



How Pointing is Integrated into Language: Evidence From Speakers and Signers

Kensy Cooperrider^{1*}, Jordan Fenlon², Jonathan Keane², Diane Brentari² and Susan Goldin-Meadow¹

¹Department of Psychology, University of Chicago, Chicago, IL, United States, ²Department of Linguistics, University of Chicago, Chicago, IL, United States

OPEN ACCESS

Edited by:

Olga Capirci,
Italian National Research Council, Italy

Reviewed by:

Brendan Costello,
Basque Center on Cognition, Brain
and Language, Spain
Chiara L. Rivolta,
Basque Center on Cognition,
Brain and Language, Spain,
in collaboration with BC
Olivier Le Guen,
Center for Research and Higher
Studies in Social Anthropology, Mexico
N. J. Enfield,
The University of Sydney, Australia

*Correspondence:

Kensy Cooperrider
kensycoop@gmail.com

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: 30 May 2020

Accepted: 29 January 2021

Published: 03 May 2021

Citation:

Cooperrider K, Fenlon J, Keane J,
Brentari D and Goldin-Meadow S
(2021) How Pointing is Integrated into
Language: Evidence From Speakers
and Signers.
Front. Commun. 6:567774.
doi: 10.3389/fcomm.2021.567774

When people speak or sign, they not only describe using words but also depict and indicate. How are these different methods of communication integrated? Here, we focus on pointing and, in particular, on commonalities and differences in how pointing is integrated into language by speakers and signers. One aspect of this integration is *semantic*—how pointing is integrated with the meaning conveyed by the surrounding language. Another aspect is *structural*—how pointing as a manual signal is integrated with other signals, vocal in speech, or manual in sign. We investigated both of these aspects of integration in a novel pointing elicitation task. Participants viewed brief live-action scenarios and then responded to questions about the locations and objects involved. The questions were designed to elicit utterances in which pointing would serve different semantic functions, sometimes bearing the full load of reference ('load-bearing points') and other times sharing this load with lexical resources ('load-sharing points'). The elicited utterances also provided an opportunity to investigate issues of structural integration. We found that, in both speakers and signers, pointing was produced with greater arm extension when it was load bearing, reflecting a common principle of semantic integration. However, the duration of the points patterned differently in the two groups. Speakers' points tended to span across words (or even bridge over adjacent utterances), whereas signers' points tended to slot in between lexical signs. Speakers and signers thus integrate pointing into language according to common principles, but in a way that reflects the differing structural constraints of their language. These results shed light on how language users integrate gradient, less conventionalized elements with those elements that have been the traditional focus of linguistic inquiry.

Keywords: pointing, gesture, sign language, deixis, demonstratives, language

INTRODUCTION

When people communicate—whether by speaking or signing—they interweave at least three different methods (Clark, 1996; Enfield, 2009; Clark, 2016; Ferrara and Hodge, 2018). First, they use the categorical, highly conventionalized symbols—words—used within their language community. That is, they *describe*. These communicative resources are the traditional focus of linguistic inquiry. But people also use resources that are gradient and less conventionalized: they iconically represent images, actions, and sounds—they *depict*—and they draw attention to locations,

objects, and people—they *indicate*. This three-method framework—which builds on the semiotic theories of Charles Peirce (Peirce, 1940)—presents a powerful lens through which to understand the complexity and heterogeneity of human communication. But it also prompts a key question: How are these disparate methods of communication integrated? How do resources like depictions, points, and words come together into a coherent and fluent stream of meaning? And how do the mechanisms of integration differ in signed vs. spoken communication?

To shed light on these issues, we focus here on the case of pointing. Pointing—in which one person directs another's attention to a target location—is a ubiquitous form of indicating (Clark, 2003; Kendon, 2004; Cooperrider et al., 2018). It has been described as “a basic building block” (Kita, 2003a) of human communication: it is universal across cultures (e.g., Eibl-Eibesfeldt, 1989); is among children's first communicative acts (e.g., Bates, 1979; Liszkowski, 2006; Capirci and Volterra, 2008; Liszkowski et al., 2012); and is pervasive and multifunctional in both spoken (Clark, 2003) and signed (e.g., McBurney, 2009; Johnston, 2013) interaction. For these and other reasons, pointing has been widely examined across the cognitive and linguistic sciences (e.g., contributions in Kita, 2003b), and has been a focus of recent efforts to compare gesture and sign (e.g., Cormier et al., 2013; Johnston, 2013; Meier and Lillo-Martin, 2013; Goldin-Meadow and Brentari, 2017; Fenlon et al., 2018; Fenlon et al., 2019). Yet little work to date has closely examined how pointing is integrated with words (for some exceptions, see: Bangerter, 2004; Cartmill et al., 2014; Cooperrider, 2016; Floyd, 2016). The question of how pointing is integrated with the more highly conventionalized components of language is multilayered, and we focus on two important aspects of it.

First, how is pointing integrated with the intended meaning of the broader utterance in which it is embedded? We term this *semantic integration*, and several observers have noted that a point's form reflects this type of integration. Kendon (2004) noted that speakers point differently depending on how the target of the point “is presented in the speaker's discourse” (p. 201). Enfield et al. (2007) examined the size of a pointing gesture in relation to its function. They found that speakers of Lao were more likely to produce big points—that is, points with a greater degree of arm extension—when those points convey “a primary, foregrounded part of the message” (p. 1723). (They also suggested that, beyond involving greater arm extension, such points tended to be longer in duration, but they did not analyze this systematically [p. 1728]). The researchers described these contexts as “location-focus” utterances. The central example, they observed, occurs when someone answers a question about where something is; speakers will often respond to a “where” or “which” question by producing a point along with a demonstrative such as “here” or “there.” Thus, they find that the form of a pointing gesture is integrated with utterance-level meaning.

Building on the proposals in Enfield et al. (2007), we suggest that a point's form might also be integrated with meaning at a finer level: the lexical semantics of the phrase in which the point is embedded. Within location-focus utterances, pointing can carry

more or less of the burden of specifying location. For example, the question “Where did you park?” could be answered simply with a wordless point, or with a point along with a demonstrative, such as “there.” In such cases, the point bears the full load of specifying location; we call these instances “load-bearing points.” The same question could also be answered, however, with a point plus a longer description of location, such as “Over on the left, in the far back.” In this example, the point shares the load of specifying location with lexical material; we call such cases “load-sharing points.” To date, beyond the work of Enfield et al. (2007) on Lao co-speech gesture, we are not aware of any studies that have closely examined either of these types of semantic integration, utterance-level (location-focus or not) or lexical-level (load-bearing or load-sharing). A major goal of the present study is to determine whether pointing signs (points that signers produce) reflect semantic integration in the same way as pointing gestures (points that speakers produce).

Second, how is pointing as a manual signal structurally integrated with other signals involved in the utterance? We term this *structural integration*. Our focus is manual pointing, although non-manual forms are prominent in some communities (see, e.g., Cooperrider et al., 2018). This aspect of integration has the potential to differ for speakers and signers because spoken and signed communication are put together differently. Spoken communication involves different articulatory channels operating in parallel, with speech produced in tandem with gestures of the hands and movements of the face and body. Importantly, in spoken communication, much of the describing—that is, the use of highly conventionalized symbols—is done with the mouth, and much of the depicting and indicating is done with the hands. Signed communication similarly involves different articulatory channels working in parallel: manual signs are produced in tandem with movements of the face and body. But, in contrast to spoken communication, the bulk of the describing, depicting, and indicating all occur in the hands (though enriched with critical information in the face and body).

