COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 45 (2021) e13061

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13061

Mapping Word to World in ASL: Evidence from a Human Simulation Paradigm

Allison Fitch,[a,b] Sudha Arunachalam,[c] Amy M. Lieberman[a]

[a]*Deaf Education and Deaf Studies, Boston University*
[b]*Psychology, Rochester Institute of Technology*
[c]*Communicative Sciences and Disorders, New York University*

Received 18 December 2020; received in revised form 27 September 2021; accepted 5 October 2021

Abstract

Across languages, children map words to meaning with great efficiency, despite a seemingly unconstrained space of potential mappings. The literature on how children do this is primarily limited to spoken language. This leaves a gap in our understanding of sign language acquisition, because several of the hypothesized mechanisms that children use are visual (e.g., visual attention to the referent), and sign languages are perceived in the visual modality. Here, we used the Human Simulation Paradigm in American Sign Language (ASL) to determine potential cues to word learning. Sign-naïve adult participants viewed video clips of parent–child interactions in ASL, and at a designated point, had to guess what ASL sign the parent produced. Across two studies, we demonstrate that referential clarity in ASL interactions is characterized by access to information about word class and referent presence (for verbs), similarly to spoken language. Unlike spoken language, iconicity is a cue to word meaning in ASL, although this is not always a fruitful cue. We also present evidence that verbs are highlighted well in the input, relative to spoken English. The results shed light on both similarities and differences in the information that learners may have access to in acquiring signed versus spoken languages.

Keywords: American Sign Language; Human Simulation Paradigm; Iconicity; Word learning

1. Introduction

Language acquisition researchers have extensively investigated the mapping problem: How do infants take arbitrary word forms from the input and map each onto the correct meaning, despite a vast hypothesis space of possible meanings? This research has uncovered a variety of cues that children may employ, including joint attention (e.g., Baldwin, 1991; Tomasello

Correspondence should be sent to Allison Fitch, Boston University, 18 Lomb Memorial Drive, Rochester, NY 14623, USA. E-mail: ahfgsh@rit.edu

& Farrar, 1986) and iconicity (e.g., Imai & Kita, 2014; Imai, Kita, Nagumo, & Okada, 2008; Perry, Perlman & Lupyan, 2015). These cues typically increase referential transparency; that is, they make the mapping between form and meaning more transparent. For example, if the child and interlocutor share attention on an object while the interlocutor labels it, the child is more likely to learn the label than if the interlocutor labels an object that is not currently present in the room (Akhtar et al., 1991; Baldwin, 1993; Tomasello & Farrar, 1986). Although this body of work has been extremely influential to our understanding of language acquisition, these studies have investigated the acquisition of spoken languages (primarily English). These cues may not be equally applicable across languages or modalities, particularly with regard to cues that rely on visual attention. The research presented here addresses this gap by providing evidence that word learning in American Sign Language (ASL) can be supported by cues that are shared with those that support word learning in a spoken language, but that there are important differences that are unique to the modality and structure of ASL.

One cue we consider is iconicity, or the degree to which a word resembles its referent. For example, onomatopoeic words (e.g., choo choo, oink) provide referential transparency because they sound like what they reference; children typically acquire these words early (Frank, Braginsky, Yurovsky & Marchman, 2016; Laing, 2019; Motamedi et al., 2020; Nielsen & Dingemanse, 2020; Perry et al., 2015). While all languages contain some degree of iconicity, the visual nature of sign languages allows for more information to be communicated through iconic means (Perniss, Thompson & Vigliocco, 2010). Further, iconic word forms in sign languages may be particularly transparent because they directly resemble the visual properties of their referents (a feature that is not available to spoken languages). Thus, one might expect children acquiring a sign language like ASL to rely more heavily on iconicity to learn words than children acquiring spoken languages.

The degree to which children can actually use the referential transparency afforded by iconicity is a matter of some debate. It is relatively uncontroversial that cross-linguistically, iconic words appear earlier in vocabulary than noniconic words. Indeed, recent evidence from sign language acquisition work suggests that deaf children acquire signs with particular iconic properties early (Caselli & Pyers, 2017; 2020; Ortega et al., 2017; Thompson et al., 2012). Additionally, theorists seem to agree that iconicity promotes the acquisition of iconic word forms (a so-called "local learning effect;" Nielsen & Dingemanse, 2020). Within the visual modality, a proposed mechanism for this comes from one study on British Sign Language demonstrating that deaf parents tend to modify iconic signs in ways that highlight the shared features between the sign and its referent (Perniss, Lu, Morgan & Vigliocco, 2018).

However, the use of iconicity as a cue is more difficult to reconcile given that ASL is not a wholly iconic language and there are many signs in children's early vocabularies that are not iconic (Anderson & Reilly, 2002). Additionally, an iconic strategy may be limited even for iconic words. The resemblance between an iconic sign and its referent may not always be accessible, and may be context-driven. According to structure-mapping theory, in order to leverage iconicity, learners must be able to identify the similarities between the sign and its referent (e.g., Emmorey, 2014, Taub, 2001). Iconic signs sometimes resemble only certain aspects of their referents (e.g., one sign for CHICKEN in ASL resembles a beak; Fig. 1). In order to "see" that resemblance, those features must be in one's mental

Fig 1. An example of an ASL sign for CHICKEN. Image drawn from the ASL Signbank (ASL Signbank, 2020).

representation of the referent. Thus, the sign CHICKEN may not be as referentially transparent in the context of a chicken dinner as it is in the context of a henhouse. This theory has been used to explain a peculiar finding: Although adult nonsigning participants perform better at learning and remembering iconic isolated signs relative to arbitrary ones (Baus et al., 2013; Morett, 2015; Ortega & Morgan, 2015), they are typically not successful at guessing the meanings of iconic signs in isolation (Griffith, Robinson & Panagos, 1981; c.f. Ortega et al., 2019). Being able to "see" the resemblance between the sign and referent is potentially even more difficult for children, who may lack a robust mental representation for a newly learned concept.

Iconicity in and of itself may not fully explain the earlier acquisition of iconic signs. It may be the case that signs that are particularly amenable to iconic productions are also high in concreteness (Caselli & Pyers, 2020). Concrete (in contrast to abstract) referents tend to be easy to observe and thus may have a higher degree of referential transparency. Concreteness also may be less contextually variable than iconicity: the referent of the concrete sign CHICKEN is equally easy to observe in chicken dinners and henhouses, while the referent of the abstract sign WORRY is not easily observed in most contexts. Across spoken languages, words for concrete referents are indeed acquired earlier than abstract ones (Braginsky, Yurovsky, Marchman & Frank, 2017). However, recent work on ASL suggests that children acquire iconic forms earlier, even when controlling for concreteness (Caselli & Pyers, 2020). Therefore, despite that the evidence surrounding the role of iconicity in learning ASL is nuanced, we expect that overall, the iconicity of a particular form should predict how easily it is learned.

A second set of cues that increases referential transparency involves the extralinguistic situation, including copresence of, and attention to, the referent being labeled. Object presence is the minimum criterion for joint attention, which is consistently demonstrated to be an important cue for word learning (e.g., Tomasello & Farrar, 1986). Moreover, a present object can be pointed to or held; these manual cues can direct attention to an object and have sometimes been shown to be more important than gaze cues for word learning when participating in a joint attention bout (Yu & Smith, 2015, 2017). The timing of these cues is important as well. Mutual attention to an object between a child and an interlocutor at the moment that object is labeled provides complete overlap between the visual referent and its label. Language learners are sensitive to this overlap, and are less successful at word learning if the timing of these cues is offset even slightly (Trueswell et al., 2016). Overlap reduces the uncertainty of the interlocutor's intended referent (Akhtar & Tomasello, 2010; Brooks & Meltzoff, 2005), and it may also highlight that something about the object is interesting and worth attending to (Yu & Smith, 2016; Yu, Suanda & Smith, 2019).

However, these extralinguistic cues look different during a signed interaction relative to a spoken interaction. The overlap of mutual attention and labeling that is so helpful for hearing children is counter-productive in ASL. Children are unable to simultaneously gaze at an object and receive linguistic input from an interlocutor through the visual modality. Instead, they must shift their gaze or locus of attention (depending on how close the interlocutor is) between the interlocutor and the object (Lieberman, Hatrak & Mayberry, 2014), and the timing of mutual attention relative to labeling must therefore be asynchronous. Likewise, because ASL is primarily articulated with the hands, object manipulation (holding and touching) may occur asynchronously with label production. Thus, in order for joint attention to support word learning in ASL, the child's interlocutor must be sensitive to this asynchrony, either by waiting for a child to make a gaze shift or by using an attention-getting strategy (e.g., touching the child), before providing linguistic information. Indeed, deaf infants with native ASL exposure have recently been shown to have enhanced gaze following abilities and an increased number of shifts between an object and interlocutor relative to hearing infants (Brooks, Singleton & Meltzoff, 2020). However, it is not currently known what the optimal timing dynamics of this interaction are.

