Language. *Human Development*, 53, 285-300.
27
28
29 Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental*
30
31 *Psychology*, 18, 643-662.
32
33
34 Talmy, L. (2003). The representation of spatial structure in signed and spoken language. In K.
35
36 Emmorey (Ed.), *Perspectives on classifier constructions* (pp. 169–195), Mahwah, New
37
38 Jersey: Lawrence Erlbaum Associates.
39
40
41 Tversky, B., & Hard, B. (2009). Embodied and disembodied cognition: Spatial perspective-
42
43 taking. *Cognition*, 110, 124-129.
44
45
46 Wilson, M. (2001). Perceiving imitable stimuli: Consequences of isomorphism between input
47
48 and output. *Psychological Bulletin*, 127, 543-553.
49
50
51 Wraga, M., Creem, S.H., & Proffitt, D.R. (1999). The influence of spatial reference frames on
52
53 imagined object- and viewer rotations. *Acta Psychologica*, 102, 247-264.
54
55
56
57
58
59
60

FIGURES



A) Signer's viewpoint



TABLE

LOCATED_HERE_(left)

B) Perceiver's (addressee) viewpoint



TABLE

LOCATED_HERE_(right)

Figure 1. ASL description of the location of a table in the picture a) from the signer's egocentric perspective as he views the picture and b) from the addressee's viewpoint as he/she views the signs.

A) Non-shared space



signer



addressee

B) Shared space



signer



addressee

Figure 2. (A) When the signer talks about non-present referents or non-jointly viewed referents, the signer describes the spatial array from his own viewpoint and the addressee interprets the utterance non-egocentrically. (B) When signer and addressee jointly view an environment their different vantage points on the scene map to the same, shared physical space. The description of the scene in sign space by the signer or the addressee matches the real-world locations of the objects.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
A.)

B.)



Figure 3. The eight test pictures used in Experiment 1. The orientation of the man was not coded.

Therefore, the four pictures in Set A were considered correct matches for each other, and the four pictures in Set B were considered correct matches for each other.

Review Only

45
46
47
48
49
50
51

45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

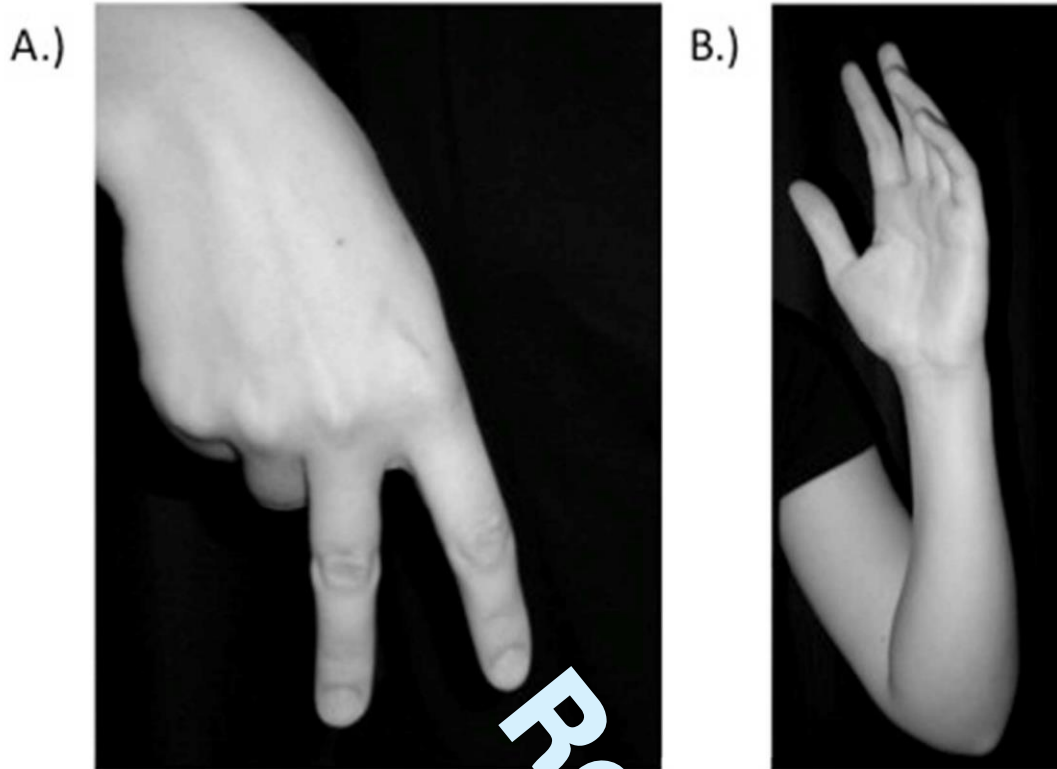


Figure 4. The two gestures taught to hearing sign-naïve participants in Experiment 1 to represent A) the man and B) the tree.