These differences have clear consequences for the structural integration of pointing: manual pointing gestures must integrate with spoken words, which are produced in a different articulatory channel, whereas manual pointing signs must integrate with signs within the same articulatory channel. Fenlon et al. (2019) described this situation as a “same-channel constraint” and suggested that the constraint may account for certain differences between points that accompany sign language and points that accompany spoken language. For instance, in their comparison of pointing signs and pointing gestures in two conversational corpora, Fenlon et al. (2019) found that pointing signs were much shorter in duration overall than pointing gestures, which may have been related to the pressure to slot those points in to a stream of manual signals. Speakers, of course, do not face the same-channel constraint; their pointing gestures are therefore free to take more time, spanning across much of an utterance or even bridging over adjacent utterances. Fenlon et al.'s (2019) account of these tendencies, however, needs to be further investigated. After all, signers do sometimes point with their non-dominant hand, thus allowing those points to span

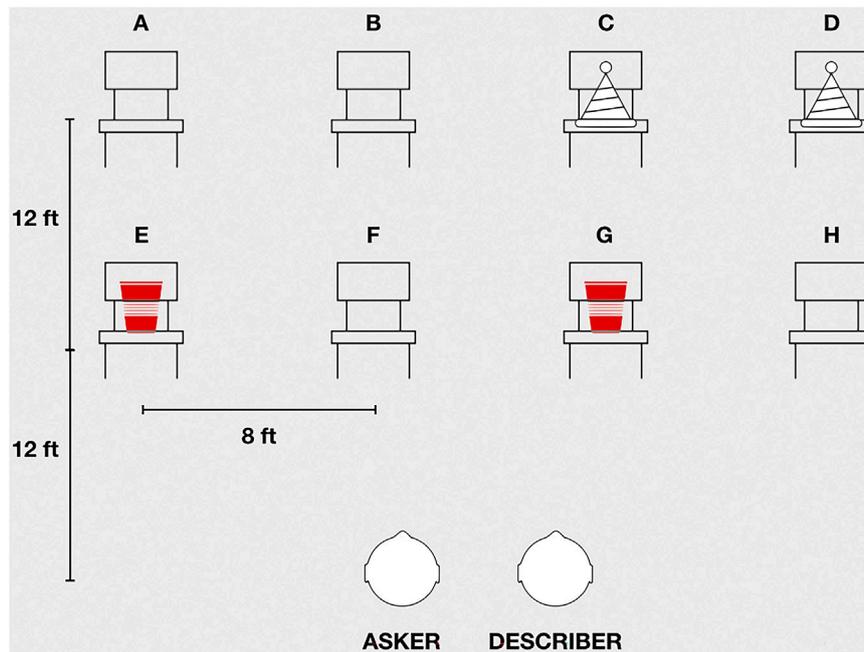


FIGURE 1 | A schematic depiction of the task layout (not to scale). Participants sat before an array of chairs (labeled **A–H**) in a large performance space. After watching an actor carry out actions involving two pairs of objects (e.g., hats, cups), the describer answered a series of questions posed by the asker.

across signs produced with their dominant hand (Johnston, 2013). Moreover, speakers occasionally slot their gestures in to breaks in speaking, a phenomenon occasionally called “component gestures” (Clark, 1996). Thus, a second goal of the present work is to examine these aspects of structural integration more closely for both pointing gestures and pointing signs.

In the present study, we sought to investigate both semantic and structural aspects of integration in speakers and signers. To this end, we developed a novel pointing elicitation task. English speakers and ASL signers watched brief live-action vignettes in which an actor interacted with objects at different locations in an eight-location grid. Participants were then asked questions about the locations and objects in the vignettes. Some questions were designed to elicit utterances in which location would be focal (e.g., an answer specifying where a particular object was located), whereas other questions were designed to elicit utterances in which location would be mentioned but not focal (e.g., what action was performed at a given location). The elicited utterances also provide a data set within which to examine how points were integrated with other signals, whether manual or vocal. We thus extend Enfield et al. (2007) by examining the semantic integration of pointing in another spoken language (English) and, for the first time, in a sign language (ASL). We also extend Fenlon et al. (2019) by focusing on points to objects and location rather than persons; by using a controlled task rather than conversational data; and by looking at the consequences of the “same-channel” constraint in further detail. The overarching goal is to shed light on how speakers and signers integrate disparate forms of communication into coherent streams of discourse.

METHODS

Participants

24 adults (mean age = 30.3, 8 women) from the Chicago area participated in exchange for payment. One group, referred to as ‘speakers’ ($n = 12$; 5 women), consisted of hearing participants who were native or near-native speakers of English. None of the speakers reported knowing ASL. The other group, referred to as ‘signers’, consisted of deaf participants ($n = 12$, 3 women) who were native or near-native signers of American Sign Language (ASL). The study was conducted in pairs of participants from the same group.

Materials and Procedure

The study was conducted in a large performance space. The participants sat in two chairs near the stage, facing out toward where the audience would typically be seated. In the middle of the room were two rows of four chairs, one row approximately 12 feet from participants’ seats and the other approximately 12 feet beyond that. Two rows at different distances offered the possibility of analyzing whether target distance affects pointing form. Within each row, the chairs were spaced approximately eight feet apart (Figure 1).

For each trial, one participant served as the *describer* and the other as *asker*. Before the beginning of a trial, the asker left the room; the describer stayed seated and watched as an experimenter acted out a scripted scenario. On the chairs were four objects of two types (e.g., two identical cups and two identical party hats). Each scenario would involve the experimenter carrying out a series of four actions, which involved the four objects at different

chairs. For example, in one scenario (see **Figure 1**), the four objects were located on four different chairs: two identical red cups on chairs E and G, and two identical white party hats on chairs C and D. At the start of the scenario—signaled by the experimenter raising her hand—the experimenter moved to chair E (far left chair, front row) and retrieved one of the cups (action 1); she then moved to chair B (left middle, back row) and took a sip from the cup (action 2); she then walked to chair H (far right, front row) and placed the cup on the chair (action 3); finally, she walked to chair D (far right, back row) and put on the party hat (action 4). Finally, the experimenter signaled the end of the scenario by raising her hand.

At the end of each scenario, the asker re-entered the room and was given a single piece of paper with five questions to ask the describer. The first four questions asked for “which”/“where” information. These questions were designed to elicit utterances focused on a single, specific location (location-focus utterances). For example, the question “Which cup was used?” might elicit a location-focus utterance such as “That one.” The final question, identical for all scenarios, was: “Could you please explain everything that happened, in the order that it happened?” This prompt was designed to elicit utterances that may have mentioned locations, but were focused on explaining the actions and the sequence in which they took place (“explanatory utterances”). For example, in response to this question a speaker might begin a longer explanation with the utterance, “So she picked up the cup from that chair.”

Each pair completed two practice scenarios—one per participant—followed by eight primary scenarios. The participants switched roles after every scenario, with each participant completing four scenarios as describer and four as asker. The chairs remained in the same position across all scenarios; the four objects varied, as did the relevant locations and the actions performed at each location. All locations were used across the scenarios but not equally. Full information about the scenarios, as well as a video of the experimenter performing them, is available in the online Supplementary Material: <https://osf.io/wckx5/>. Finally, pointing was not mentioned in the instructions and all sessions were video-recorded.