These two types of cues—iconicity and extralinguistic cues such as object presence—may play different roles for learning different kinds of word meanings. For example, object presence and object manipulation may be most helpful for acquiring nouns that label concrete objects, as compared to other kinds of words, because their referents are more likely to be static. Indeed, nouns labeling concrete objects are typically acquired earlier than predicates, and extralinguistic cues like joint attention are likely to be at least one reason why this is so (e.g., Piccin & Waxman, 2007). However, the relative distribution of nouns to predicates in early vocabulary varies cross-linguistically (e.g., Frank et al., 2020; Tardif, 1996; Tardif et al., 2008). Mandarin, for example, shows less of a noun bias in early vocabulary than English (Tardif et al., 2008). It is believed that for Mandarin-acquiring infants, verbs may be highlighted well in the input via extralinguistic cues, and thus as referentially transparent as nouns (Snedeker, Yuan & Li, 2003; cf. Tardif et al., 1997). Of note for the current study, while children's early vocabularies in ASL are primarily composed of nouns, they contain proportionally more predicates than early English vocabularies (Anderson & Reilly, 2002). A reasonable hypothesis, then, is that extralinguistic cues may support verb learning in ASL in a way that is not supported for English. Alternatively, though, verbs may be better highlighted through solely linguistic means like iconicity, because some iconic forms for verbs are mimetic gestures of the labeled action. As of yet, although we have reason to believe that verbs may be relatively easy to acquire in ASL based on Anderson and Reilly's (2002) results, we do not know why this might be the case.

To summarize, we expect that features of interactions that lead to referential transparency in ASL will be different from those in spoken languages in a few key ways. We expect that a higher degree of iconicity, where visual word features represent visual referent features, increases referential transparency. We also expect that while some extralingustic cues are helpful, the temporal dynamics of those cues will be different in ASL than in spoken language, most notably that children's attention to an object and parent labeling will occur in sequence rather than simultaneously. We further expect that ASL may be unlike English (but like Mandarin) in privileging verbs such that situations in which verbs are uttered are equally referentially transparent as situations in which nouns are uttered. The goal of the current study is to test these hypotheses.

1.1. Current study: An adapted human simulation paradigm for ASL

One way to test a word's referential transparency is through the Human Simulation Paradigm (HSP; Gillette et al., 1999, for a review see Trueswell & Gleitman, 2018). This paradigm asks adult participants to watch vignettes of interactions in which a mystery word is used, and they are asked to guess the identity of the mystery word. The key manipulation is that their access to the linguistic signal is blocked (i.e., the sound is off for spoken languages), so their guesses are only informed by observation of the extralinguistic context. Some vignettes result in much higher guessing accuracy than others. Vignettes in which most participants make accurate guesses can be foraged for cues in the extralinguistic context that may have helped inform their decision—in other words, cues that made the form-meaning mapping more transparent. Thus, referential transparency in this paradigm is operationalized as guess accuracy. This paradigm also allows us to examine multiple hypotheses about referential transparency at one time.

The HSP is not a perfect simulation of language learning. Adult participants come into the task with world knowledge and experiences that young children do not have in the language acquisition process. Similarly, language-acquiring children have access to information outside of the vignette that they may use to inform the word-learning process (e.g., objects that are just out of frame, events that occurred immediately before the vignette). Still, HSP findings have largely mirrored findings from direct studies of children's word learning, including the importance of joint attention and other social cues (Trueswell et al., 2016), better accuracy for nouns than other word types in English (Gillette et al., 1999; Snedeker et al., 2003), and improved performance given information about the linguistic context in which the word occurred (Gillette et al., 1999). Additionally, preschoolers and school-aged children perform similarly to adults on these tasks (Medina, Snedeker, Trueswell & Gleitman, 2011; Piccin & Waxman, 2007). Thus, the HSP paradigm is a fruitful approach to explore whether or not a given hypothesis about children's word learning may be worthy of further study.

Our understanding of referential transparency in ASL is still in its infancy, but has strong theoretical and applied implications. Disentangling the cues that promote referential transparency in ASL will provide both a better understanding of word-meaning mapping across modalities, and a snapshot of what successful word learning looks like in ASL. This is particularly important for the language acquisition of deaf children born into hearing families. These children comprise 95% of the population of deaf children (Mitchell & Karchmer, 2004) and are at increased risk for language deprivation, given that their families lack sign language experience. Thus, knowledge of what leads to the greatest referential clarity is critical for promoting word learning in these children.

To that end, the research presented here describes two studies using the HSP in ASL in order to answer the following research questions:

1. Does iconicity enhance referential transparency (i.e., guess accuracy) in ASL? ASL is more iconic than spoken English, but individual signs in ASL vary in iconicity. We predict that iconicity will enhance referential transparency in naturalistic interactions, as it sometimes does with isolated signs (Ortega et al., 2019).
2. What is the role of extralinguistic cues in promoting referential transparency? We consider the presence of referent objects and event, object manipulation, and the timing of these cues relative to the linguistic signal.
3. Are certain lexical classes of ASL signs more referentially transparent than others? Specifically, we ask if nouns are more referentially transparent than verbs, as they are in spoken English (Gillette et al., 1999; Snedeker et al., 2003), or whether verbs are more referentially transparent than nouns, as they appear to be in Mandarin (Snedeker et al., 2003).

In order to investigate naturally occurring interactions in typically developing children, the interactions that served as the basis for the study involved native-signing deaf children interacting with their deaf parents. In Study 1, we focus on iconicity, and explore this question at the level of the sign, asking if particular signs are more referentially transparent in their form-meaning mappings than others. In Study 2, we examine the properties of each naming event in more detail, looking for both extralinguistic cues (e.g., object presence for nouns or event presence for verbs) that make particular naming events more referentially transparent than others.

2. Study 1

The primary goal of Study 1 was to examine the role of a linguistic cue, namely iconicity, on form-meaning mapping. Hearing, nonsigning adults participated in an HSP involving videos of deaf children interacting with their deaf parents in ASL. We coded the adults' accuracy in guessing the signs, comparing their performance on nouns and verbs, and examining whether iconicity predicted success.

Because HSPs have historically focused on spoken language, we modified the task to be appropriate to both the modality and our research questions. Unlike in spoken language HSP studies, in which the language can be made inaccessible to the participant by simply muting the sound while keeping extralinguistic information relatively intact, we could not block access to the ASL signal without also blocking access to the very cues that might support success in the task. In ASL, linguistic information is encoded in the hand shape, palm orientation, location, movement, and nonmanual markers of a sign, which involve not only the signer's hands but also facial expressions and eye gaze. Thus, blurring or otherwise visually blocking the sign would mean preventing access to gaze and object manipulation cues, as well as iconicity. Further, the location of the sign cannot be blocked without blocking out most of the signer's body, making the task extremely unnatural in addition to restricting access to potentially important cues. To circumvent this problem, we recruited hearing, nonsigning participants. These participants were able to observe the full vignette, including potentially important referential cues, but did not have access to the linguistic input, even though it was fully visible to them. This in some ways resembles the experience of a language-learning child: the linguistic signal is still present, but they are not necessarily able to forage it for information.

Another modification we made to the traditional HSP is that we collected qualitative data about the participant's strategy in the form of justifications. By design, the HSP makes several assumptions about the kinds of strategies participants (and by proxy, children) might use to determine the meanings of words (e.g., iconicity). These strategies are likely to vary in their efficacy across all words and word-to-word. For example, drawing on iconicity to determine word meanings may lead to an overall increase in correct word mappings, but it is also a better strategy for more iconic words than less iconic words. However, naïve participants are unlikely to draw on the same strategy on every trial and may employ different strategies than we hypothesize. We therefore periodically asked them to explain how they arrived at their response. These justifications allowed us to determine if particular strategies were more successful than others, or if participants used strategies outside of what we would expect (linguistic and/or extralinguistic cues). They also allowed us to see whether or not participants drew on strategies that could not be available to young children (e.g., preconceived notions or biases about sign languages).

2.1. Method

2.1.1. Participants

We recruited 36 adults aged 18–29 years (29 female, seven male). All participants were undergraduate or graduate students and were fluent English speakers. All participants reported that they did not have any ASL knowledge or experience and none reported experience with any other sign language. One participant was excluded from analysis for contributing too few trials (<25% of total trials) and two participants were unable to complete the study due to technological failure. The final sample size was 33 adults (27 female, six male).

2.1.2. Apparatus and stimuli

Stimuli were 189 40-s video clips extracted from recordings of standardized dyadic play interactions (Fieldsteel, Bottoms, & Lieberman, 2020) between seven deaf children (18–42 months of age) and their five deaf parents (all mothers; two sibling pairs participated in the play interactions, but each sibling had a separate interaction). During these interactions parents were instructed to play with the toys and interact with their child as they typically would, and were not given any more specific or additional instructions. Dyads played with the same set of toys that included a picnic set (plastic miniature food, utensils, and tablecloth), a school bus, a set of miniature human figures, a baby doll with accessories for feeding and bathing, a camping set (an empty canteen, lantern, and binoculars), a Magna Doodle, and three books. During all interactions, communication occurred in ASL and was transcribed by native or fluent ASL signers. Maternal utterances were then analyzed for type frequency across dyads, and sign types were ranked from most to least frequent. The target signs selected for the study were the 24 most frequent nouns and 24 most frequent verbs used in the child-directed signing across these interactions, with the additional constraint that the sign must have been produced in at least two dyads, to reduce idiosyncrasies associated with a particular interaction. As some nouns and verbs are phonologically similar in ASL, judgments about which lexical category a sign belonged to were made for each individual production by native signers, who had access to the full context of the utterance, during the initial corpus transcription. Disagreements among coders were managed through group discussion (for further details, see Fieldsteel et al., 2020). The final 48 target signs are listed in Table 1, and still images of the target signs are in Figure S1. Thirty-seven of the target signs are on the ASL-CDI (Anderson & Reilly, 2002), and all of them are reportedly understood by 21 months of age according to CDI norms. Of the 24 target nouns, 14 referred to objects with a visible referent in the toy set (i.e., it was one of the toys, was part of one of the toys, or described a subset of toys). All verbs could be associated with at least one, but often several different referents within the toy set (e.g., driving could be associated with the bus or car; sleeping could be associated with the baby or dinosaur).