52
53
54
55
56
57
58
59
60

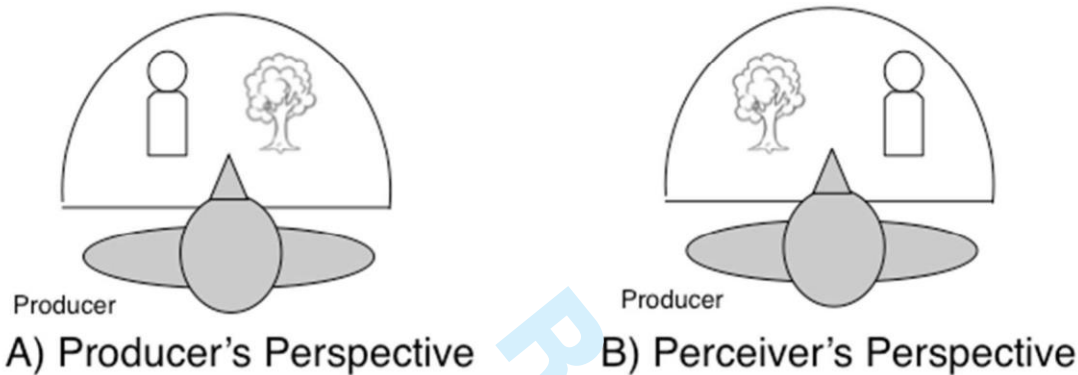


Figure 5. Diagram of how the Producer's utterances were coded for viewpoint. A) Signers or gesturers were coded as using the "Perceiver's Perspective" when they mapped the relative locations of the sign or gesture for 'man' and the sign or gesture for 'tree' directly onto the locations represented in the stimulus. In this case, the man is to the left of the tree from the Producer's perspective, and the Producer locates the man on his left side and the tree on his right side. B) Signers or gesturers were coded as using the "Producer's Perspective" when they reversed the relative locations of the sign or gesture for 'man' and the sign or gesture for 'tree' such that when the Perceivers viewed the manual production, the manual production mapped directly onto the spatial scene viewed by the Perceiver. In this case the man is to the left of the