Analysis

The videos were segmented into scenarios and analyzed using ELAN annotation software (Lausberg and Sloetjes, 2009). Only the describer’s behaviors are discussed here. The primary coder was a hearing signer with experience analyzing bodily communication; this coder performed all the pointing and language analyses, with another experienced coder performing reliability as described below.

Pointing Gesture Analysis

Following prior work (e.g., Fenlon et al., 2019), points were defined as movements toward a region of space that were intended to direct attention to that region. Points produced by the speakers (pointing gestures) and points produced by the signers (pointing signs) that were directed to any of the eight target locations—i.e., the chairs—were identified for further analysis (see ‘Pointing analysis’ section of the coding manual

in the online **Supplementary Material**). We only considered those points that exclusively conveyed location, sometimes called “pure points” (Kendon, 2004). Pointing gestures that simultaneously conveyed location and other information (e.g., gestures directed to a location but also conveying the action of picking up an object or moving points showing the trajectory the experimenter took between locations) were excluded, as were points judged to be non-communicative (e.g., produced as part of a private rehearsal of the scenario).

Points were coded for two features: extension and duration. These features were selected, following prior studies, because they seem to vary according to a point’s communicative importance and function (Enfield et al., 2007; Peeters et al., 2015). For extension, we distinguished points with full extension—an arm that was straight across the elbow—from points with partial extension—an arm that had a bend at the elbow. (We initially tried to distinguish a third category of “minimal extension” points, which had a bend at the elbow and no raising of the upper arm, but were unable to achieve high reliability for this distinction.) For duration, we coded the onset of the point—defined as the first frame of motion from rest or from a prior gesture or sign—and the offset—defined as the last frame of the hold phase (or frame of fullest extension, if no hold), before returning to rest or beginning another gesture/sign. Duration was then determined based on the time between pointing onset and offset.

Reliability was assessed by having a second coder (another hearing signer with experience analyzing bodily communication) analyze one randomly selected scenario from each participant (of the four completed; i.e., 25% of the data). Agreement on the presence of pointing was 80%. Discrepancies resulted from: one coder identifying a point that the other did not; one coder identifying a single point where the other identified two or more points; and one coder considering a point to be iconic or non-communicative, whereas the other considered it a “pure point.” For those points identified by both coders as pure points ($N = 238$; other points were not coded for extension), they agreed on whether extension was partial or full 93% of the time (Cohen’s $K = 0.87$). Further, the durations attributed to points by the two coders were highly correlated with each other ($r = 0.94$).

Language Analysis

For each point, we identified the utterance in which it occurred. Utterances were segmented using a combination of grammatical criteria (i.e., clause boundaries) and visible/audible criteria (e.g., pauses) (Du Bois et al., 1993; Sandler et al., 2005; Fenlon et al., 2007; Brentari et al., 2011) (for full details, see ‘Language analysis’ section of the coding manual). These utterances thus closely aligned with the notion of an “intonational phrase” (Nespor and Vogel, 1986), but we use “utterance” for simplicity. After each utterance containing a point was demarcated, its contents were transcribed; transcriptions were crosschecked by the second coder. As part of this transcription process, the primary coder marked, within the string of words or signs, the hold onset of each point to the nearest word boundary (with an open bracket) and the hold offset of each point to the nearest word boundary (with a

TABLE 1 | Examples of utterance types.

Utterance type	Examples from speakers	Examples from signers
Location-focus		
<i>Load-bearing</i>	[that] chair	HAT []
	right [there]	[]
	[] that one	BAG []
<i>Load-sharing</i>	the [back] left chair	BACK ROW [] THIRD []
	that [far corner] brush	LEFT []
	[front] far right	BACK [] ALL-THE-WAY []
Explanatory		
	then she moved back one [chair]	DRINK [] FINISH
	she picked up the cup from [that] chair	WALK [FRONT SECOND]
	and she put the ball in [the] chair	PUT-DOWN-CUP []

Note: Points not produced along with words or signs are marked as []. For points that were produced along with words or signs, we indicate the onset and offset of the hold with brackets around the overlapping lexical material.

closed bracket). This step allowed for subsequent analysis of different aspects of structural integration.

These transcribed utterances were then coded for their semantic function at two levels of granularity (Table 1). At one level, we coded whether the utterances were “location-focus” or “explanatory.” “Location-focus” utterances served to specify *which* or *where* information, and were most often sentence fragments with no verb (e.g., “In the back row”). “Explanatory” utterances were defined as those that included a verb describing the actions of the experimenter (e.g., “Then walked to that chair”). At a second, more fine-grained level, we further divided the location-focus utterances based on the semantic role of the point vis-à-vis the other lexical material in the same utterance¹. If the point specified location entirely on its own, it was considered a “load-bearing point”; if the utterance in which the point occurred contained any other spoken words or lexical signs that helped specify location (e.g., *back*, *front*, *left*, *right*), the point was considered a “load-sharing point.” We also coded whether speakers’ points were co-produced with a demonstrative—*this*, *that*, *here*, or *there* (e.g., Diessel, 2006). (The same analysis was not carried out for signers because the lexical item *THAT* was used rarely—by only three of the signers—and primarily served an anaphoric function.)

Points were also coded for whether, and how, they were structurally integrated with the other lexical material. We first noted whether the utterances that included a point also included lexical material. Points within utterances that did contain other

material were further analyzed as follows. Points were considered to slot in if they were not produced concurrently with other words (e.g., “[] that one”) or signs (e.g., “PUT BOWL []”). Points were considered to span across if they were produced concurrently with words (e.g., “that [far corner] brush”) or signs (if produced with one hand while the other hand produced additional signs; e.g., “WALK [FRONT SECOND]”) (as indicated by square brackets/underlining containing words or glosses); and they were considered to bridge over if they were held across adjacent utterances (speaker example: “So, walked over to this [chair /picked up the ball]”; signer example: “WALK [] FRONT THIRD [/ PUT DOWN HAT]”). On the basis of the transcriptions, utterances were further coded for the number of words or signs, in addition to the point, that they contained.

RESULTS

General

Pointing—though not mentioned in the instructions or modeled in any way—was a prominent part of how both speakers and signers carried out the task (see examples, Figure 2). Speakers produced a total of 513 points (mean per participant = 42.8) (again, including only pure points to target locations and excluding points with iconic aspects, as described earlier). At least one point was produced in 93% of responses to the “which”/“where” questions and 98% of responses to the explanation prompts. Signers produced a total of 508 points (mean per participant = 42.2), and at least one point occurred in 90% of responses to the “which”/“where” questions and 90% of responses to the explanation prompts. Despite the fact that pointing is usually considered optional in spoken communication but

¹A similar, more fine-grained analysis of explanatory utterances was not possible because this class was more heterogeneous than the location-focus utterances. Some of the speakers’ points in explanatory utterances, for instance, did not occur with location references in speech.