Individual vignettes were created by identifying every production of a target sign in the series of play interactions, and then extracting the 30 s prior to the production's onset, and 10 s after the production's onset, in line with prior HSP studies (Cartmill et al., 2013; Gillette et al., 1999; Medina et al., 2011; Snedeker et al., 1999). To signal the sign's onset, a beep sound was added to the video and aligned with the sign's onset and offset. In the event that multiple productions of the target sign occurred in the 40 s, additional beeps were added.

For each target sign, four vignettes (representing four individual productions) were randomly selected for use in the study. Each of the four vignettes was assigned to a stimulus list, yielding four lists, each containing a unique production of each of the 48 target signs. There were three exceptions to this: MAN, PHONE, and WATER. For these signs, only three

Table 1
Target signs used in Studies 1 and 2. Nouns marked with a * are visible referents in the play session toy set

Nouns	Verbs
BABY*	CAN
BATTERIES	CAN'T
BIRD*	CLEAN
BOOK*	COME
BUS*	CRY
BUS(FS)*	DO-DO
CAR	DON'T-WANT
CHICKEN*	DRAW
COLOR	DRINK
DINOSAUR*	DRIVE
DOOR	EAT
DUCK*	ERASE
EYE*	cl(FLASHING-LIGHT)
FOOD*	GOix
HAMBURGER*	HAVE
HOTDOGs*	HELP
LIGHT*	LOOK
MAN*	OPEN
MOM	PLAY
NAPKIN	SEE
PHONE	SIT
SCHOOL	SLEEP
STORY	STOP
WATER	WANT

productions could be identified that were sufficient for use in the study. Thus, two lists shared the same production of each of these signs—sharing was balanced across lists. Vignettes were arranged across lists such that no particular dyad was more strongly represented than another in any given list. Vignettes did not contain any observable mouthing of English translations.

The paradigm was programmed in PsychoPy 2 (Peirce et al., 2019) and displayed on an iMac with a 27-inch screen.

2.1.3. Procedure

After providing written consent, participants were given an online demographic survey that asked about their hearing status, ASL experience, and language background. Participants were then assigned to a list (based on the order in which they participated) and presented with the HSP.

The participants read on-screen instructions in which they were told they would watch a video of a parent–child interaction in ASL. During this video, they would hear a beep corresponding to a target sign. After the video they were to type in the meaning of that target sign. To ensure that participants were treating the task as a linguistic one, they were told that ASL is a full, complex language like English, that signs can denote both single and compound

Fig 2. Schematic of trials in Studies 1 and 2. Images have been altered from their original format to protect the dyad's confidentiality.

words, and that all beeps within a video referred to the same sign.[1] They were then presented with the same practice trial using a video clip where the target sign was THANK-YOU. Finally, they were told that after some video clips they would be asked to explain why they provided the response that they did.

Video clips appeared on a grey screen in a random order. After each video clip they were asked to type their response at the prompt "*Meaning?*" which was printed in white font in the center of the screen (Fig. 2). Although participants could advance to the next trial themselves, the program automatically advanced to the next trial after 30 s. This was to ensure that participants were responding with their first thought and not second-guessing themselves. To better understand participants' strategies, on every sixth clip they were prompted to justify their answer with the prompt "*Explain*" in the same white font; this was not time limited.

2.1.4. Analysis

Participants who completed the experiment contributed a minimum of 12 and a maximum of 48 codable sign meanings and eight justifications. Sign meaning responses were excluded if participants did not provide an answer or said that they did not know ($n = 20$, 1.3% of responses) or if the response contained greater than one content morpheme (e.g., if they provided a phrase; $n = 239$, 15% of responses). Remaining responses ($n = 1{,}325$) were coded (by AF) as correct or incorrect, where correct was defined as the exact transcription for the target sign or an acceptable equivalent meaning as judged by a native ASL signer. Incorrect responses were further coded for semantic relatedness to the target (e.g., the response “food” for EAT), and syntactic relatedness, or belonging to the same word class as the target (i.e., a verb response for a verb sign or a noun response for a noun sign). All responses were identified as a noun, a verb, or neither. Responses that could be interpreted as either a noun or a verb were coded based on the most frequent usage based on the iWeb corpus (Davies, 2018). All sign meaning responses were double coded by a second trained coder. Agreement was calculated at 96%.

2.2. Results

The focus of Study 1 was on the role of iconicity in referential transparency. Specifically, we analyze the role of iconicity in overall accuracy, as well as the interaction between iconicity and word type (i.e., nouns vs. verbs). Accuracy, iconicity, concreteness, and word class data for each sign are available in Table S1.

2.2.1. Iconicity

One way that iconicity may play a role in referential transparency is in participants’ overall accuracy. Because ASL contains a higher degree of iconicity than spoken languages, and uses the visual modality to represent visual features, participants may be overall quite successful at mapping form to meaning. In order to determine overall accuracy, each participant received a score of the percent of correct responses they gave out of the number of coded (i.e., not excluded) responses. For each participant, we tallied an overall percentage, as well as percentages for the subset of noun targets and the subset of verb targets. On average, participants accurately guessed the meaning of the target sign just 14.29% of the time ($SD = 9.07\%$). This is in line with the accuracy levels in prior work from English (Gillette et al., 1999) and Mandarin (Snedeker et al., 2003). Among the incorrect responses, 19.18% ($SD = 12.75\%$) were semantically related to the target sign. Phonologically similar signs (BIRD/CHICKEN, DOOR/OPEN, FOOD/EAT, CAR/DRIVE) were not any more or less accurate than signs that are phonologically distinct (see Table S1).

To determine if accuracy was related to the iconicity of a given sign, we calculated the percentage of correct responses on each target sign (these ranged from 0% to 68.7%) and correlated these with iconicity ratings from ASL-LEX (Caselli, Sehyr, Cohen-Goldberg & Emmorey, 2017; Sehyr, Caselli, Cohen-Goldberg & Emmorey, 2021), which were available for 42 of the 48 target signs. In brief, ASL-LEX iconicity ratings were obtained from a sample of hearing nonsigners[2] who rated sign productions on a scale of 1 (does not look like what it means) to 7 (looks very much like it means). A Spearman’s correlation (Fig. 3) revealed that accuracy was indeed positively correlated with iconicity overall ($\rho = .598$, $p < .001$) and for both nouns ($\rho = .48$, $p = .02$) and verbs ($\rho = .54$, $p = .01$) separately. We conducted a similar analysis using concreteness ratings, which we obtained for 46 of the target signs (Brysbaert, Warriner & Kuperman, 2014; see also Caselli & Pyers, 2020). Brysbaert et al. (2014) instructed participants to rate English words on a scale of 1 (abstract; cannot be experienced directly) to 5 (concrete; exists in reality). Concreteness was not correlated with accuracy ($\rho = -.09$, $p = .55$).

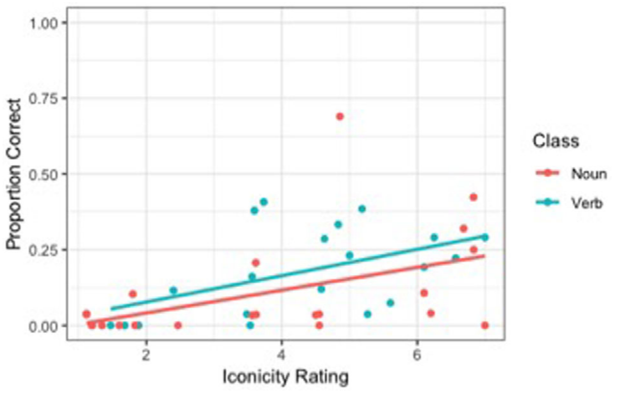

Fig 3. Correlation between iconicity ratings and accuracy by item in Study 1.

2.2.2. Word class

In order to determine if certain kinds of signs were more likely to be guessed than others, we compared participant accuracy by word class (nouns and verbs). Accuracy differed significantly by word class, with noun targets less accurate than verb targets (noun accuracy = 10.2%, verb accuracy = 19.5%): $t(32) = -3.97$, $p < .001$ (Fig. 4).

Among the incorrect guesses, we were interested in whether participants were more likely to make guesses that were in the same word class as the target (i.e., more likely to name a noun for a noun target). Indeed, of the incorrect responses, 37.47% ($SD = 9.41\%$) belonged to the same word class as the target sign, even though participants' responses could come from any word class (i.e., not just nouns and verbs). This tendency to map a meaning from the same word class as the target disproportionately advantaged verbs. Of the 663 verb trials, participants gave a verb response 68.6% of the time (455 trials). Of the 662 noun trials, participants gave a noun response only 23.4% of the time (155 trials). The majority of incorrect responses to noun trials (77%) were verbs. Incorrect responses that were from other word classes were adjectives, adverbs, and interjections.