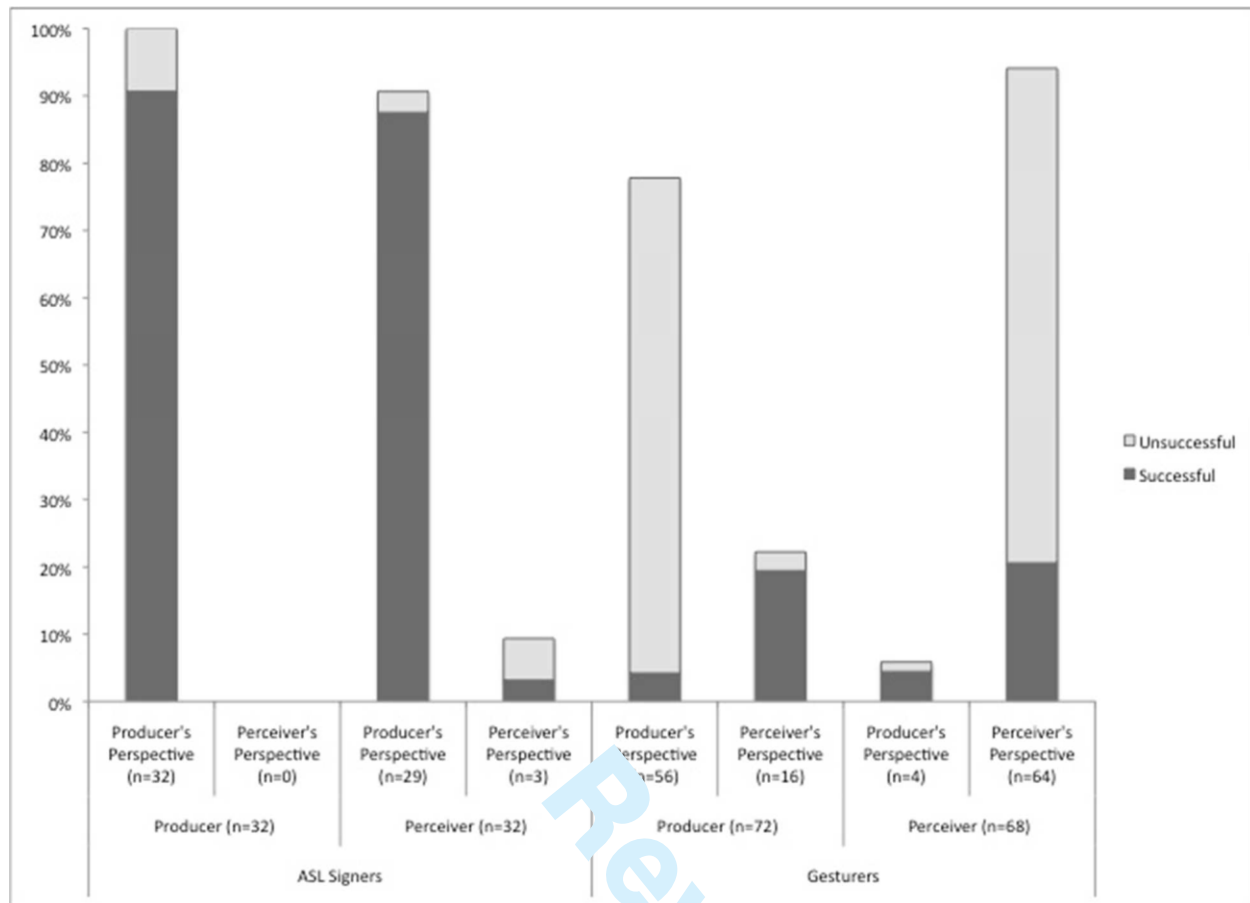


Figure 6. Percentage of successful and unsuccessful picture selections for each viewpoint strategy presented by Producers and interpreted by Perceivers in Experiment 1.

tree from the Producer's perspective, and the Producer locates the man on his right side (the Perceiver's left side) and the tree on his left side (the Perceiver's right side).