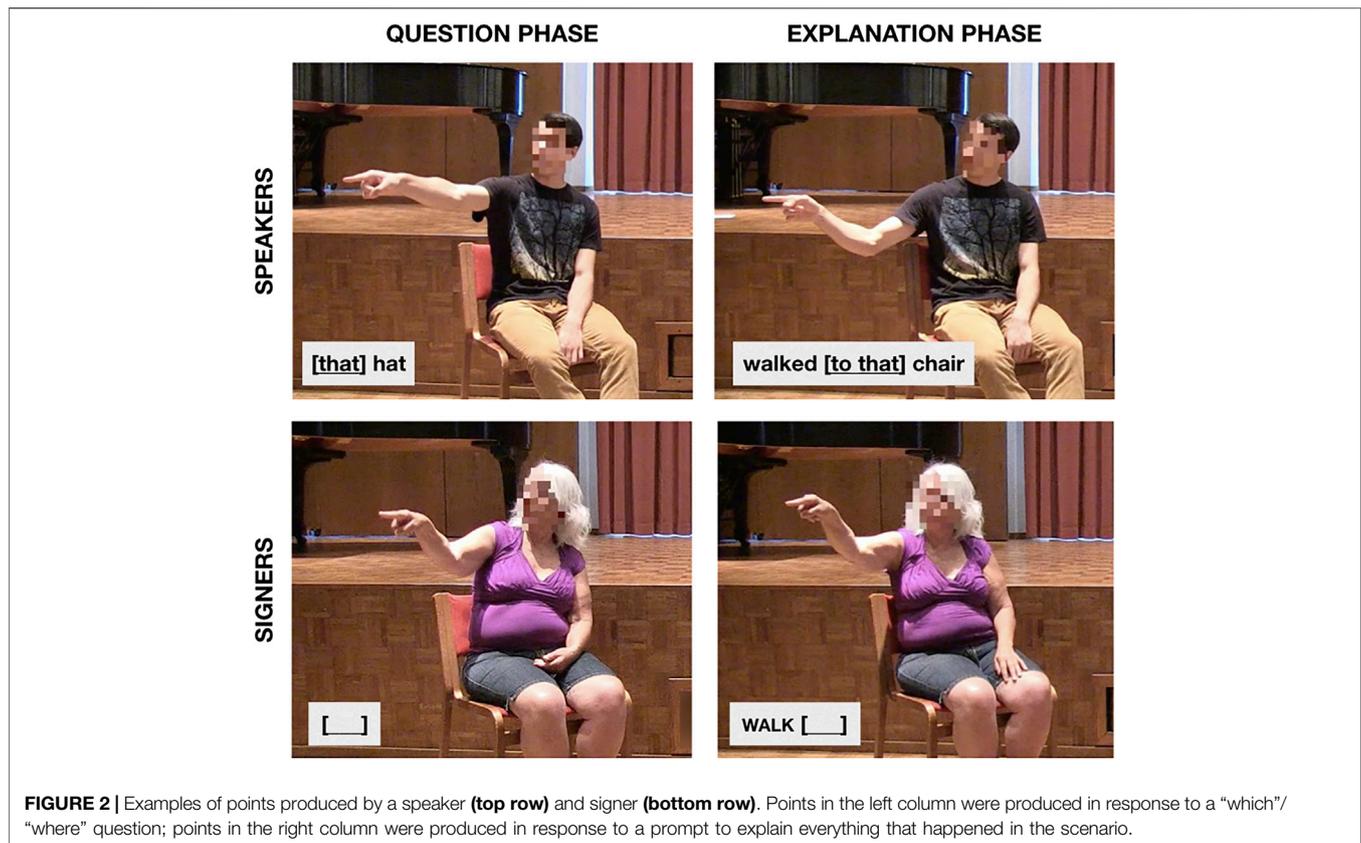


FIGURE 2 | Examples of points produced by a speaker (**top row**) and signer (**bottom row**). Points in the left column were produced in response to a “which”/“where” question; points in the right column were produced in response to a prompt to explain everything that happened in the scenario.

integral to the language in signed communication, the amount of pointing was thus comparable in the two groups.

We next confirmed that the two different question types—“which”/“where” or explanation prompt—successfully elicited points embedded in different types of utterances. Of the points produced in response to the “which”/“where” questions, most were part of location-focus utterances (speakers = 90%; signers = 86%). Conversely, of the points produced in response to the explanation prompts, most were part of explanatory utterances (gesturers = 82%; signers = 73%). These percentages confirm that the elicitation procedure worked as expected. In what follows, rather than use question type (i.e., “which”/“where” or explanation prompt) in our analyses, we use semantic function—load-bearing location, load-sharing location, and explanatory—as the key predicting variable.

All three semantic functions were common in both groups (Note that semantic integration analyses excluded the 68 points—64 in speakers, 4 in signers—that bridged over utterances, as these were difficult to assign to a single semantic function.) Speakers produced 108 load-bearing location points (24% of all speakers’ points), 146 load-sharing points (33%), and 195 explanatory points (43%). Signers produced 130 load-bearing location points (26% of signers’ points), 194 load-sharing location points (38%), and 184 explanatory points (36%). 17 participants produced points associated with all three semantic functions; the other seven participants (four gesturers and three signers) produced points with only two of the semantic functions. For

a breakdown of the number of observations of each type contributed by each participant, see **Supplementary Table S1**.

Results Bearing on Semantic Integration

We next analyzed the points for whether their form reflected their semantic integration with the surrounding language, focusing on extension and duration. For all analyses, we built hierarchical regression models in R (R Core Team, 2016) using the lme4 package (Bates et al., 2015). We used the maximal models that were both justified by theoretical concerns (e.g., including group and semantic function, and their interaction) and were able to converge. Unless otherwise specified, values in all figures and the text are estimates from these models rather than raw values. For ease of interpretability, we present these estimates in their natural interpretation space (i.e., probabilities or untransformed milliseconds) along with 95% confidence intervals (CIs), rather than coefficient estimates and standard errors (available in the **Supplementary Material**, along with *t* or *z* values).

To judge the direction and magnitude of effects, we use CIs, rather than *p* values (e.g., Gelman and Tuerlinckx, 2000; Cumming and Finch, 2005). An advantage of CIs is that they provide an intuitive visual format for judging the precision of point estimates and the size of effects, and thus discourage the “dichotomous thinking” (Cumming, 2014, p. 8) associated with *p* values. We judge whether two model estimates are different in a gradient manner. If the CIs of the estimates do not overlap at all, this indicates a robust difference; if one estimate falls outside of

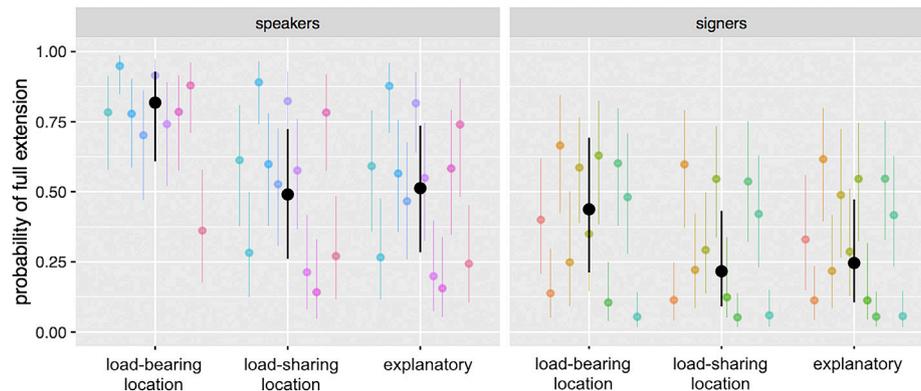


FIGURE 3 | Results from the model predicting the probability of using full arm extension, as a function of participant group and semantic function. Speakers were much more likely to use full extension overall, but in both groups load-bearing location points were more likely to involve full arm extension than other points. Black dots represent group means, and black lines represent 95% confidence intervals around those means; colored dots and lines represent individual participants.

