



Article

The Seeds of the Noun–Verb Distinction in the Manual Modality: Improvisation and Interaction in the Emergence of Grammatical Categories

Yasamin Motamedi ^{1,*} , Kathryn Montemurro ², Natasha Abner ³, Molly Flaherty ⁴, Simon Kirby ¹ and Susan Goldin-Meadow ⁵ 

- ¹ Centre for Language Evolution, University of Edinburgh, Edinburgh EH8 9AD, UK; simon.kirby@ed.ac.uk
² Department of Linguistics, University of Chicago, Chicago, IL 60637, USA; kathryn.montemurro@gmail.com
³ Department of Linguistics, University of Michigan, Ann Arbor, MI 48109, USA; nabner@umich.edu
⁴ Department of Psychology, Davidson College, Davidson, NC 28035, USA; moflaherty@davidson.edu
⁵ Department of Psychology, University of Chicago, Chicago, IL 60637, USA; sgsg@uchicago.edu
* Correspondence: yasamin.motamedi@gmail.com

Abstract: The noun–verb distinction has long been considered a fundamental property of human language, and has been found in some form even in the earliest stages of language emergence, including homesign and the early generations of emerging sign languages. We present two experimental studies that use silent gesture to investigate how noun–verb distinctions develop in the manual modality through two key processes: (i) improvising using novel signals by individuals, and (ii) using those signals in the interaction between communicators. We operationalise communicative interaction in two ways: a setting in which members of the dyad were in separate booths and were given a comprehension test after each stimulus vs. a more naturalistic face-to-face conversation without comprehension checks. There were few differences between the two conditions, highlighting the robustness of the paradigm. Our findings from both experiments reflect patterns found in naturally emerging sign languages. Some formal distinctions arise in the earliest stages of improvisation and do not require interaction to develop. However, the full range of formal distinctions between nouns and verbs found in naturally emerging language did not appear with either improvisation or interaction, suggesting that transmitting the language to a new generation of learners might be necessary for these properties to emerge.

Keywords: silent gesture; interaction; improvisation; grammatical categories; sign language emergence; iconicity



Citation: Motamedi, Yasamin, Kathryn Montemurro, Natasha Abner, Molly Flaherty, Simon Kirby, and Susan Goldin-Meadow. 2022. The Seeds of the Noun–Verb Distinction in the Manual Modality: Improvisation and Interaction in the Emergence of Grammatical Categories. *Languages* 7: 95. <https://doi.org/10.3390/languages7020095>

Academic Editors: Wendy Sandler, Mark Aronoff and Carol Padden

Received: 1 October 2021

Accepted: 22 March 2022

Published: 11 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The majority of the world's languages distinguish between the grammatical categories of noun and verb¹. Indeed, the grammatical distinction between nouns and verbs has long been considered a fundamental feature of human language (Hockett 1977; Hopper and Thompson 1985; Jackendoff 2002), thought to emerge early on in the evolution of language (Bickerton 1990; Heine and Kuteva 2007).

More recently, studies of emergent linguistic systems offer support for the fundamental nature of the noun–verb distinction. Research on homesign systems, communication systems created by deaf children without a language model, suggest that such children distinguish between nouns and verbs (Goldin-Meadow et al. 1994). Furthermore, research on emerging sign languages suggests that differences in how noun and verb forms are produced can emerge even in the earliest stages of language creation (Abner et al. 2019; see also Goldin-Meadow et al. 2014). However, this body of work has also noted that systematic noun–verb distinctions do not emerge fully formed, but become increasingly systematised and conventionalised through the development of a linguistic community.

It is difficult to study the factors that lead to systematization and conventionalization in a naturally emerging language, simply because we have no control over the conditions under which the language is developing. We turn here to an experimental paradigm that has the potential to allow us to model the emergence of noun–verb distinctions in the manual modality. To the extent that our experimental paradigms reveal a picture of language emergence that resembles the picture we get from naturally evolving languages, we will have evidence that these artificial language experiments are good models for the study of naturally emerging languages. We can then use the paradigms to experimentally explore factors that influence emergence. We focus on two processes that have the potential to shape the evolution of noun–verb categories in emerging communication systems: (i) improvisation and (ii) interaction. We examine how individual participants improvise and innovate novel gestural signals for events in which the target object is the focus (noun context) and events in which the target action is the focus (verb context), and then investigate how those gestures change when used in interaction.

1.1. Noun–Verb Distinctions in Natural Sign Languages

The grammatical categories of nouns and verbs are almost universally present across languages and modalities, and are thought to be based on pre-linguistic conceptual categories (Hurford 2007)—namely, the need to communicate about objects in the world (i.e., nouns) and the properties/actions of those objects or the relations between them (i.e., verbs). Nouns and verbs differ in how they are used in a simple proposition, with nouns typically heading subjects and objects (or other participant roles), and verbs typically heading predicates in a proposition. In addition to this functional distinction, nouns and verbs can also differ in form, either in the base form itself (e.g., *sing/song*, *process/prócess*), or in other formal properties that can map onto these functional contrasts. In spoken languages, for example, constituent order can distinguish verbs from nouns, and nouns and verbs can carry different inflections—nouns are often marked for gender, number and person, and verbs for tense, aspect and mood.