Overall, regardless of accuracy, participants were much more likely to guess that a particular target was a verb than a noun. For all trials, the average participant provided a noun response on 17.85% of the trials ($SD = 11.81\%$), and a verb response on 64.04% of the trials ($SD = 13.99\%$), despite the fact that the targets were 50% nouns and 50% verbs.

2.2.3. Open-ended responses

Participants were asked to provide justifications for their answers eight times throughout the experiment. Only justifications from included trials were analyzed. Further, several justification responses were excluded for being uninformative—these included intuitions (e.g., "Just a feeling") and statements that restated the guess without providing an explanation (e.g., "The mother said X" despite the fact that the participant did not have access to the parent's language). The remaining 182 responses were coded for the kind of cue used in generating the response. Prior to coding, we established categories of cues based on strategies

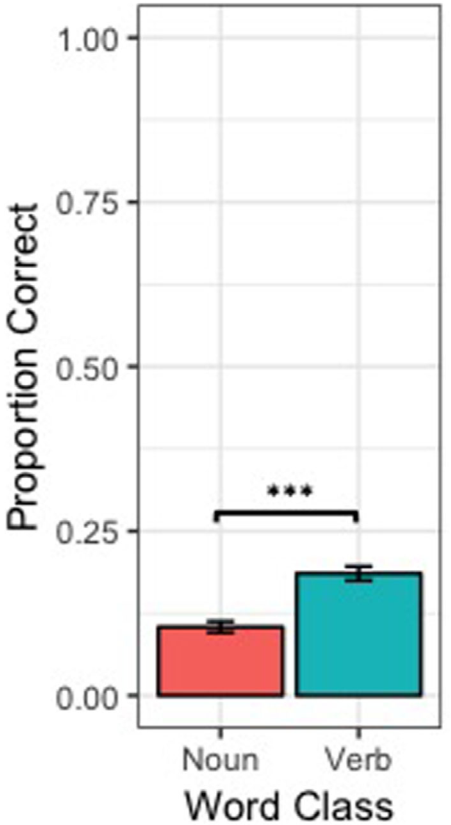

Fig 4. Proportion of accurate responses for nouns and verbs in Study 1. Error bars represent standard error.

participants might use: cues that draw on the iconicity of the sign, on the referential cues in the interaction such as eye gaze and pointing, and on the general context of the scene (e.g., which toys were being played with). Because we were interested in determining if sign and cue overlap mattered, we established categories for the relative timing of the context cue (before, during, or after the sign). Coding was not mutually exclusive; justifications were included in all categories that fit. Full details are in Table 2.

The most frequent type of response (32.8%) were those that provided an iconic explanation, meaning that they indicated that the sign looked like the object or action they provided in their response. However, on the trials in which this type of explanation was given, responses were correct just 19% of the time. When participants' explanations included timing information, most (60.7%) of them drew on context that occurred *after* the sign. This is interesting, given that only 10 of the 40 s of video context occurred after the sign itself. Like iconicity, this strategy was seemingly not any more helpful than the others—only 15% of the trials that used this type of explanation were associated with a correct response.

Table 2
Coded response justifications provided by participants in Study 1

Response Type	Example	Responses (n = 182)	% correct glosses
Iconicity	Looked like motion of holding a steering wheel	69	19% (13/69)
Joint attention	She pointed to the bus	53	13% (7/53)
Nonspecific context clue	[There was] food and woman wiping her mouth on her sleeve	11	9% (1/11)
Context BEFORE	She had just picked up the bottle	29	17% (5/29)
Context AFTER	The girl started to feed the baby after the adult made this sign	48	17% (8/48)
Context DURING	He was putting toys into the car	3	33% (1/3)

2.3. Summary

The goal of Study 1 was to determine the extent to which a naïve learner might draw on iconicity when trying to map a signed word form to its meaning in an interactive context. Findings suggested that ASL, like English, was not very transparent in its word-to-world mappings: participants were correct less than 15% of the time. The percentage of correct responses is similar to that in prior work (Medina et al., 2011; see also Gleitman & Trueswell, 2018). Thus, there was no overall benefit to the enhanced iconicity of ASL as compared to English. However, accuracy did increase with iconicity (but not concreteness) at the sign level, such that signs with high iconicity ratings tended to have greater accuracy than signs with low iconicity ratings. In their open responses, participants indicated that they were often using iconicity to guess the sign's meaning, but they were not more often correct when they used this strategy relative to other reported strategies. However, we acknowledge that participants may not have had full explicit knowledge of what informed their guesses. We return to these findings in Section 4.

Our findings also indicate that referential mapping in ASL differs from English in that accuracy was actually better for verb targets relative to noun targets. For example, Medina and colleagues (2011) reported 17% accuracy for noun trials and 6% for verb trials (Gleitman & Trueswell, 2018). This may reflect a bias in participants' responses—namely, their preference for guessing verbs over nouns. The ratio of verb responses to noun responses was more than 3:1, providing an increased chance of providing a correct response on the verb targets, which comprised just half of the target signs. We address this possibility in Study 2.

3. Study 2

The goals of Study 2 were two-fold. The first goal was to confirm findings from Study 1 with a new sample of participants. Specifically, we wanted to replicate the correlation between iconicity and accuracy, and to determine if verbs are more referentially transparent than nouns in ASL. Our findings from Study 1 indicate a possible response bias among participants in which they guessed many more verbs than nouns. This may reflect hearing adults' preconceived notions about ASL rather than something about the acquisition process. Indeed, recent studies in ASL and Sign Language of the Netherlands demonstrate a tendency to identify signs as verbs when nonsigning, hearing adults view videos of isolated signs (Emmorey & Pyers, 2017; Ortega et al., 2019). Here, we aimed to control the response rates for nouns and verbs by telling participants at the beginning of each trial whether the sign was a noun or a verb. This procedure actually makes the study more akin to certain studies using the HSP with spoken English, in which participants were also given word class information (e.g., Gillette et al., 1999; Snedeker et al., 2003). We hypothesized that including word class information would have two effects: first, participants' overall accuracy would increase. By eliminating the option to answer with a word outside of the word class, the hypothesis space of possible meanings narrows, and the chance levels (while still very low) automatically increase. Second, if indeed ASL input makes verbs particularly transparent in their form-meaning mapping, participants should still perform more accurately on verbs than nouns.

The second goal of Study 2 was exploratory: to determine which extralinguistic cues to word meaning were most helpful at the level of an individual labeling event, and whether extralinguistic cues support noun and verb identification differently. In recent HSP work, Trueswell et al. (2016) evaluated the characteristics of the vignettes for which participants often gave correct guesses, deemed highly informative vignettes. For nouns in English, they found that these highly informative vignettes share particular features: the object to which the noun refers is present, and crucially, it is highlighted via referential attention cues, such as gazing at or touching it, just before the onset of the target noun. Thus, we hypothesized that for a noun, accuracy would increase if the referent was present in the scene, and if the object was highlighted immediately prior to its sign by either the child or the parent touching it.[3] We made similar predictions for verbs: if the event described by the verb occurred during the vignette, it should be more likely to be guessed, and more so if the event occurred immediately before the sign. We further expected that object/event presence would interact with iconicity. Although iconicity and accuracy were correlated in Study 1, using an iconicity strategy did not often lead participants to determine the correct meaning. We hypothesized, in line with structural mapping theory (Emmorey, 2014), that iconicity would be more informative if the item or event to which the sign referred was concurrently visible.

3.1. Method

3.1.1. Participants

We recruited a new sample of 37 adult undergraduate and graduate students, age 18–28 years. As in Study 1, all participants were fluent English speakers. Four participants were excluded from the final sample because they reported having some degree of experience with ASL or another sign language, resulting in a final sample of 33 participants (25 female, eight male).

3.1.2. Apparatus and stimuli

The task and stimuli were identical to those in Study 1, with two exceptions. One vignette for CHICKEN and one vignette for MOM were replaced because we noticed that additional productions of those target signs within the vignette were not audibly "beeped." Additionally, the 30 s time window for responding was eliminated because some participants in Study 1 occasionally reported that they were not given enough time to respond.

3.1.3. Procedure

The procedure was virtually identical to Study 1, with the following differences: (1) After they received the instructions and practice trial, we gave them a brief noun/verb quiz in which they indicated via keyboard press whether each of 12 words (presented one at a time in English text) was a noun or verb (e.g., *upon seeing the word "girl," participants were prompted to press "n" for noun, "v" for verb*); (2) participants read on the screen whether the mystery word was a noun or a verb prior to the start of each trial; and (3) they were not asked to provide justifications for their responses; participants' explanations in Study 1 were aligned with our hypotheses, so we did not believe this warranted additional follow-up.

3.1.4. Analysis

Responses were coded for accuracy using the same criteria from Study 1. We planned to exclude any participants that did not pass the noun/verb quiz with at least 80% accuracy, but no participants fell below this threshold. Inclusion criteria for individual responses were also the same as Study 1 with the additional constraint that participants had to have provided a response in the appropriate word class. Trials in which participants responded outside the target word class were excluded. Responses that could be either a noun or a verb (e.g., "light") were coded as belonging to the target word class and were included.