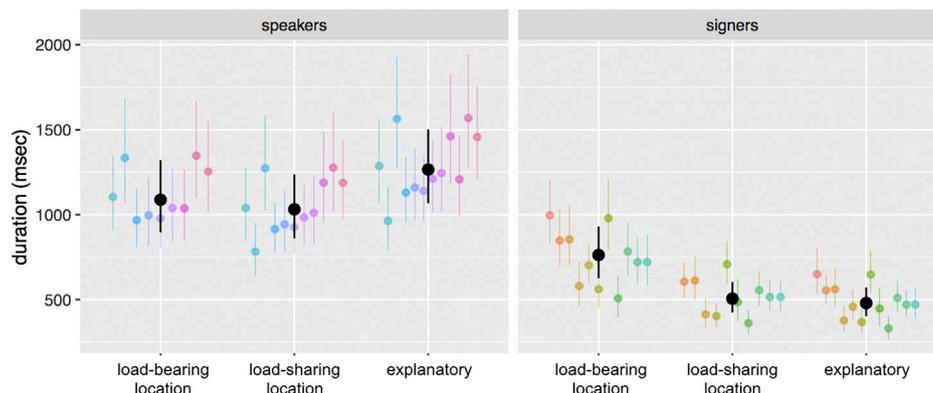


FIGURE 4 | Results from the model predicting the duration of a point as a function of participant group and semantic function. Speakers' points were longer in duration overall than signers' points, and were longer when part of explanatory utterances. In contrast, signers' load-bearing location points were longer than either of the other two types. Black dots represent group means, and black lines represent 95% confidence intervals around those means; colored dots and lines represent individual participants.

the CIs of the other (but the bounds of the CIs overlap), this suggests weaker evidence for a difference; if one estimate falls squarely inside the CIs of the other, this suggests a lack of strong evidence for a difference. We use these as guidelines rather than rigid cut-offs; they are not meant to substitute for traditional significance testing. For discussion of similar guidelines and their relation to significance testing, see Cumming and Finch (2005). Finally, in all figures, we also estimate confidence intervals for each participant using bootstrap sampling.

Extension. To analyze extension, we first built a hierarchical logistic regression model with group (speakers or signers) and semantic function (load-bearing location, load-sharing location, explanatory) as predictors, and with random effects for participant and target location (intercept only). In this and other models, these random effects were chosen because they were expected to exhibit variability that was orthogonal to our measures of interest—i.e., individuals may vary in pointing style

and different locations may prompt points with different degrees of arm extension. Overall, speakers' points were much more likely to involve full extension than signers' points, with speakers using full extension 57% of the time and signers only 36% of the time (raw percentages). However, extension in both groups was shaped by how the point fit semantically with the surrounding speech or signing, with participants in both groups being more likely to use full extension for load-bearing location points (speakers: 82% [95% CI: 61–93%]; signers: 44% [95% CI: 21–69%]), compared to either load-sharing location points (speakers: 49% [95% CI: 26–72%]; signers: 22% [95% CI: 9–43.2%]) or explanatory points (speakers: 51% [95% CI: 26–74%]; signers: 25% [95% CI: 11–47%]) (Figure 3). This pattern of data—though somewhat more robust in the speakers than in the signers—suggests that whether the utterance as a whole is location-focus or explanatory is not the important factor for predicting extension in either group. Rather, what matters is how

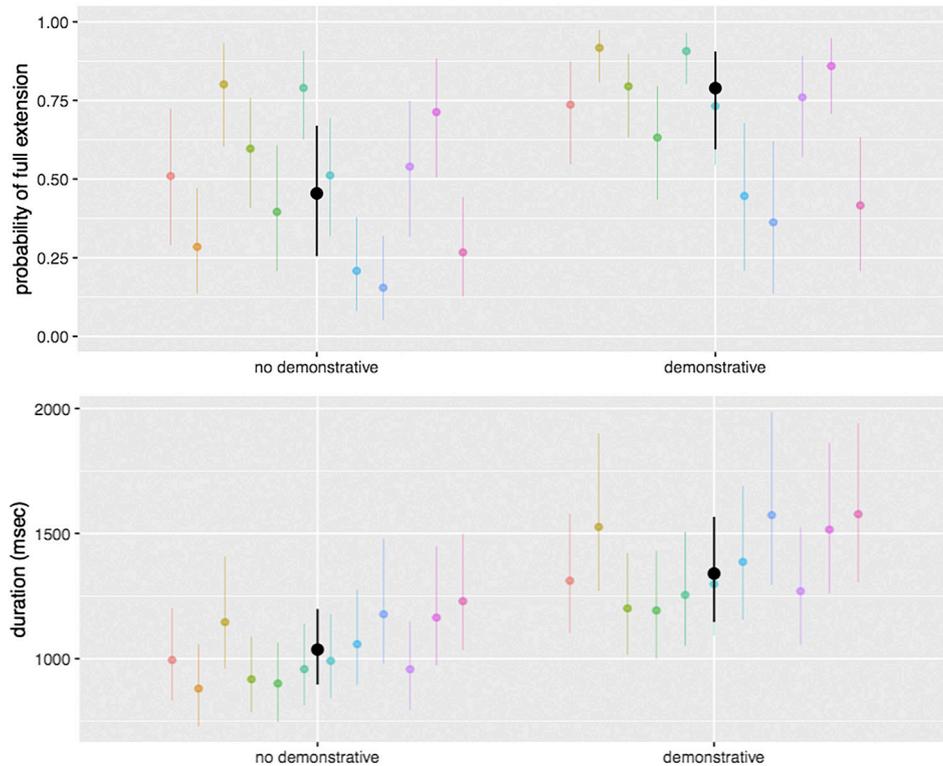


FIGURE 5 | Results from a model of speakers only, predicting the probability of full arm extension (**top**) and duration (**bottom**) as a function of whether or not a point is associated with a spoken demonstrative (*this, that, here, there*). Points that were associated with a demonstrative were more likely to involve full arm extension, and were longer in duration, than points that were not associated with a demonstrative. Black dots represent overall means and black lines represent 95% confidence intervals around those means; colored dots and lines represent individual participants.

the point relates to the lexical semantics within that utterance. We also built another version of the model with number of words in the utterance added as a predictor, but this addition did not substantially change the observed effects; on its own, number of words does not have an effect on extension. Lastly, we built a version of the model with location added as a predictor. Some target locations (i.e., B, C, and H, all far from the speaker) were associated with full-extension points more than others, but the effects of semantic function remain.

Duration. Next, we built an analogous model to analyze duration. Overall, speakers' points were much longer in duration (in milliseconds) than signers' points (speakers: $M = 1327$; signers: $M = 724$) (raw group means). Moreover, duration patterned differently across the different semantic functions in the two groups (**Figure 4**). In speakers, points embedded in explanatory utterances (explanatory = 1266 msec [95% CI: 1066–1502 msec]) were longer than those embedded in either type of location-focus utterance (load-bearing = 1088 msec [95% CI: 896–1321 msec]; load-sharing = 1032 msec [95% CI: 861–1237 msec]). In signers, the pattern was reversed; mirroring the results for extension, signers' points were longer when embedded in load-bearing location utterances (load-bearing location = 763 msec [95% CI: 625–930 msec]) than in either of the other two types of utterances (load-sharing location = 506 msec [95% CI: 424–603 msec]; explanatory = 479 msec

[95% CI: 402–572 msec]). As in our extension analyses, we built a model in which we added number of words in the utterance as a predictor. Adding this factor does not substantially change the observed effects. Moreover, there was no strong relationship between duration and number of words on its own, though in speakers there is a weak trend toward longer point durations as the number of words in the utterance increases; in signers there is no evidence of such a trend.