In sign languages, constituent order, syntactic distribution, and morphosyntactic marking also distinguish between nouns and verbs. In addition, sign languages frequently signal grammatical categories by altering the sign form in nouns and verbs that have similar underlying forms (see Abner 2021 for an overview). For example, verb and noun signs can be distinguished by size or length of movement: verbs tend to be articulated with a larger movement (Kimmelman 2009; Pizzuto and Corazza 1996), or longer duration (Hunger 2006; Pizzuto and Corazza 1996) than nouns. They can also be distinguished by the manner of movement: verbs may be articulated with continuous movement while nouns are articulated with more restrained movement in both American Sign Language (ASL) and Australian Sign Language (Auslan) (Johnston 2001; Supalla and Newport 1978). In addition, nouns in both ASL and Auslan, as well as Russian Sign Language, tend to be articulated with repeated movements, whereas verbs exhibit variability based on their meaning (Johnston 2001; Kimmelman 2009; Supalla and Newport 1978). Finally, sign languages such as British Sign Language (BSL) can borrow mouthings from the ambient spoken language, and use these mouthings to distinguish nominal and verbal forms, with noun forms more likely to be accompanied by mouthing than verb forms in some languages (Hunger 2006; Johnston 2001; Kimmelman 2009; Tkachman and Sandler 2013). There are, however, cross-linguistic differences in how some of these strategies are implemented; for example, evidence from Turkish Sign Language (TİD), Sign Language of the Netherlands (NGT) and homesign shows that, in at least these cases, repetition is *less* likely to occur in nouns than verbs (Goldin-Meadow et al. 1994; Kubus 2008; Schreurs 2006).

Nevertheless, most documented sign languages demonstrate a set of noun–verb pairs distinguished by altering properties of a shared underlying form and display striking cross-linguistic commonalities in how these distinctions are made (Tkachman and Sandler 2013). Such distinctions have been shown primarily to operate over subsets of noun–verb pairs associated with concrete objects and instrumental actions (e.g., WINDOW/CLOSE-

WINDOW) where the base form is iconically motivated (e.g., the two hands representing two panes of a window), though [Abner \(2017\)](#) provides evidence that this alternation is not limited to concrete object nouns in ASL but is also available to derive abstract, result-denoting nouns (e.g., ACCEPTANCE derived from ACCEPT). For practical reasons, experiments eliciting this contrast (including those detailed in the current manuscript) are limited to the concrete object portion of this paradigm, as these are easier to depict in video and pictorial stimuli.

Within sign language research, researchers have suggested that some of the strategies used to mark the noun–verb distinction—in particular, differences in the manner of movement and repetition—are based on iconic affordances of the categories ([Abner et al. 2019](#); [Aronoff et al. 2005](#); [Johnston 2001](#); [Kimmelman 2009](#); [Tkachman and Sandler 2013](#); [Wilbur 2008](#); [Wilcox 2004](#)). This iconic relationship has been suggested in particular to relate to the event structure of the verb ([Wilbur 2003](#)). [Supalla and Newport \(1978\)](#) observed that, while nouns in noun–verb form pairs are consistently distinguished in the same way, the specific form of the verb will depend on the properties of its event structure, consistent with the *Event Visibility Hypothesis*, which states that formal properties of predicates in sign languages reflect the semantics of event structure. For example, [Tkachman and Sandler \(2013\)](#) suggest that the continuous/restrained mapping for the manner of movement to verbs and nouns, respectively, represents a mapping of continuous and temporal aspects of the event structure of verbal forms. Similarly, [Kimmelman \(2009\)](#) suggests that verbal forms exploit embodied iconicity to signal events (i.e., that differences in sign movement might signal a difference in the movement of the event itself), which is less inherent in the noun mapping. In this way, systematic noun–verb distinctions that evolve over time may be seeded by the iconic properties of the underlying event descriptions. To further understand how these grammatical distinctions emerge, we turn to the evidence offered from studies of emerging linguistic systems.

1.2. The Emergence of Grammatical Categories

If the grammatical categories of nouns and verbs are fundamental to human languages, then we might expect them to emerge early in the creation of a novel linguistic system ([Bickerton 1990](#); [Heine and Kuteva 2007](#)). Currently, homesign systems and emerging sign languages provide some of the only natural examples of language creation and emergence, allowing us to observe novel linguistic systems through their earliest generations.

Homesign systems are gestural communication systems developed by children who do not have access to a conventional language model (i.e., profoundly deaf children born to hearing, non-signing parents). These systems are typically used within the immediate family and allow the child to communicate with other hearing family members (albeit with limited shared understanding; ([Carrigan and Coppola 2017](#))). The homesign systems developed by children demonstrate properties found in natural languages—stable lexicons ([Goldin-Meadow 2003](#)), grammatical roles ([Coppola and Newport 2005](#)), displaced references ([Morford and Goldin-Meadow 1997](#)), and relational marking ([Goldin-Meadow and Feldman 1977](#)). Studying homesigns can inform us about the distinctions that language creators introduce into languages without the benefit of a conventional language model. For instance, [Goldin-Meadow et al. \(1994\)](#) studied David, a deaf homesigner, who initially distinguished nouns from verbs by using completely different sign forms, even for related meanings. For example, he used *twist* for the verb form (twist-open), and *round-shape* for the noun form (jar). Later he used *twist* for both the verb and noun form, but marked the distinction in gesture form; he placed the *twist* serving the role of verb near an object (similar to inflecting a verb), and produced only one rotation for the *twist* serving the role of the noun (abbreviating the noun).