Additionally, each vignette was coded for several extralinguistic cues by a trained coder who was naïve to the study hypotheses, but knew the target signs. Noun items were coded for whether or not the referent object was present (binary), and verbs were coded for whether or not the referent event occurred (binary). When the object was present, noun targets were also coded for object manipulation (binary; included any contact between hand and object, including holding or displacing a sign onto an object), if manipulated then by whom (parent or child), and the time stamp of the first touch of the object relative to the target label. When the referent event occurred, verb items were similarly coded for who carried out the event (parent or child), and the time stamp of that event. Verb items in which an event is never visible (e.g., CAN, WANT) were not coded for presence and excluded from relevant analyses.

A second coder (AF) double coded 30% (57/189) of the vignettes. Interrater reliability was calculated at 87% agreement, Cohen's $\kappa = .7$. The seven disagreements were resolved through discussion between the two coders.

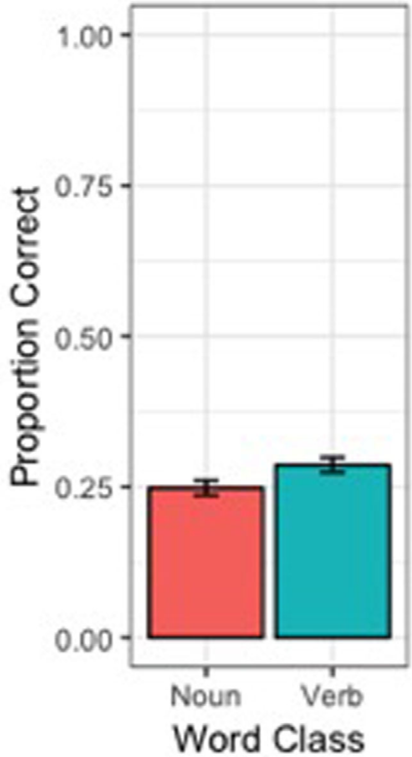

Fig 5. Proportion of accurate responses for nouns and verbs in Study 2. Error bars represent standard error.

3.2. Results

3.2.1. Linguistic cues

As in Study 1, we calculated percentage accuracy across (1) all responses, (2) noun items, and (3) verb items. Findings demonstrated that providing the word class information significantly improved overall accuracy—participants were correct on 26.5% ($SD = 8.77\%$) of their responses, which was significantly better than in Study 1 [$t(64) = -6.60$, $p < .001$, 95% CI $= -.16$ to .09]. Improvements were evident in both nouns ($M = 24.88\%$, $SD = 9.07\%$) and verbs ($M = 28.41\%$, $SD = 12.68\%$). Unlike in Study 1, accuracy for nouns and verbs did not differ [$t(32) = -1.45$, $p = .16$; Fig. 5].

Additionally, we conducted the same Spearman correlations as in Study 1 between accuracy (at the vignette level), iconicity, and concreteness. Again, iconicity was positively correlated with accuracy ($\rho = .55$, $p < .001$), and concreteness was not $\rho = .09$, $p = .21$).

3.2.2. Extralinguistic cues

While linguistic cues are the same across all vignettes with the same target sign (i.e., properties of the sign CHICKEN are consistent with each exemplar of CHICKEN), extralinguistic cues may vary widely across vignettes (e.g., the sign CHICKEN may or may not be preceded by a point to the plastic chicken wing). Thus, the analyses that follow are at the level of the vignette. We evaluated each vignette according to the percent of correct guesses that were produced after watching that vignette. The percentage accuracy for each individual vignette is listed in a table in the Supporting Information (Table S2). Many vignettes (43.7%) yielded zero correct responses; few vignettes yielded many correct responses. In the following analyses, we attempt to disentangle what cues facilitate referential transparency for noun and verb vignettes, and are therefore most likely to lead to successful word learning.

Nouns Perhaps the most likely candidate for separating transparent vignettes from those that are less transparent is the presence or absence of the referent object—it should be more difficult to determine the meaning of a sign when its referent is not visible. Further, referential attention and iconicity are likely to interact with object presence. Specifically, iconicity and object presence can both be useful cues on their own, but iconicity is likely to be more useful when the object is present because parallels between the object and sign can be observed. Object manipulation requires the object to be present, but may or may not boost referential transparency above and beyond what object presence confers alone. This may depend on the timing object manipulation relative to the sign.

Objects were present in 65 noun vignettes from 19 of the 24 noun types. To determine if object presence and/or its interaction with iconicity promote referential transparency, we used a generalized linear mixed-effects model (GLMM) to predict correct responses from the presence of the object, the iconicity of the sign, and the interaction term. Subject and sign type were included as random factors. Findings demonstrated a significant effect of iconicity ($\beta = .72$, $z = 2.74$, $p = .006$) but contrary to our expectations, no significant effect of object presence ($\beta = 2.00$, $z = 1.47$, $p = .14$) and no significant interaction ($\beta = -.21$, $z = -.82$, $p = .41$).

A potential explanation for why object presence did not predict accuracy on noun targets is because the target object was not in the focus of attention at a critical moment for word learning (e.g., immediately before or after the beep). Therefore, we analyzed the agent and timing of object manipulation. For the subset of vignettes in which the object was present, we used GLMM (fit to the full data) to predict correct responses from who touched the object (parent or child), the timestamp (in seconds) of that manipulation relative to the beep (i.e., a touch five seconds before the beep is coded as −5), and the interaction between them. Subject and sign were again included as random factors. Again contrary to our predictions, analysis showed no significant effects of who touched the object ($\beta = -.15$, $z = -.55$, $p = .58$) or time of touch ($\beta = -.13$, $z = -.69$, $p = .49$), and no significant interaction ($\beta = .26$, $z = .91$, $p = .36$).

Verbs To determine what cues may have made verb vignettes transparent, we repeated the same GLMMs, replacing the object's presence with the occurrence of the event associated with the verb. Events were present for 40 verb vignettes from 14 of the 24 verb types. The first model analyzed the extent to which correct responses were predicted by event occurrence, sign iconicity, and the interaction term. Findings showed that iconicity ($\beta = .89$, $z = 3.39$, $p < .001$) and event occurrence ($\beta = 3.37$, $z = 2.87$, $p = .004$), but not the interaction ($\beta = -.33$,

Table 3
Contingency tables for informativity and object presence (for nouns) or event presence (for verbs) in Study 2 for (A) nouns and (B) verbs

A			
	Object present	Object absent	Total vignettes
High informative	19	4	23
Low informative	44	25	69
Total vignettes	63	29	92
B			
	Event present	Event absent	Total vignettes
High informative	23	6	29
Low informative	14	24	38
Total vignettes	37	30	67

$z = -1.45$, $p = .15$), predicted correct responses to verb items. For the subset of vignettes in which the event occurred, we ran a second model to analyze whether correct responses were predicted by the agent who performed the action (parent or child), the timestamp of the event onset, and the interaction between them. Here, the agent who performed the action significantly predicted accuracy: events in which the child, but not the parent enacted the referent significantly predicted correct responses ($\beta = 2.41$, $z = 4.03$, $p < .001$). Neither the timing of the event ($\beta = .24$, $z = .58$, $p = .56$), nor the interaction between agent and timing ($\beta = -.41$, $z = -1.00$, $p = .32$) were significant predictors.

Informativity analysis Given the puzzling finding that for noun vignettes, referent object presence did not predict a sign's referential transparency, we pursued additional analyses. Borrowing from Medina and colleagues (2011), we considered vignettes where accuracy was $<30\%$ as "low informative" (LI) vignettes, because the context was not very informative for mapping that word to its meaning. Vignettes that yielded a large percentage ($>50\%$) of correct responses were labeled "high informative" (HI) vignettes. Vignettes that fell in the middle were excluded from this analysis. This classification yielded 52 HI vignettes (23 noun tokens from 10 of the 24 noun types, 29 verb tokens from 13 of the 24 verb types) and 107 LI vignettes (69 noun tokens from 21 noun types, 61 verb tokens from 22 verb types). For HSPs, where response accuracy tends to be quite low, this system has greater validity than a traditional median split.[4]

Once vignettes were categorized, we looked for relationships between informativity (high/low) and object/event presence. To do this, we conducted chi-square tests for independence (one for each of noun and verb vignettes; Table 3A and B). Our goal for this analysis was to determine whether object/event presence was necessary and/or sufficient for identifying the target sign. If presence were necessary and sufficient for mapping word meaning, we would expect all HI vignettes to have the object present, and most LI vignettes to have the object absent. On the other hand, if object presence is neither necessary nor sufficient for guessing the sign, we would expect there to be no significant differences in object presence in LI vs HI vignettes. Note that the vignettes from the six never visible verb types (CAN, CAN'T, DO-DO, DON'T-WANT, HAVE and WANT) were not included in this analysis.

For nouns, the chi-square test did not reach significance (χ^2 (1) = 2.84, p = .09), suggesting that object presence was neither necessary nor sufficient for guessing the object. Numerically, almost all HI vignettes contained the target object, but the majority of LI vignettes also contained the target object (in line with a necessary, but not sufficient hypothesis). For verbs, we saw a different pattern: in most HI events the event occurred, and in most LI events the event did not occur. The chi-square test was significant (χ^2 (1) = 12.00, $p < .001$). This suggests that informativity and event presence are not independent factors, but rather factors that coincide. We therefore argue that our results align most closely with the hypothesis that object/event presence is necessary, but not sufficient, for determining form-meaning mapping in ASL.