Demonstratives. Finally, we also conducted a separate analysis for speakers, asking whether points that were associated with spoken demonstratives ($N = 175$, or 39% of all points) differed from points that were not. (Again, we did not attempt to analyze this in signers; although ASL signers do make occasional use of a lexical item glossed as *THAT*, only three signers in our study used this sign and the predominant function appeared to be anaphoric rather than to support a co-occurring pointing sign.) We built analogous models predicting extension or duration, but adding in demonstrative presence as a predictor (and removing function and participant group, as we were only looking at speakers). We found that pointing gestures that were associated with demonstratives were more likely to be fully extended (demonstrative present = 79% [95% CI: 59–91%]) than pointing gestures that were not associated with demonstratives (demonstrative absent = 45% [95% CI: 26–67%]). Points that were associated with demonstratives were also longer in duration

speakers

A



[] on that chair

B



or [behind] that

signers

C



BACK SECOND CHAIR []

D



WALK [FRONT SECOND]

FIGURE 6 | Examples of different structural integration possibilities. Each example consists of an utterance containing a point and three words; each frame shows the participant during production of the word (or standalone point) transcribed below. In some cases, the points slotted in to the string of words, a pattern observed rarely in speakers (**A**) but very commonly in signers (**C**). In other cases, the points spanned across words in the utterance, a pattern observed very commonly in speakers (**B**) but rarely in signers (**D**).

TABLE 2 | Structural integration.

Type of point	Speakers	Signers
	count (raw mean duration in msec)	
Stand-alone points	20 (717)	110 (963)
Points contained within an utterance	429 (1356)	394 (657)
Spanning across words	399 (1381)	26 (1125)
Slotting in to words	30 (1020)	368 (624)
Points bridging over adjacent utterances	64 (2801)	4 (2736)
Total	513	508

(demonstrative present = 1340 msec [95% CI: 1146–1566 msec]) than points that were not associated with demonstratives (demonstrative absent = 1036 msec [95% CI: 896–1198 msec]) (Figure 5). We built additional models adding in semantic function as a predictor, which attenuates the effect of demonstrative presence on extension but not on duration.

Results Bearing on Structural Integration

We turn now to issues of structural integration. We first analyzed whether points were stand-alone in that they were the only signal within an utterance; whether they were contained within an utterance that included other lexical material; or whether they bridged over adjacent utterances. In both speakers and signers, the majority of points were produced within an utterance that also contained other lexical material (speakers = 84%; signers = 78%). Signers produced a greater percentage of stand-alone points than speakers (speakers = 4%; signers = 22%), and speakers produced a greater percentage of bridging over points than signers (speakers = 12%; signers = 1%) (Table 2).

Zooming in on those points produced as part of an utterance that contained other lexical material, we next considered how points are integrated with that material. In particular, we were interested in whether points “slotted in” (that is, fit into gaps in the string of words or signs) or “spanned across” (that is, overlapped with other words or signs in the utterance) (Figure 6). Although speakers’ points did occasionally slot in to the lexical string (7%), the overwhelming majority of speakers’ points spanned across part (or all) of that string (93%). Signers showed the opposite pattern, with their points slotting in to other signs (93%) more often than spanning across other signs (7%).

In sum, all the structural integration categories we considered were deployed by both groups. Yet each group nonetheless exhibited a clear characteristic profile, likely because of the same-channel constraint that exerts pressure on signers but is absent in speakers. Moreover, informal inspection of the durations of points associated with these different categories reveals other patterns (Table 2). In both groups, points that span across lexical material are numerically longer than points that slot in; and, similarly, points that bridge over utterances are substantially longer than points that are contained within an utterance (though there are few examples of bridging over points in signers). Thus overall differences in pointing duration between

speakers’ and signers’ points may be due largely to the contrasting structural integration profiles that each group exhibits.

DISCUSSION

Human communication—whether spoken or signed—involves the fluent integration of pointing into language. Here, we used a novel elicitation task to better understand how this integration is achieved, at both semantic and structural levels. To investigate semantic integration, we used different prompts expected to elicit utterances in which pointing served different semantic roles vis-à-vis linguistic elements. At the broadest level, these points sometimes supported utterances in which location was focal and other times supported utterances that were more explanatory; at the finer lexical level, these points sometimes bore the full load of communicating location and sometimes shared that load with words. We found that both speakers and signers were more likely to produce points with full arm extension when the point bore the full load. (The duration of the points proved more nuanced, a finding we discuss in more detail below.) To investigate structural integration, we used the resulting utterances from this elicitation task to better understand how pointing coordinates with surrounding words/signs and, in particular, how it slots in to, or spans across, other spoken or manual signals. We found that, while speakers and signers both integrate pointing with other signals in diverse ways, they also show strong characteristic profiles, with pointing gestures usually spanning across words—and sometimes bridging across utterances—and pointing signs usually slotting in between signs—and rarely spanning across them or bridging across utterances. These differences, we argue, are best understood in light of the differing modality constraints faced by speakers and signers. Put together, our results show how speakers and signers integrate pointing into language in response to common pressures, while at the same time navigating constraints that are particular to spoken or signed communication.

Our findings about semantic integration conceptually replicate and build on earlier observations. We extended Enfield et al.’s (2007) core finding to another spoken language and, for the first time, to a signed language. However, it is important to note key differences between the present findings and these prior ones. One is that we further subdivided location-focus utterances into two finer categories—those in which the point uniquely specified location (i.e., ‘load-bearing points’) and those in which the point was supplemented by other lexical material that helped specify location (i.e., ‘load-sharing points’). In our data, the significant difference proved to be between these two finer subtypes rather than between the coarser-grained distinction between location-focus and not, as Enfield et al. reported. A possible reason for this discrepancy is that the difference between location-focus and other points was more subtle in our task than it was in Enfield et al. study. Our points all concerned a small set of possible locations, all relatively near the interlocutors. By contrast, their data—which was from naturalistic interviews—likely yielded more heterogeneous

pointing behaviors, produced for a wider variety of pragmatic purposes and toward referents at different distances. If we had elicited a broader range of utterances, we might well have seen a basic difference between location-focus points and other points. Our findings about extension also fit with earlier observations that gestures with heightened communicative status exhibit a greater degree of effort (Peeters et al., 2015; Cooperrider, 2017). For instance, one study found that participants were more likely to point with their arms fully extended when their interlocutors could see them, compared to when they could not (Bangerter and Chevalley, 2007).

Although we found that the extension of points reflected semantic integration in a way common to speakers and signers, we did not find the same pattern for duration. Signers' points were longer in duration when they were load-bearing (much as their load-bearing points were more likely to involve full arm extension). But speakers' points did not fit this pattern—there was no difference in duration between load-bearing and load-sharing points. (This is another way that our data appear to depart from Enfield et al. (2007). They observed—but did not formally analyze—a pattern in which big points tended to be held for longer.) Although at first puzzling, this pattern becomes intelligible in light of our findings about structural integration. Signers' points are highly constrained by the need to produce other signs in the manual channel. Speakers' points are not so constrained, particularly because speakers do not often produce more than one gesture per utterance (McNeill, 1992). As a result, during longer utterances—such as the explanatory utterances we observed here—speakers' points may “stretch out” for longer periods of time. Further studies will be needed to identify the precise factors that determine how long pointing gestures are maintained. One possibility is that duration in speakers' points reflects how the point relates to the speech it accompanies—i.e., how it “takes scope” over some relevant portion of the utterance. The upshot is that, in signers, duration serves as a reliable cue to the importance of the point in the utterance but, in speakers, duration is more complexly determined.