Homesign systems can be studied alongside emerging sign languages to further understand the impact that having a linguistic community has on marking grammatical distinctions. For example, Nicaraguan Sign Language (NSL) began to develop in the late 1970s when a new government policy established a school for deaf students in Managua.

The school allowed deaf children, who had developed homesign systems with their hearing families and had no access to other deaf individuals, to come together for the first time and share their homesign systems. Nicaraguan Sign Language (NSL) was born in this first cohort. As new deaf children entered the school, they learned the language (which changed in the course of learning) from the older children, thus forming a second cohort of NSL users.

Goldin-Meadow et al. (2014) analysed the consistency of handshape forms used for nominal and predicate constructions in Nicaraguan homesigners and the first and second cohorts of NSL. Overall, handshapes in nominal signs were less variable than handshapes in predicate signs, and the variability played different roles in the two types of signs. There was no variability across grammatical contexts (e.g., an agent vs. no-agent context) in nominals but considerable variability in predicates. Moreover, the variability in predicates was systematic across agent vs. no-agent contexts, suggesting that handshape functions as a productive morphological marker on predicate signs, even in homesign. In nominals, there was no variability across grammatical contexts (agent vs. no-agent). All of the groups, including homesigners, thus distinguished between forms playing nominal vs. predicate roles.

Similarly, Abner et al. (2019) analysed differences in form between nouns and verbs in three groups: ASL users, NSL users, and Nicaraguan homesigners, focusing on pairs of nouns and verbs with the same underlying form (e.g., camera vs. taking a photo). They analysed signs based on some of the properties that have been previously shown to mark noun–verb distinctions in natural sign languages (e.g., size, repetition). All three groups marked a distinction between nouns and verbs using utterance position (verbs were placed at the end of an utterance, nouns earlier in the utterance) and movement size (verbs were made with bigger movements, nouns with smaller movements).

There was, however, variation in whether a base hand and movement repetition were used to mark the noun–verb contrast. This variation offers insight into the pressures that influence the development of a linguistic system, and into cross-linguistic variation in the signed modality (ASL vs. NSL). The first cohort of NSL uses movement repetition and base hand just like homesigners do, but different from the second cohort who entered the NSL community later and learned a *pre-existing* system. This finding suggests that intergenerational transmission to new learners (not just sharing a language with other signers) plays a key role in the development of these particular devices. These results demonstrate not only the importance of the noun–verb distinction in human communication, but also how this distinction emerges and develops in a new (sign) language.

The evidence thus far suggests that distinctions between nouns and verbs are present in the earliest stages of novel linguistic systems. However, these distinctions may not initially be fully conventionalised or codified, but may instead become conventionalised through use with communication partners in the linguistic community. We present two experimental studies that aim to explore how the improvisation of novel signals by individuals, and the interaction between users of an emerging system, affect the noun–verb distinction. Using a silent gesture task in which hearing, non-signing participants are asked to communicate using only their hands, we assess whether participants spontaneously improvise distinctions between forms playing noun-like and verb-like roles, and whether those distinctions reflect those found in naturally emerging sign languages. We also introduce shared communication into our paradigm to explore whether communicative interaction affects the development of the distinctions. In this way, we investigate the extent to which distinctions between nouns and verbs in naturally emerging languages represent natural conceptual categories, and the extent to which they do and do not depend on shared communication.

1.3. Experimentally Modelling the Noun–Verb Distinction

Previous experimental research has demonstrated how methods such as silent gesture, artificial language learning, and experimental semiotics can be used to investigate

the pressures that shape language—specifically, pressures from the cognitive biases of individuals, and pressures from the social forces within a linguistic community (Beckner et al. 2017; Fay et al. 2010; Kirby et al. 2015; Motamedi et al. 2019; Nölle et al. 2018; Raviv et al. 2019; Silvey et al. 2019; Verhoef et al. 2014). These experiments elicit novel forms from participants across different media—gestures, drawings, non-linguistic vocalisations—to understand how participants create signals, how participants produce and interpret signals in the presence of a partner, and how signals evolve as they are used in interaction. For example, experiments investigating the creation of novel signals have shown that, in the absence of existing conventions, participants may rely on highly iconic forms to ground shared reference, but that these forms can become increasingly symbolic as they are used and conventionalised through communication (Fay et al. 2010, 2013; Garrod et al. 2007; Perlman et al. 2015; Sulik 2018; Theisen et al. 2010).