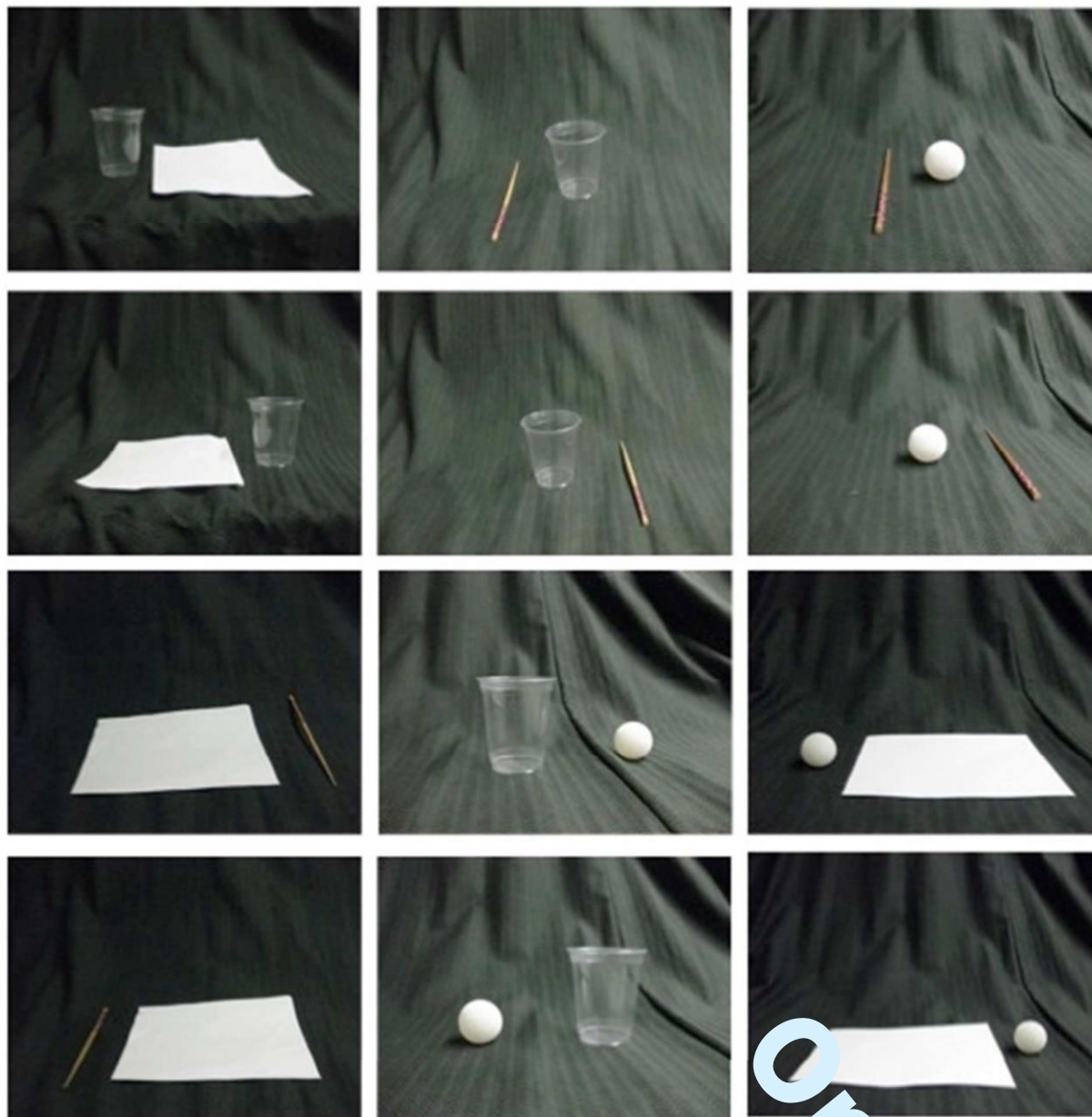


Figure 7. The Perceiver's layout of the twelve test pictures used in Experiment 2.