We also analyzed, in the speakers, whether points that were associated with demonstratives differed in form from points that were not. (This analysis could not be carried out in signers because ASL does make regular use of lexical signs that serve to highlight co-occurring points in the same way.) We found that points associated with demonstratives were more likely to involve full arm extension, and were longer in duration, than points not associated with demonstratives. One way to understand these findings is that speakers face a semantic integration problem that signers do not. As discussed, signers' points occur in the same articulatory channel as the majority of the referential content—the hands. Speakers' points, by contrast, occur in a different channel from the majority of the referential content; speakers thus face the task of stitching manual content in with the spoken content when it is critical to do so. Spoken demonstratives support this stitching by signaling that there is critical content in the secondary, manual channel. In turn, fully extending the point and holding it for a longer period of time further supports this stitching by enhancing the gesture's salience. The fact that points associated with demonstratives are longer in duration than points not associated with demonstratives further

underscores the fact that, in speakers, the duration of a pointing gesture is complexly determined by its coordination with speech. Sometimes it may reflect constraints of structural integration—or the *absence* of a same-channel constraint, in this case—and other times it may reflect constraints of semantic integration—the need to make a point salient in order to help stitch it in with speech.

A final set of findings concerned structural integration. In a recent study, we described a same-channel constraint that exerts pressure on pointing signs, but not on pointing gestures (Fenlon et al., 2019). In this earlier work, we did not analyze the mechanics of this constraint in detail, but we identified several hallmarks of pointing signs that might reflect it. Here, we sought to confirm these broader patterns while also delving more deeply into how structural integration actually plays out in speakers and signers. As in our earlier work (Fenlon et al., 2019), we found here that speakers' points were overall much longer in duration than signers' points—in both studies about twice as long, despite a number of differences between the data sets. We also found that signers were much less likely to produce full-extension points than speakers, perhaps reflecting a general economy of effort that the same-channel constraint encourages.

Both of the broader patterns in speakers and signers just mentioned make sense when we look more closely at how these points articulate with other signals. Speakers' points showed a strong tendency to span across neighboring lexical material (spoken words), and occasionally even bridge over adjacent utterances. Although signers' points are free to span across lexical content in this way—and sometimes did—their strong tendency was to slot in between neighboring lexical material (manual signs). The pressure to slot in encourages shorter, less effortful points in signers; the absence of this constraint makes room for longer lasting, more effortful points in gesturers. A question that arises is whether these broad tendencies are intrinsic to these types of communication—spoken vs. signed—or are part of the conventional practices one comes to master. One way to investigate this would be to see whether young speakers and signers show these same strong tendencies from the start, or whether they are gradually acquired like other discourse conventions.

CONCLUSION

Human communication is increasingly understood as composite in nature—as an activity that integrates different types of communicative elements (e.g., Clark, 1996; Enfield, 2009; Vigliocco et al., 2014). These elements include not only the highly conventionalized symbols that have traditionally been the focus of linguistic inquiry, but also more tailored, gradient elements—from depicting signs and ideophones, to size-specifying constructions and pointing gestures. An emerging question within this framework concerns how such elements are integrated into seamless, fluent discourse (e.g., Davidson, 2015; Clark, 2016; Dingemans and Akita, 2017; Lu and Goldin-Meadow, 2018). Here, we used pointing—a ubiquitous and multi-functional act—as a paradigm case. We have attempted to shed light on how pointing is integrated into language—and, moreover, how this

integration differs across spoken and signed communication. Of course, there are a number of aspects of integration that we did not touch on, even for the case of pointing. For example, work remains to be done on how pointing might be coordinated with specific grammatical structures or sequential environments. Work also remains to be done on whether aspects of the integration of pointing into language, such as those described here, generalize—for instance, to other types of pointing (e.g., points to invisible entities; Flack et al., 2018) or to communities with different articulator preferences (e.g., those who rely heavily on non-manual pointing; Cooperrider et al., 2018) and practices (e.g., points to the sun's arc to refer to time of day; Floyd, 2016). Beyond pointing, a host of questions await about the integration of other types of signals into language. Such questions will become more and more central as linguistic inquiry broadens its focus to account for the heterogeneous nature of communication—and how it coheres.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: <https://osf.io/wckx5/> 200601.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Social and Behavioral Sciences Institutional Review Board, University of Chicago. The participants provided their written informed consent to participate in this study.

Written informed consent was obtained from the individuals for the publication of any potentially identifiable images or data included in this article.

REFERENCES

- Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychol. Sci.* 15, 415–419. doi:10.1111/j.0956-7976.2004.00694.x
- Bangerter, A., and Chevalley, E. (2007). “Pointing and describing in referential communication: when are pointing gestures used to communicate?,” in *CTIT proceedings of the workshop on multimodal output generation (MOG)*. Editors I. Van Der Sluis, M. Theune, E. Reiter, and E. Kraemer (Scotland: Aberdeen).
- Bates, D., Mächler, M., Ben, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67 (1), 1–48. doi:10.18637/jss.v067.i01
- Bates, E. (1979). *The emergence of symbols: cognition and communication in infancy*. New York: Academic Press.
- Brentari, D., González, C., Seidl, A., and Wilbur, R. (2011). Sensitivity to visual prosodic cues in signers and nonsigners. *Lang. Speech* 54, 49–72. doi:10.1177/0023830910388011
- Capirci, O., and Volterra, V. (2008). Gesture and speech: the emergence and development of a strong and changing partnership. *Gesture* 8 (1), 22–44. doi:10.1075/gest.8.1.04cap
- Cartmill, E., Hunsicker, D., and Goldin-Meadow, S. (2014). Pointing and naming are not redundant: children use gesture to modify nouns before they modify nouns in speech. *Dev. Psychol.* 50 (6), 1660–1666. doi:10.1037/a0036003
- Clark, H. H. (1996). *Using language*. (Cambridge: Cambridge University Press).
- Clark, H. H. (2003). “Pointing and placing,” in *Pointing: where language, culture, and cognition meet*. Editor S. Kita (Mahwah, NJ: Lawrence Erlbaum), 243–268.

AUTHOR CONTRIBUTIONS

All authors conceived of the research questions and study design. JF supervised the data collection, with input from all authors. KC and JF supervised the gesture and language coding, with input from all authors. JK conducted all statistical analyses, with input from KC. KC drafted the manuscript, with critical revisions provided by all authors. All authors approved the final version of the manuscript.

FUNDING

This article was supported by a grant from the Neubauer Collegium at the University of Chicago, awarded to Diane Brentari and Susan Goldin-Meadow.

ACKNOWLEDGMENTS

We gratefully acknowledge support for this research from the Neubauer Family Collegium for Culture and Society and the Center for Gesture Sign and Language at the University of Chicago. We also thank Averí Ayala and Zena Levan for research assistance and three reviewers for their valuable feedback.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomm.2021.567774/full#supplementary-material>.