The experiments we present use the silent gesture paradigm to explore the evolution of a communication system in the manual modality. Hearing participants with no knowledge of sign language are asked to communicate using only gesture (without speech), a paradigm that has been shown to have limited influence from participants' existing linguistic knowledge (Gershkoff-Stowe and Goldin-Meadow 2002; Goldin-Meadow et al. 2008; Özçalışkan et al. 2016; Singleton et al. 1995). Silent gesture is a paradigm that has been widely used to understand the preferences participants have when creating novel signals. For example, a number of silent gesture studies have investigated word order in speakers of languages that exhibit different word order patterns, asking hearing participants who know no sign language to describe a series of events. Goldin-Meadow et al. (2008) found that participants with different linguistic backgrounds all produced verb-final word orders that mapped onto a Subject-Object-Verb (SOV) order when describing events in which an animate agent acts on an inanimate patient (e.g., MAN-GUITAR-PLAYS). More recent studies suggest that the preference for SOV may be mediated by a variety of factors, such as the semantics of the events (Schouwstra and de Swart 2014), the reversibility of the events (Gibson et al. 2013; Hall et al. 2013), or the possibility of iconic representation (Christensen et al. 2016; Meir et al. 2014).

Silent gesture is also a valuable tool to model the emergence of linguistic properties in the manual modality because it allows comparison with data from naturally emerging sign languages. By embedding silent gesture into an interactive framework in which participants use the gestures they create to communicate with a partner, we can model the processes enacted in the early emergence of sign languages—when signers bring their own homesign systems to a community of deaf individuals, each of whom also has their own homesign system. Previous experimental research that has embedded silent gesture into an interactive framework has shown that novel manual systems adapt to the pressures involved in interaction, and result in conventionalised and communicatively efficient signals (Fay et al. 2013; Motamedi et al. 2019; Nölle et al. 2018; Schouwstra et al. 2016).

Here, we model the processes of improvisation (the creation of novel signals) and interaction (use of signals with others who are also using signals) to understand how conventionalised noun-verb distinctions emerge in a manual communication system. We compare the systems resulting from these processes to two stages in the emergence of Nicaraguan Sign Language: (1) homesign, where children without a language model improvise their own communicative systems, and (2) interaction in the first cohort, where the formation of a linguistic community leads to the conventionalisation of signals from the improvised communicative systems. We asked participants to improvise gestures for a set of event scenes devised by Abner et al. (2019), and then use those gestures in interaction with a partner. We analysed the gestures participants produced using the coding system developed by Abner et al. (2019). We predict that the strategies used to distinguish nouns and verbs that have been found in the earliest stages of language creation (i.e., in homesign: the preference for verb-final ordering) will be present in the gestures that our participants create. However, the strategies that are found only in later cohorts of NSL and in ASL may be absent from the gestures that our participants create. We further predict that the

distinctions participants improvise will rely on the iconic affordances of the modality, as suggested by [Wilbur \(2008\)](#) and [Kimmelman \(2009\)](#), for example, with gesture size and repetitions iconically representing properties of individual items.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Experiment 1 was conducted at the University of Edinburgh and recruited participants from the university's Careers Hub website. Forty participants (aged 18 to 27, median age = 20; 13 male) took part in a study that required them to produce and interpret gestures for a set of video scenes that depicted objects being used either in a scenario that was either typical or atypical for the object. Participants were paid GBP 7 to take part in the experiment. All participants were self-reported to be right-handed "native" English speakers, with no knowledge of any sign language. Participants took part in the experiment concurrently with another participant, who acted as their communication partner in stage 2 of the task (see Section 2.1.3), giving a total of 20 pairs. Data from 5 pairs were not included in the analyses due to technical errors in video recording.

2.1.2. Materials

We used a set of video scenes designed to show target objects used in either a typical or atypical context. For example, a scene in which a man takes a photo with a camera shows a camera being used in a typical context; a scene in which a man digs with a camera shows a camera being used in an atypical context (Figure 1). Typical scenarios are expected to elicit gestures for typical actions ('take picture with camera'), and thus more verb-like productions. Atypical scenes are expected to elicit gestures related to the target object ('camera'), and are thus more noun-like productions. We selected a set of 24 vignettes, showing 12 objects used in typical contexts and 12 used in atypical contexts. The video scenes we used were a subset of those used in the study reported by [Abner et al. \(2019\)](#), for which objects were selected that would be familiar to participants in the United States and in Nicaragua and which would elicit different types of movements. The subset of vignettes was selected such that each type of atypical use (e.g., drop in bin, drop in water glass) was used with at least 2 objects².