3.3. Summary

The goals of Study 2 were to determine whether the findings from Study 1 would be replicated when participants were told ahead of time whether the target sign was a noun or verb. We also sought to uncover which cues facilitate referential transparency, and if they vary by word class. With regard to the first goal, findings showed that providing word class information improved performance on the HSP task by $\sim$12% overall, and that there were now no statistical differences in accuracy between noun and verb targets. This implies that the verb advantage seen in Study 1 was indeed the result of response bias, such that participants were more likely to guess verbs than nouns overall. However, this finding remains different from those seen in spoken English HSP studies, where nouns take the lion's share of accuracy. Additionally, we replicated the finding from Study 1 that at the level of the sign, iconicity correlates with accuracy.

Findings related to the second goal were more nuanced. For nouns, we did not identify any significant predictor of accuracy on the vignette-level—iconicity (a sign-level cue) was the only factor that significantly predicted accuracy. That said, nearly all HI noun vignettes contained the referent object, and most of the LI vignettes did as well. This suggests that object presence is necessary for meaning mapping, but may not be sufficient on its own. For verbs, both iconicity and event occurrence predicted successful meaning mapping, although these factors did not interact. More specifically, when the referent event occurred, vignettes were more referentially transparent when the child carried out the action, although the timing of when they did so did not predict the outcome.

4. Discussion

The studies presented here are a first step in uncovering cues that may support word learning in ASL. To our knowledge, this is the first HSP study to be conducted in ASL or any sign language. In Study 1, we discovered that nonsigning, hearing adults were rarely able to determine the meaning of an unknown sign from a 40-s interaction in child-directed sign. The signs they did guess correctly tended to be highly iconic, and were much more likely to be verbs than nouns. In Study 2, when participants were informed about the sign's word class, accuracy increased. This parallels findings from participants in other HSP studies (Snedeker, Gleitman & Brent, 1999), as well as findings from behavioral studies with children, who are more successful at word learning when provided information about a word's class (e.g., Brown, 1957; Waxman, Lidz, Lavin, & Braun, 2009). Participants in Study 2 were also equally successful at guessing nouns and verbs. Iconicity remained the strongest (and for nouns, the only) predictor of successful mapping. For verbs, referential context—specifically the child enacting the referent event—also played a role. We discuss these findings in the context of each of our research questions in turn.

4.1. Does iconicity enhance referential transparency in ASL?

A feature that sets sign languages apart from spoken languages is its higher degree of iconicity, and the fact that in ASL, iconicity is a visual form representing a visual meaning. Our findings underscore this difference. Across both studies, iconicity was related to successful meaning mapping, and in Study 2, iconicity was the strongest predictor of correct responses. Concreteness, in contrast, was not related to accuracy in either study. While the HSP does not permit direct inferences to be drawn from adult participants to children acquiring ASL as a first language, these findings align with prior work demonstrating that in children, iconicity is related to acquisition in ASL and other sign languages (Caselli & Pyers, 2017, 2020; Ortega et al., 2017). We therefore propose that these highly iconic forms are also highly referentially transparent. On the face of it, this makes sense: a sign that looks like its referent will be more visible in its meaning. Indeed, this is the strategy that participants most reported using in Study 1.

Our findings robustly support the notion that iconicity advantages word learning at the local level (that is, iconicity is beneficial for iconic words), but speak little to the use of iconicity for the acquisition of noniconic words, which has previously been proposed (Imai & Kita, 2014). We saw that participants often overused iconicity as a strategy in Study 1, which sometimes misled them for both arbitrary and iconic targets, due to mismatches in sign and referent features and overlap between potential referents. For example, one participant reported that the sign for DRAW appeared as though the parent was making an X on their hand (they therefore guessed the sign meant "stop"). Another participant guessed "pacifier" for FOOD, reporting that the location and motion near the mouth were clues. Sehyr and Emmorey (2019) term these signs that yield incorrect but consistent guesses highly "transparent," having found the same effect in a study of nonhearing signers guessing the meanings of isolated signs.

Prior literature has suggested that a second problem with relying on iconic strategies is that even when a sign is iconic, it may not activate the mental representation of its referent concept (i.e., a fried chicken wing will probably not call the chicken's beak to mind). Structural mapping theory argues that this is because iconicity benefits processing (and perhaps, by extension, learnability) only when there is clear alignment between the form and the mental representation of the referent (Emmorey, 2014; Taub, 2001). This may especially disadvantage young language-learning children, who may lack the referent concepts, and thus will fail to activate them when presented with an iconic sign. We hypothesized that a solution to mapping unknown or not activated concepts is referent presence. Being able to see a parent drinking a glass of juice should make the mapping of the iconic sign DRINK particularly straightforward, because it resembles a glass being held to the mouth with a drinking motion. It has been argued that parents actually highlight these parallels when producing iconic signs (Perniss et al., 2018). However, our data do not support this—the presence of an object or event did not interact with iconicity to predict success. It may be the case that for highly iconic signs, the context alone was sufficient to activate the depicted concept. For example, a picnic setting in which people are eating may activate the concept of drinking, even absent cups or juice. Note that this contextual explanation does not resolve the issue of unknown concepts—children would have to know what occurs at a picnic to make the same connection. Clearly, future work on the mechanism that supports iconic word-meaning mapping is warranted.

4.2. What is the role of extralinguistic referential cues (object/event presence, manipulation, cue timing) in promoting referential transparency?

Another factor that we expected would be different from spoken language is the use of extralinguistic cues to meaning. Specifically, while referential attention cues (gaze, object manipulation) are likely to be important for word learning cross-linguistically, the relative timing of these cues differs for a sign language articulated by the hands. We anticipated that referent presence (object or event presence) would enhance referential transparency because it is a bare minimum requirement for other referential attention cues. We further hypothesized that within vignettes where the object is present, object manipulation would be helpful in determining the meaning of a sign (because it highlights the referent). Further, we hypothesized that the timing of the referential cue relative to the object label would be sequential, and not simultaneous as is seen in spoken language (Trueswell et al., 2016), because of the need in sign language to be looking at the interlocutor at the moment of labeling.

For nouns, presence of the referent object was insufficient for mapping. This is perhaps because the object was present in most vignettes—both the highly informative and low informative vignettes. This mirrors findings from Trueswell et al.'s (2016) English HSP study—object presence was not a significant predictor of performance. However, those researchers found that parent or child attention to the object highlighted its presence at the moment the word was uttered, and this led to accurate mapping. This was not the case for our HSP study. Neither highlighting the object by the parent or child touching it, nor the timing of that first occurrence relative to the sign onset, predicted accuracy on nouns in Study 2. Our measure of attention differed from Trueswell et al.'s, which may explain some of this discrepancy. Specifically, they included eye gaze, deictic gestures, and body orientation in addition to touch/manipulation—by using only touch, our measure is narrower and may not capture strategies parents used to highlight the referent object. However, we anticipate that extralinguistic cues that do not vary by modality (e.g., body orientation) are likely to pattern in the same way. We also examined when the object was first highlighted in the interaction (because that is likely when it grabs attention) rather than examining the cues frame-by-frame. However, our finding potentially highlights an important difference between spoken and sign languages: While in spoken languages, complete overlap between touch and gaze to the object and label can be effective because visual attention remains on the object (e.g., Baldwin, 1991), ASL requires the child's visual attention to be redirected to the interlocutor's sign input. Therefore, it may be equally effective to label an object before or after it has been attended to.

In contrast, verb items told a different story. The occurrence of the referent event at any point during the vignette was a significant predictor of accuracy, suggesting that this was a sufficient cue for meaning mapping, along with iconicity. Like nouns though, the time at which the event occurred relative to the sign onset did not predict accuracy. Thus, exactly how events are highlighted by ASL input is not fully revealed by our data. Participants were more likely to respond correctly when the child had performed the action labeled by the verb, suggesting that participants may expect parents to label the child's (and thus, their focus of attention) after it has begun. However, this account would predict that the timing of the action (occurring *before* the label) would be a significant predictor of success, especially in ASL-based interactions where timing is critical to enable children to shift attention (see also Tomasello & Kruger, 1992), contra our findings. Future work should pursue the precise relationship between timing and labeling in more nuanced measures.

4.3. Do ASL nouns and verbs differ in referential transparency?

A striking finding from Study 1 was the discrepancy between nouns and verbs. Despite the concrete nature of the nouns in the study (many of which referred to objects in the scene), participants succeeded in guessing verbs almost twice as often as nouns. This advantage seemed to be due to response bias. When we provided participants with word class information—that is, whether the target sign was a noun or verb—in Study 2, there was no verb advantage—nouns and verbs were guessed equally well.

The response bias itself provides meaningful evidence about how naïve learners viewed these ASL interactions. It implies that without additional information, nonsigners seem to assign event meanings rather than object meanings to signs. This aligns with recent evidence from other studies on hearing nonsigners showing a very similar bias (Emmorey & Pyers, 2017; Ortega et al., 2019). What causes this bias is unclear. However, several production studies have demonstrated that in both sign language and gesture, deaf and hearing individuals depict objects by describing/indicating how the objects are handled or manipulated (Ortega, Sumer & Ozyurek, 2017; Pettenati, Sekine, Congestri & Volterra, 2012; Van Nispen et al., 2017). The bias to guess event meanings for signs may therefore reflect a belief that actions are more likely to be produced in the manual modality.