- Clark, H. H. (2016). Depicting as a method of communication. *Psychol. Rev.* 123 (3), 324–347. doi:10.1037/rev0000026
- Cooperrider, K. (2016). The co-organization of demonstratives and pointing gestures. *Discourse Process.* 53 (8), 632–656. doi:10.1080/0163853X.2015.1094280
- Cooperrider, K. (2017). Foreground gesture, background gesture. *Gesture* 16 (3), 176–202. doi:10.1075/gest.16.2.02coo
- Cooperrider, K., Slotta, J., and Núñez, K. (2018). The preference for pointing with the hand is not universal. *Cogn. Sci.* 42 (4), 1375–1390. doi:10.1111/cogs.12585
- Cormier, K., Schembri, A., and Woll, B. (2013). Pronouns and pointing in sign languages. *Lingua* 137, 230–247. doi:10.1016/j.lingua.2013.09.010
- Cumming, G. (2014). The new statistics: why and how. *Psychol. Sci.* 25 (1), 7–29. doi:10.1177/0956797613504966
- Cumming, G., and Finch, S. (2005). Inference by eye: confidence intervals and how to read pictures of data. *Am. Psychol.* 60 (2), 170–180. doi:10.1037/0003-066X.60.2.170
- Davidson, K. (2015). Quotation, demonstration, and iconicity. *Linguistics Philos.* 38 (6), 477–520. doi:10.1007/s10988-015-9180-1
- Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cogn. Linguistics* 17 (4), 463–489. doi:10.1515/COG.2006.015
- Dingemans, M., and Akita, K. (2017). An inverse relation between expressiveness and grammatical integration: on the morphosyntactic typology of ideophones, with special reference to Japanese. *J. Linguistics* 53 (3), 501–532. doi:10.1017/S00222671600030X
- Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., and Paolino, D. (1993). Outline of discourse transcription. In Edwards & Lampert (Eds.), *Talking data: transcription and coding in discourse research* (pp. 45–48). Hillsdale, NJ: Erlbaum.

- Eibl-Eibesfeldt, I. (1989). *Human ethology*. New York: Aldine de Gruyter. doi:10.1017/S0140525X00060416
- Enfield, N. J., Kita, S., and de Ruiter, J. P. (2007). Primary and secondary pragmatic functions of pointing gestures. *J. Pragmatics* 39 (10), 1722–1741. doi:10.1016/j.pragma.2007.03.001
- Enfield, N. J. (2009). *The anatomy of meaning: speech, gesture, and composite utterances*. Cambridge: Cambridge University Press. doi:10.1111/j.1467-9655.2010.01661_28.x
- Fenlon, J., Cooperrider, K., Keane, J., Brentari, D., and Goldin-Meadow, S. (2019). Comparing sign language and gesture: insights from pointing. *Glossa*. 4 (1), 1–26. doi:10.5334/gjgl.499
- Fenlon, J., Denmark, T., and Campbell, R. (2007). Seeing sentence boundaries. *Sign. Lang. Linguistics* 10 (2), 177–200. doi:10.1075/sll.10.2.06fen
- Fenlon, J., Schembri, A., and Cormier, K. (2018). Modification of indicating verbs in British Sign Language: a corpus-based study. *Lang.* 94 (1), 84–118. doi:10.1075/sll.10.2.06fen
- Ferrara, L., and Hodge, G. (2018). Language as description, indication, and depiction. *Front. Psychol.* 9, 1–15. doi:10.3389/fpsyg.2018.00716
- Flack, Z. M., Naylor, M., and Leavens, D. A. (2018). Pointing to visible and invisible targets. *J. Nonverbal Behav.* 42 (2), 221–236. doi:10.1007/s10919-017-0270-3
- Floyd, S. (2016). Modally hybrid grammar? Celestial pointing for time-of-day reference in Nheengatú. *Lang.* 92 (1), 31–64. doi:10.1353/lan.2016.0013
- Gelman, A., and Tuerlinckx, F. (2000). Type S error rates for classical and Bayesian single and multiple comparison procedures. *Comput. Stat.* 15, 373–390. doi:10.1007/s001800000040 10.1007/s001800000040
- Goldin-Meadow, S., and Brentari, D. (2017). Gesture, sign and language: the coming of age of sign language and gesture studies. *Behav. Brain Sci.* 1, 1–82. doi:10.1017/S0140525X15001247
- Johnston, T. (2013). Formational and functional characteristics of pointing signs in a corpus of Auslan (Australian Sign Language): are the data sufficient to posit a grammatical class of “pronouns” in Auslan?. *Corpus Linguistics Linguistic Theor.* 9 (1), 109–159. doi:10.1515/clt-2013-0012
- Kendon, A. (2004). *Gesture: visible action as utterance*. Cambridge: Cambridge University Press.
- Kita, S. (2003a). “Pointing: a foundational building block of human communication,” in *Pointing: where language, culture, and cognition meet*. Editor S. Kita (Mahwah, NJ: Lawrence Erlbaum), 1–8.
- S. Kita (Editor) (2003b). *Pointing: where language, culture, and cognition meet*. Mahwah, NJ: Lawrence Erlbaum.
- Lausberg, H., and Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behav. Res. Meth. Instrum. Comp.* 41 (3), 841–849. doi:10.3758/BRM.41.3.841
- Liszkowski, U. (2006). “Infant pointing at 12 months: communicative goals, motives, and social-cognitive abilities,” in *Roots of human sociality*. Editors N. J. Enfield and S. C. Levinson (Oxford, England: Berg Publishers), 153–178.
- Liszkowski, U., Brown, P., Callaghan, T., Takada, A., and de Vos, C. (2012). A prelinguistic gestural universal of human communication. *Cogn. Sci.* 36, 698–713. doi:10.1111/j.1551-6709.2011.01228.x
- Lu, J., and Goldin-Meadow, S. (2018). Creating images with the stroke of a hand: depiction of size and shape in sign language. *Front. Psychol.* 9 (1276), 1–15. doi:10.3389/fpsyg.2018.01276
- McBurney, S. L. (2009). “Pronominal reference in signed and spoken language: are grammatical categories modality-dependent?,” in *Modality and structure in signed and spoken languages*. Editors R. P. Meier, K. Cormier, and D. Quintos-Pozos (Cambridge: Cambridge University Press), 329–369.
- McNeill, D. (1992). *Hand and mind: what gestures reveal about thought*. Chicago: University of Chicago Press.
- Meier, R. P., and Lillo-Martin, D. (2003). The points of language. *Humana. Mente J. Philosoph. Stud.* 24, 151–176.
- Nespor, M., and Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris Publications. doi:10.1007/978-94-009-2733-9_16
- Peeters, D., Chu, M., Holler, J., Hagoort, P., and Özyürek, A. (2015). Electrophysiological and kinematic correlates of communicative intent in the planning and production of pointing gestures and speech. *J. Cogn. Neurosci.* 27 (12), 2352–2368. doi:10.1162/jocn_a_00865
- Peirce, C. S. (1940). *Philosophical writings of Peirce*. Editor J. Buchler (New York: Dover).
- R Core Team (2016). R: a language and environment for statistical computing. Vienna, Austria. Available at: <https://www.R-project.org/>.
- Sandler, W., Meir, I., Padden, C., and Aronoff, M. (2005). The emergence of grammar: systematic structure in a new language. *Proc. Natl. Acad. Sci. USA* 102 (7), 2661–2665. doi:10.1073/pnas.0405448102
- Vigliocco, G., Perniss, P., and Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* 369, 20130292. doi:10.1098/rstb.2013.0292

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Cooperrider, Fenlon, Keane, Brentari and Goldin-Meadow. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.