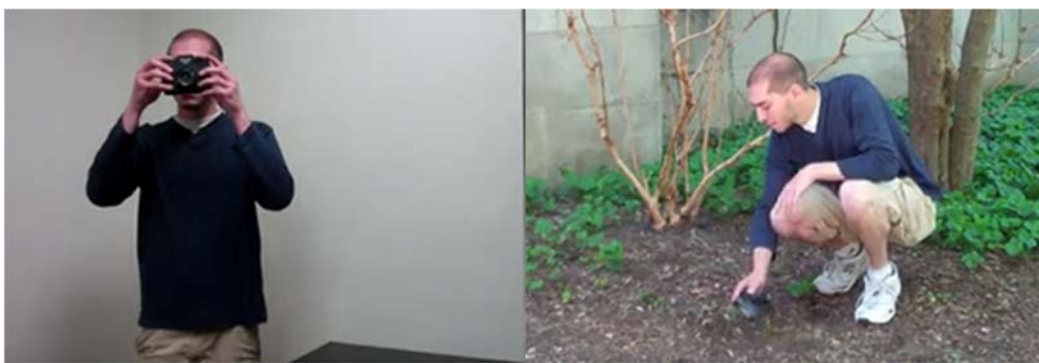


Figure 1. Example of scenes used in the study. (Left) a typical-use scene showing a person taking a photo with a camera. (Right) an atypical-use scene, showing a person digging with a camera.

Participants completed the experiment in individual experiment booths for the duration of the experiment. The experiment was run on an Apple Thunderbolt monitor, attached to an Apple Macbook Air laptop. Video recording was done using a Logitech webcam, also attached to the laptop, and the experiment ran using Psychopy ([Peirce 2007](#)). Video streaming and recording used VideoBox, custom software designed to enable streaming and recording between networked computers ([Kirby 2016](#)).

2.1.3. Procedure

The experiment comprised three stages. In the first stage, the *improvisation* stages, participants produced gestures for each vignette individually, without communicating with another participant. In the second stage, termed the *interaction* stage, participants communicated with their partners, producing and interpreting a gesture for each vignette. In the third stage, another *improvisation* stage, participants again produced gestures individually so that we could see whether any changes introduced in stage two were retained in stage 3 (Figure 2). Throughout the experiment, participants communicated using only manual gestures. Participants were instructed not to use speech when gesturing (audio was not recorded), nor to use fingerspelling of any kind. Participants were also asked to remain seated throughout the task.

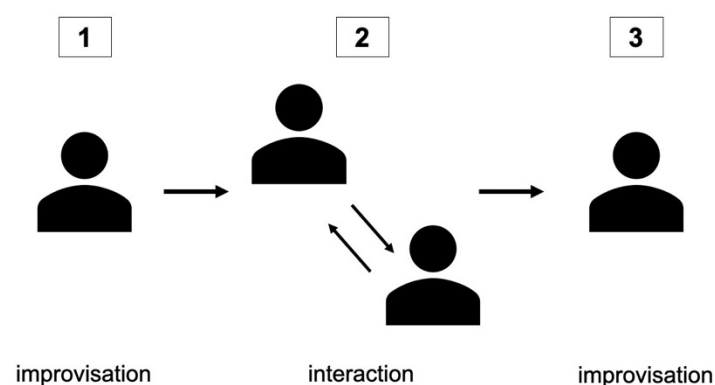


Figure 2. Stages in experiment 1. Participants take part in 3 stages: first, they take part in an improvisation stage, producing gestures to describe each vignette. They then take part in an interaction stage, producing and interpreting gestures in interaction with a partner. Finally, they complete a second improvisation stage.

In the first and third stages, participants were presented with each vignette, in random order, and asked to produce gestures to communicate each scene. One vignette was shown and a gesture was elicited at each trial. Participants were given a 3 s countdown to prepare them for the beginning of each trial. The vignette was shown on the screen, playing through twice, before participants were instructed to communicate the scene they had watched to the camera, using only gestures. Participants were again shown a 3 s countdown, this time to prepare them for recording. When recording began, participants saw themselves onscreen (mirrored) in the VideoBox window. Instructions were shown onscreen throughout the trial, informing participants to press the space bar to stop recording and move on to the next trial. Participants completed trials for all 24 vignettes. The procedure was identical for both improvisation stages.

In the intervening interaction stage, participants took turns with a partner to produce and interpret gestures, in a director–matcher task. Participants both produced and interpreted gestures for each vignette, giving a total of 48 trials in the interaction stage (i.e., each participant acted as director and receiver for all 24 vignettes). Participants switched roles at each trial, and the presentation of the scenes in each trial was randomised. Participants remained seated in individual experiment booths, and communication was enabled by streaming video between networked computers.

As director, the participant was asked to produce a gesture to communicate the vignette to their partner. After a 3 s countdown, participants were shown a vignette, twice through, as in the improvisation stages. They were then instructed to communicate the scene they had just watched to their partner. A 3 s countdown prepared them for recording and streaming to their partner. The participant's gesture was streamed to the networked computer operated by the matcher; the director saw themselves mirrored onscreen at the same time. Either director or matcher could stop the recording and streaming by pressing the space bar. When streaming was terminated, the director had to wait for the matcher to

guess what the gesture meant. Both participants were given feedback, and the experiment continued to the next trial.