Only

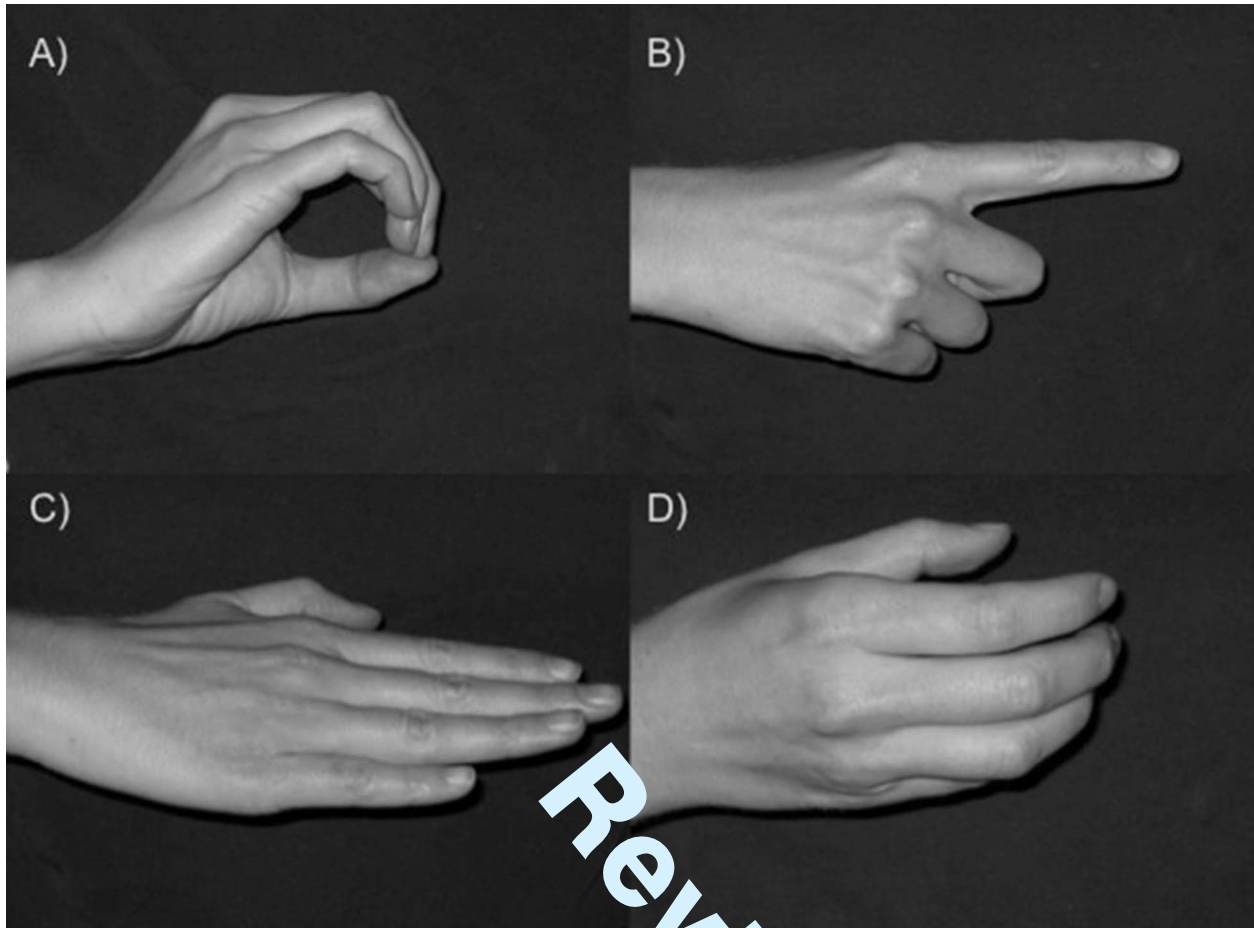


Figure 8. The four gestures taught to hearing sign-naïve participants in Experiment 2 to represent A) the ball, B) the chopstick, C) the paper, and D) the cup.

Viewpoint in the Visual-Spatial Modality

42

Table 1

Means (Standard Deviations) of Perspective Choices in Experiment 1 and in the Unassigned Condition of Experiment 2

Group	Pairs(N)	Producer		Perceiver		Correct Picture Selection(#) ^a
		Producer's Perspective(%)	Perceiver's Perspective(%)	Producer's Perspective(%)	Perceiver's Perspective(%)	
Experiment 1						
ASL Signers	4	100.0 (0.00)	0.0 (0.00)	90.6 (6.25)	9.4 (6.25)	7.3 (0.50)
Gesturers	9	77.8 (34.67)	22.2 (34.67)	5.6 (6.59)	88.9 (14.58)	1.9 (2.67)
Experiment 2						
Gesturers						
Unassigned Perspective	11	82.2 (28.41)	17.8 (28.41)	35.6 (43.32)	64.4 (43.32)	5.2 (4.77)

^aOut of 8 possible for Experiment 1 and 12 possible for Experiment 2.

Viewpoint in the Visual-Spatial Modality

Table 2
Means (Standard Deviations) and Inferential Statistics of Response Variables in the Experimental Conditions of Experiment 2

Condition	Pairs(N)	Producer			Perceiver		
		Correct Viewpoint Present (%)	First trial correct productiontime(s)	Average Correct Productiontime(s)	Correct Viewpoint Interpreted(%)	First trial correct selectiontime(s)	Average Correct SelectionTime(s)
Producer's Perspective	17	98.6 (3.12)	2.8 (1.49)	3.1 (1.55)	80.4 (36.08)	3.6 (1.16)	3.0 (.75)
Perceiver's Perspective	19	78.5 (34.74)	4.2 (2.26)	3.8 (1.72)	85.53 (31.3)	2.90 (1.8)	2.73 (.2)
Producer's vs Perceiver's Conditions	t, df, p, d	2.51 ^a , 18.32, .02, .81	2.04, 26.00, .05, .75	0.46, 34.00, .65, .39	0.46, 34.00, .65, .15	1.07, 22.00, .30, .27	1.23, 32.00, .23, .55

^a Assuming unequal sample variances, F(18,16) = 125.67, p < .0001