When participants were given word class information, accuracy was equivalent for nouns and verbs. This is akin to other findings from spoken language—just not English. Snedeker and others previously hypothesized that the relative proportion of nouns and verbs in a child's early vocabulary may be driven by extralinguistic factors in their input. Specifically, in languages where children acquire more predicates early on, caregivers may highlight verbs in a way that languages like English do not. This hypothesis drew on evidence that in the HSP, nouns outperform verbs when the input is in English, but they equal verbs when the input is in Mandarin (Snedeker & Li, 2000; Snedeker et al., 2003). Here, we provide converging evidence; our participants' performance mirrored that of the Mandarin HSP. This could be related to the fact that like Mandarin, early vocabularies in ASL contain proportionally more predicates than early English vocabularies (although they still contain many nouns; Anderson & Reilly, 2002); thus, there is something in the input to ASL learners that may make verb learning relatively easier. One possibility that deserves future attention is the recent finding that verbs are particularly frequent in child-directed input in ASL (Fieldsteel et al., 2020). Additionally, verb performance was specifically predicted by (in addition to iconicity), whether or not the verb's referent event had occurred in the vignette. In contrast, the presence of a noun's referent object did not predict success. This suggests that the verb referents in our study are indeed better highlighted by the extralinguistic input than the noun referents. However, Snedeker et al.'s exploration of the cues that adults might have used to succeed with noun and verb targets does not fully align with our own. For example, they found that object presence predicted noun performance, while we did not. Moreover, they found minimal differences between Mandarin and English in the cues that predicted success, meaning that we still have not identified a strong explanation for cross-linguistic differences in performance.

4.4. Limitations

The use of the HSP is an early step toward understanding word learning through child-directed input in ASL. Our use of this paradigm provided us with many benefits of the HSP, but like other HSP studies also may not generalize to language acquisition in vivo. Specifically, adults are not perfect proxies for children acquiring language. Adults have a vast amount of conceptual knowledge, preexisting biases, and have already acquired a language. On the converse, children may be able to draw on certain learning strategies that support language development that adults do not have access to. Further, the task differs from real-world learning in that the adult participant is in the position of only learning one word at a time, but also only having one opportunity to learn that word. Their third-party view of the dyadic interaction further does not capture the first-person view of a language-acquiring child (although see Akhtar, Jipson & Callanan, 2001; Akhtar, Tolins & Fox Tree, 2019). Further, observation of the extralinguistic context alone may parallel acquisition of the very first words in a child's vocabulary, but by the second year, children are able to use the words they know to reduce referential ambiguity in the word learning process (see, e.g., Fisher, Jin, & Scott, 2020; He & Arunachalam, 2017). Despite these limitations, HSPs have provided evidence for which research avenues deserve further pursuit. This is critical in a field where data collection with a large number of young children is time consuming and resource intensive. This is especially true when studying deaf children, where this research is sorely needed.

In addition to sharing limitations with traditional HSPs, our HSP is necessarily methodologically different from prior work, and introduces new challenges. The fact that participants could see the language being articulated is more akin to a foreign language HSP with the sound on (which to our knowledge, has not been done). Hearing, nonsigning participants may have biases (correct or not) about the structure of ASL that children are not likely to have (including, perhaps, that signs are more likely to be verbs than nouns). These biases may have had less influence if signs were not visible. However, just as in spoken language HSPs, some of our findings mirror those of children's vocabulary acquisition patterns in ASL (e.g., Anderson & Reilly, 2002; Caselli & Pyers, 2017; Thompson et al., 2012). We therefore believe that this methodology has merit in the visual modality and provides a starting point for future work in children acquiring ASL.

4.5. Conclusions

This study sheds light on the word learning process in ASL, and more generally across languages. Specifically, we demonstrated that as in spoken languages, observation alone is insufficient for form-meaning mapping in ASL. This underscores the idea that children are using multiple strategies when acquiring language. We also provided evidence that there are language-specific features that are relevant to word learning. For example, iconicity is likely to be relevant for visual languages that are rich in iconic cues, and far less relevant for spoken languages, particularly those with a low degree of sound symbolism. Additionally, different kinds of words may be highlighted differently cross-linguistically (e.g., verbs in ASL or Mandarin). Finally, we added to a growing body of evidence that referential attention cues may be important in early verb learning (Childers, Vaughan & Burquest, 2007; Roseberry, Hirsh-Pasek & Golinkoff, 2014; Tomasello & Kruger, 1992). Each of these areas is ripe for further exploration—in ASL, in sign languages more broadly, and across modalities.

Notes

1 Two vignettes, one for MOM and one for CHICKEN, did not contain an audible beep for at least one of the additional productions. Excluding these vignettes from analysis in Study 1 did not meaningfully change the results, so they are included here.
2 A recent expansion of ASL-LEX includes iconicity ratings from deaf signers, which are highly correlated with ratings from hearing nonsigners (Sehyr et al., 2021).
3 Although gaze cues can also highlight an object or event in the scene, we were unable to accurately code gaze in this study. Additionally, pointing was not coded because points occupy multiple functions in ASL and are not purely extralinguistic (Meier & Lillo-Martin, 2013).
4 In our data, median accuracy across vignettes was only 14%.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

Open Research Badges

This article has earned Open Data badge. Data are available at https://osf.io/5te46/.

References

Akhtar, N., Dunham, F., & Dunham, P. J. (1991). Directive interactions and early vocabulary development: The role of joint attentional focus. *Journal of Child Language*, *18*(1), 41–49. https://doi.org/10.1017/S0305000900013283

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development*, *72*(2), 416–430.

Akhtar, N., Tolins, J., & Tree, J. E. F. (2019). Young children's word learning through overhearing: Next steps. In J. von Koss Torkildsen, & J. Horst (Eds.), *International handbook of language acquisition* (pp. 427–441). Abingdon, UK: Routledge.

Akhtar, N., & Tomasello, M. (2000). The social nature of words and word learning. In K. Hirsh-Pasek, & R. M. Golinkoff (Eds.), *Becoming a word learner: A debate on lexical acquisition*. Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195130324.003.005

Anderson, D., & Reilly, J. (2002). The MacArthur communicative development inventory: normative data for American Sign Language. *Journal of Deaf Studies and Deaf Education*, *7*(2), 83–106.

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, *20*(2), 395–418.

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, *62*(5), 875–890. https://doi.org/10.1111/j.1467-8624.1991.tb01577.x

Baus, C., Carreiras, M., & Emmorey, K. (2013). When does iconicity in sign language matter? *Language and Cognitive Processes*, *28*(3), 261–271. https://doi.org/10.1080/01690965.2011.620374

Braginsky, M., Yurovsky, D., Marchman, V., & Frank, M. C. (2017). Consistency and variability in word learning across languages. *Open Mind*, *3*, 52–67

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*(6), 535–543. https://doi.org/10.1111/J.1467-7687.2005.00445.X

Brooks, R., Singleton, J. L., & Meltzoff, A. N. (2020). Enhanced gaze-following behavior in deaf infants of deaf parents. *Developmental Science*, *23*(2), 1–10. https://doi.org/10.1111/desc.12900

Brown, R. W. (1957). Linguistic determinism and the part of speech. *The Journal of Abnormal and Social Psychology*, *55*(1), 1–5.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911.

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, *110*(28), 11278–11283. https://doi.org/10.1073/pnas.1309518110

Caselli, N. K., & Pyers, J. E. (2017). The road to language learning is not entirely iconic: Iconicity, neighborhood density, and frequency facilitate acquisition of sign language. *Psychological Science*, *28*(7), 979–987. https://doi.org/10.1177/0956797617700498

Caselli, N. K., & Pyers, J. E. (2020). Degree and not type of iconicity affects sign language vocabulary acquisition. *Journal of Experimental Psychology: Learning Memory and Cognition*, *46*(1), 127. https://doi.org/10.1037/xlm0000713

Caselli, N. K., Sehyr, Z. S., Cohen-Goldberg, A. M., & Emmorey, K. (2017). ASL-LEX: A lexical database of American Sign Language. *Behavior Research Methods*, *49*(2) 784–801. https://doi.org/10.3758/s13428-016-0742-0

Childers, J. B., Vaughan, J., & Burquest, D. A. (2007). Joint attention and word learning in Ngas-speaking toddlers in Nigeria. *Journal of Child Language*, *34*(2), 199–225. https://doi.org/10.1017/S0305000906007835

Davies, M. (2018). The 14 billion word iWeb corpus. [Database].

Emmorey, K. (2014). Iconicity as structure mapping. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *396*(1651), 20130301. https://doi.org/10.1098/rstb.2013.0301

Emmorey, K., & Pyers, J. E. (2017). Cognitive biases in construing iconic mappings. *11th International Symposium on Iconicity in Language and Literature, Brighton, UK.*

Fieldsteel, Z., Bottoms, A., & Lieberman, A. M. (2020). Nouns and verbs in parent input in American Sign Language during interaction among deaf dyads. *Language Learning and Development*, *16*(4), 351–363.