As matcher, participants were given a 3 s countdown to signal the start of the trial, but were shown text on the screen reading “Waiting for partner” whilst the director watched the vignette. The matcher then received a synchronised 3 s countdown to prepare them for the start of streaming and recording. The matcher saw their partner’s gesture, unmirrored, on screen. The matcher could terminate streaming by pressing the space bar when they felt they had understood their partner’s gesture. Once streaming had been terminated, the matcher saw a set of 4 vignettes and made their guess. The 4 vignettes were chosen from vignettes used throughout the experiment, and comprised the target vignette (correct response) and three foils, determined as follows:

1. **Target object–foil context:** a vignette sharing the target object but showing the non-target action. For example, if the target vignette shows the typical camera context (taking a photo), then this foil would show an atypical camera context (e.g., dig with camera, shown in Figure 3, image 2).
2. **Foil object–target context:** a vignette sharing the target context (typical or atypical) for a different object. For example, if the target vignette shows the typical camera context, this foil would show the typical context for another object (e.g., cut with scissors, shown in Figure 3, image 3).
3. **Foil object–foil context:** a vignette that does not match the target vignette on either object or context, but does match the other foils. For example, if the target vignette shows the typical camera context, and the first foil shows the typical context for scissors, then this foil would show the atypical context for scissors (shown in Figure 3, image 1).

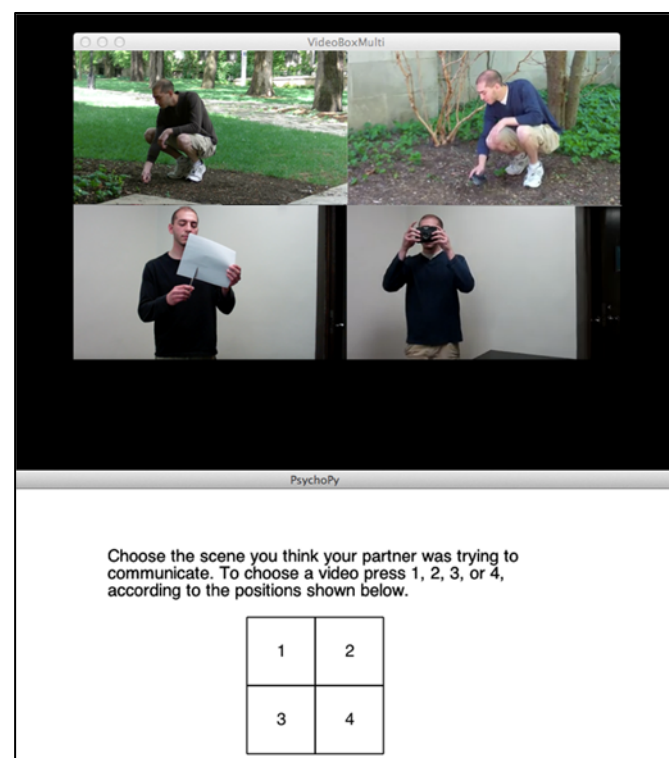


Figure 3. Example of a matching trial. The participant is shown a target and 3 foil videos playing in a loop on screen, and asked to select the video they think their partner was trying to communicate.

The target and 3 foils were presented as a grid of 4 looping videos. The matcher made their guess by pressing the number (from 1–4) of the corresponding video, as indicated in a dummy grid presented below the videos (see Figure 3). Once the matcher responded,

both participants were given full feedback. If the matcher's guess was correct, they saw the target video highlighted in green, and the director saw the target video on screen. If the matcher's guess was incorrect, the selected video and the target video were highlighted on the screen in red and green, respectively. In this case, the director saw the target video and the selected video. Both participants also received text feedback on screen, reading either "Correct" or "Incorrect". Feedback showed onscreen for 8 s before the experiment software automatically continued to the next trial, giving participants enough time to see both the target and the selected videos.

2.1.4. Gesture Coding

Here, we analyse the gestures produced in the two improvisation rounds, the first round (before interaction) and the final round (after interaction). Gesture sequences produced at each trial (describing single vignettes) were glossed and coded using ELAN (Sloetjes and Wittenburg 2008) by members of the research team. Individual gestures in a sequence were given a gloss describing each gesture (e.g., take photo), and then category codes were assigned to each gesture denoting 4 main categories:

- **Target action:** gestures representing the action related to the object used in the stimulus; a functional action showing how the target object is used. For example, for the target object camera, the target action would be taking a photo with a camera
- **Target other:** gestures related to the target object, which are not the functional actions associated with that object. For example, for nail polish, a gesture showing the action of opening the nail polish bottle in order to perform the target action of painting the nails.
- **Not related:** gestures not related to the target object, but some other component of the scene, such as the glass of water or bin used in some of the videos.
- **Verb:** gestures representing the atypical verb. For example, drop, dig, cover.

Following Abner et al. (2019), our goal is to analyse some of the formal features that distinguish noun and verb signs across natural sign languages (e.g., size, number of repetitions), in gestures that share a similar underlying form. For example, in Figure 4a, the participant produces two different gestures for typical and atypical scenes featuring the target item egg: in the left-hand panel, she gestures the target action of cracking an egg; in the right-hand panel, she positions her right hand as if holding an egg. Because the participant has chosen two distinct forms to represent the egg, we cannot compare features of the gestures in the typical and atypical contexts. In contrast, in Figure 4b, the participant produces gestures that have the same underlying form for typical and atypical scenes featuring the target item hammer: the participant's hand (or hands) moves as if manipulating a hammer in both cases. By comparing gestures with the same underlying form, we can examine if, across typical and atypical contexts, participants selectively use different features to distinguish productions in contexts designed to elicit noun forms vs. contexts designed to elicit verb forms. Therefore, we take the **target action (TA)** gestures produced in a sequence to be the participant's representation of the intended target (i.e., camera), and we compare TA gestures for the same object that the participant produced in its typical and atypical context³. We code these TA gestures for the following formal features known to distinguish nominal and verbal signs in natural sign languages:

- **Base hand use:** the use of a non-dominant hand in a stationary gesture acting as a ground for the dominant hand (e.g., representing the wall in a hammering gesture). Only two-handed asymmetrical gestures (such as representing hammering a wall) can be coded for base hand use (i.e., symmetrical two-handed gestures cannot be articulated with a base hand).
- **Gesture location:** We note the location of the gesture as either placed on the body (specified as eyes, mouth, ear, shoulder, torso) or in neutral space (specified in different zones related to height and laterality of the gesture).

- **Gesture size:** We code gestures as comprising local movement only (articulated using the wrist, thumb or finger joints) or path movement (the elbow, shoulder and trunk are involved in the movement; note that this code subsumes local movements).
- **Repetitions:** We note whether or not there is a repetition within a single gesture unit (target action).

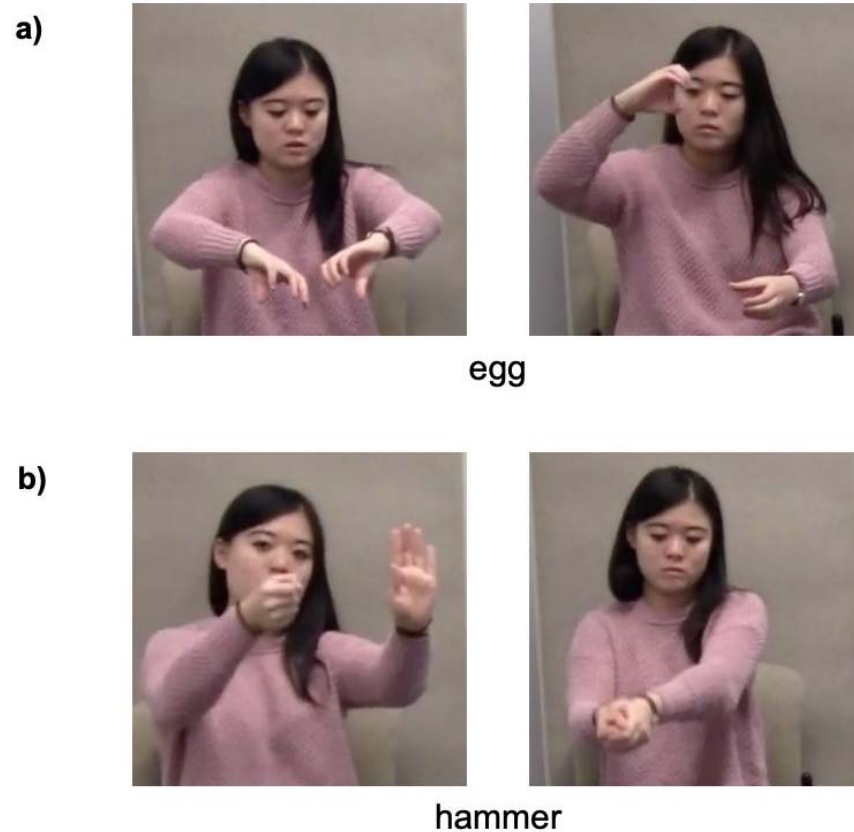


Figure 4. Example of gestures representing targets: (a), representing the target item egg where the two forms (left, right) have different underlying representations (b) representing the target item hammer, where the two forms (left, right) have the same underlying representations.

Two coders completed coding for data from study 1. A subset of 20% of the data (spanning data from each coder) was second-coded by KM and reliability between this sample and the original coding was calculated using Cohen's Kappa (Cohen 1960) for target action coding and for each of the formal parameters. We found very high agreement for our variables of interest: first target action ($\kappa = 0.93$), base hand ($\kappa = 0.85$), gesture size ($\kappa = 0.89$), gesture location ($\kappa = 0.88$) and repetitions ($\kappa = 0.88$). The full coding scheme can be found at <https://osf.io/qzgjt> (accessed on 21 March 2022).

2.2. Results

We analyse our measures using mixed effects models, implemented with R (R Core Team 2013) and lme4 (Bates et al. 2015), including context (typical/atypical) and round (first/final) as deviation-coded binary predictors, as well as their interaction. We use the maximal model (including all slopes and intercepts) that allows convergence, including intercepts for item and participant, nested in pairs. Where models do not converge, we (i) test model fit with different optimizers, (ii) remove correlations between slopes and intercepts, and (iii) remove slopes with the lowest variance. The full specification for each model can be found at <https://osf.io/qzgjt> (accessed on 21 March 2022).