Fisher, C., Jin, K., & Scott, R. M. (2020). The developmental origins of syntactic bootstrapping. *Topics in Cognitive Science*, *12*(1), 48–77. https://doi.org/10.1111/tops.12447

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2016). Wordbank: A repository for vocabulary data. *Journal of Child Language*, *44*(3), 677–694.

Frank, M. C., Braginsky, M., Marchman, V. A., & Yurovsky, D. (2020). *Variability and consistency in early language learning: The Wordbank project*. Cambridge, MA: MIT Press.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135–176. https://doi.org/10.1016/S0010-0277(99)00036-0

Gleitman, L. R., & Trueswell, J. C. (2018). Easy words: Reference resolution in a malevolent referent world. *Topics in Cognitive Science*, *12*(1), 22–47. https://doi.org/10.1111/tops.12352

Griffith, P. L., Robinson, J. H., & Panagos, J. M. (1981). Perception of iconicity in American Sign Language by hearing and deaf subjects. *Journal of Speech and Hearing Disorders*, *46*(4), 388–397. https://doi.org/10.1044/jshd.4604.388

He, A. X., & Lidz, J. (2017). Verb learning in 14- and 18-month-old English-learning infants. *Language Learning and Development*, *13*(3), 335–356. https://doi.org/10.1080/15475441.2017.1285238

He, A. X., & Arunachalam, S. (2017). Word learning mechanisms. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(4), e1435. https://doi.org/10.1002/wcs.1435

Hochgesang, J. A., Crasborn, O. & Lillo-Martin, D. (2020). ASL Signbank. New Haven, CT: Haskins Lab, Yale University. https://aslsignbank.haskins.yale.edu/

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130298. https://doi.org/10.1098/rstb.2013.0298

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, *109*(1), 54–65. https://doi.org/10.1016/j.cognition.2008.07.015

Laing, C. E. (2019). Phonological motivation for the acquisition of onomatopoeia: An analysis of early words. *Language Learning and Development*, *15*(2), 177–197. https://doi.org/10.1080/15475441.2019.1577138

Lieberman, A. M., Hatrak, M., & Mayberry, R. I. (2014). Learning to look for language: Development of joint attention in young deaf children. *Language Learning and Development*, *10*(1). https://doi.org/10.1080/15475441.2012.760381

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(22), 9014–9019. https://doi.org/10.1073/pnas.1105040108

Meier, R. P., & Lillo-Martin, D. (2013). The points of language. *Humanamente*, *24*, 151.

Mitchell, R. E., & Karchmer, M. A. (2004). Chasing the mythical ten percent: Parental hearing status of deaf and hard of hearing students in the United States. *Sign Language Studies*, *4*(2), 138–163. https://doi.org/10.1353/sls.2004.0005

Morett, L. M. (2015). Lending a hand to signed language acquisition: Enactment and iconicity enhance sign recall in hearing adult American Sign Language learners. *Journal of Cognitive Psychology*, *27*(3), 251–276. https://doi.org/10.1080/20445911.2014.999684

Motamedi, Y., Murgiano, M., Perniss, P., Wonnacott, E., Marshall, C., Goldin-Meadow, S., & Vigliocco, G. (2020). Linking language to sensory experience: onomatopoeia in early language development. *Developmental Science*, *24*(3), e13066. https://doi.org/10.1111/desc.13066

Nielsen, A. K. S., & Dingemanse, M. (2020). Iconicity in word learning and beyond: A critical review. *Language and Speech*, *64*(1), 52–72. https://doi.org/10.1177/0023830920914339

Ortega, G. (2017). Iconicity and sign lexical acquisition: A review. *Frontiers in Psychology*, *8*, 1280. https://doi.org/10.3389/fpsyg.2017.01280

Ortega, G., & Morgan, G. (2015). The effect of iconicity in the mental lexicon of hearing non-signers and proficient signers: evidence of cross-modal priming. *Language, Cognition and Neuroscience*, *30*(5), 574–585. https://doi.org/10.1080/23273798.2014.959533

Ortega, G., Schiefner, A., & Ozyurek, A. (2019). Hearing non-signers use their gestures to predict iconic form-meaning mappings at first exposure to signs. *Cognition*, *191*, 103996. https://doi.org/10.1016/j.cognition.2019.06.008

Ortega, G., Sümer, B., & Özyürek, A. (2017). Type of iconicity matters in the vocabulary development of signing children. *Developmental Psychology*, *53*(1), 89. https://doi.org/10.1037/dev0000161

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Perniss, P., Lu, J. C., Morgan, G., & Vigliocco, G. (2018). Mapping language to the world: The role of iconicity in the sign language input. *Developmental Science*, *21*(2), e12551. https://doi.org/10.1111/desc.12551

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, *1*, 227. https://doi.org/10.3389/fpsyg.2010.00227

Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS ONE*, *10*(9), e0137147. https://doi.org/10.1371/journal.pone.0137147

Pettenati, P., Sekine, K., Congestrì, E., & Volterra, V. (2012). A comparative study on representational gestures in Italian and Japanese children. *Journal of Nonverbal Behavior*, *36*(2), 149–164. https://doi.org/10.1007/s10919-011-0127-0

Piccin, T. B., & Waxman, S. R. (2007). Why nouns trump verbs in word learning: New evidence from children and adults in the human simulation paradigm. *Language Learning and Development*, *3*(4), 295–323. https://doi.org/10.1080/15475440701377535

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype me! Socially contingent interactions help toddlers learn language. *Child Development*, *85*(3), 956–970. https://doi.org/10.1111/cdev.12166

Sehyr, Z. S., & Emmorey, K. (2019). The perceived mapping between form and meaning in American Sign Language depends on linguistic knowledge and task: Evidence from iconicity and transparency judgments. *Language and Cognition*, *11*(2), 208–234. https://doi.org/10.1017/langcog.2019.18

Sehyr, Z. S., Caselli, N., Cohen-Goldberg, A. M., & Emmorey, K. (2021). The ASL-LEX 2.0 Project: A Database of Lexical and Phonological Properties for 2,723 Signs in American Sign Language. *The Journal of Deaf Studies and Deaf Education*, *26*(2), 263–277.

Snedeker, J., Gleitman, L. R., & Brent, M. R. (1999). The successes and failures of word-to-world mapping. In A. Greenhill, M. Hughs, H. Littlefield, & H. Walsh (Eds.), *Proceedings of the Twenty-Third Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Snedeker, J., & Li, P. (2000). The Limits of Observation: Can the Situations in Which Words Occur Account For Cross-Linguistic Variation in Vocabulary Composition?. J. Tai & Y. Chang, *Proceedings of the Seventh International Symposium on Chinese Language and Linguistics*.

Snedeker, J., Li, P., & Yuan, S. (2003). Cross-cultural differences in the input to early word learning. *25th Annual Conference of the Cognitive Science Society, August 2, Boston, MA*, p. 25.

Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian, and Mandarin. *Journal of Child Language*, *24*(3), 535–565.

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology*, *32*(3), 492. https://doi.org/10.1037/0012-1649.32.3.492

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby's first 10 words. *Developmental Psychology*, *44*(4), 929. https://doi.org/10.1037/0012-1649.44.4.929

Taub, S. F. (2001). *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge, England: Cambridge University Press.

Thompson, R. L., Vinson, D. P., Woll, B., & Vigliocco, G. (2012). The road to language learning is iconic: Evidence from British Sign Language. *Psychological science*, *23*(12), 1443–1448.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*(6), 1454–1463. https://doi.org/10.1111/J.1467-8624.1986.Tb00470.X

Tomasello, M., & Kruger, A. C. (1992). Joint attention on actions acquiring verbs in ostensive and non-ostensive contexts. *Journal of Child Language*, *19*(2), 311–333. https://doi.org/10.1017/S0305000900011430

Trueswell, J. C., Lin, Y., Armstrong, B., Cartmill, E. A., Goldin-Meadow, S., & Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent-child interactions. *Cognition*, *148*, 117–135. https://doi.org/10.1016/j.cognition.2015.11.002

van Nispen, K., van de Sandt-Koenderman, M., Sekine, K., Krahmer, E., & Rose, M. L. (2017). Part of the message comes in gesture: How people with aphasia convey information in different gesture types as compared with information in their speech. *Aphasiology*, *31*(9), 1078–1103.

Waxman, S. R., Lidz, J. L., Braun, I. E., & Lavin, T. (2009). Twenty four-month-old infants' interpretations of novel verbs and nouns in dynamic scenes. *Cognitive Psychology*, *59*(1), 67–95. https://doi.org/10.1016/j.cogpsych.2009.02.001

Yu, C., & Smith, L. B. (2015). Linking joint attention with hand-eye coordination - A sensorimotor approach to understanding child-parent social interaction. *CogSci … Annual Conference of the Cognitive Science Society. Cognitive Science Society (U.S.). Conference*, *2015*, 2763–2768.

Yu, C., & Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Current Biology*, *26*(9), 1235–1240. https://doi.org/10.1016/j.cub.2016.03.026

Yu, C., & Smith, L. B. (2017). Hand–eye coordination predicts joint attention. *Child Development*, *88*(6), 2060–2078. https://doi.org/10.1111/cdev.12730

Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science*, *22*(1), e12735. https://doi.org/10.1111/desc.12735

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting Information