Sequence length First, we analyse the overall length of gesture sequences for typical and atypical scenes (Figure 5a⁴), using a mixed effects Poisson regression model for

count data. A model including both round and scene context, as well as their interaction, demonstrated a better fit than a reduced model ($\chi^2 = 8.47$, $p = 0.003$). The model revealed a significant main effect of context, such that typical scenes were shorter than atypical scenes ($\beta = -0.39$, $SE = 0.08$, $z = -5.14$, $p < 0.001$), and an interaction between round and context ($\beta = -0.21$, $SE = 0.07$, $z = -2.92$, $p = 0.003$). That is, participants produce longer gesture sequences for atypical compared to typical scenes, but this difference reduces over rounds once participants converge on conventional ways to communicate targets in the atypical contexts.

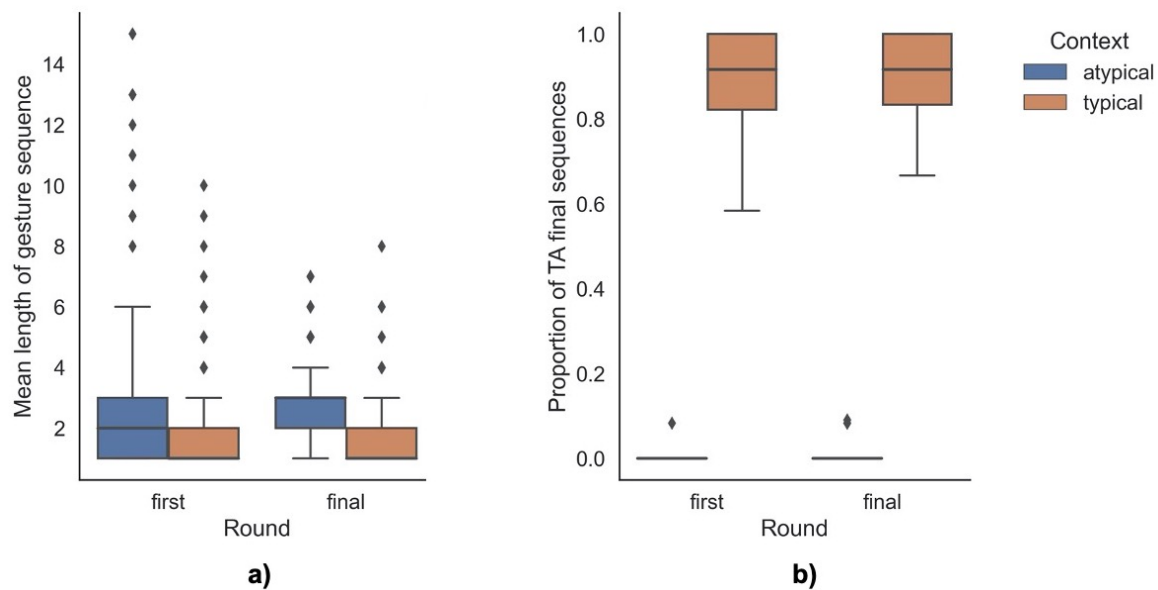


Figure 5. Mean gesture sequence length in terms of the number of individual gestures produced in a sequence (a) and Proportion of gesture sequences with a target action (TA) in final position (b), shown for each round and each context (typical/atypical).

Target action position. We assess differences in how target actions are positioned in a gesture sequence, using a logistic mixed effects model to analyse how often target action gestures appear in the final position in a sequence (Figure 5b). For example, in a camera event, does the target action gesture (taking a photo with a camera) appear at the end of a gesture sequence or elsewhere in the sequence? We present here a model including only context as a fixed effect, as including round did not improve model fit ($\chi^2 = 1.21$, $p = 0.27$). Participants show a strong preference for producing target actions at the end of the sequence in typical contexts, and rarely produced target actions at the end of the sequence in atypical contexts ($\beta = 10.97$, $SE = 1.72$, $z = 6.36$, $p < 0.001$).

In our remaining analyses, we focus on gestures that are directly comparable across typical and atypical scenes—those coded as TA gestures. Though some responses did include multiple TA gestures, we include only the first instance of each TA gesture produced in a sequence (only ~11% of all trials contained more than one TA gesture within the same sequence).

In the following measures, we analyse how often participants' productions differ between typical and atypical contexts based on the four formal properties of gestures we coded: base hand use, gesture location, gesture size, and repetitions. If participants produce distinctions based on scene type, we expect typical contexts to elicit verb-like gestures and atypical contexts to elicit noun-like gestures, varying the gesture properties in ways similar to those found in natural sign languages (i.e., more base hand use for verbs, more repetitions for nouns).

Base hand use. The proportion of scenes in which participants use a base hand for each round and context is illustrated in Figure 6a. We analysed the presence of base hand gestures at each trial using a logistic mixed effects model; the model including round did

not show improved fit over the model including only context ($\chi^2 = 0.79$, $p = 0.37$). The model revealed a significant main effect of context, with base hand use more common in typical than atypical scenes ($\beta = 3.19$, $SE = 0.96$, $z = 3.32$, $p < 0.001$).

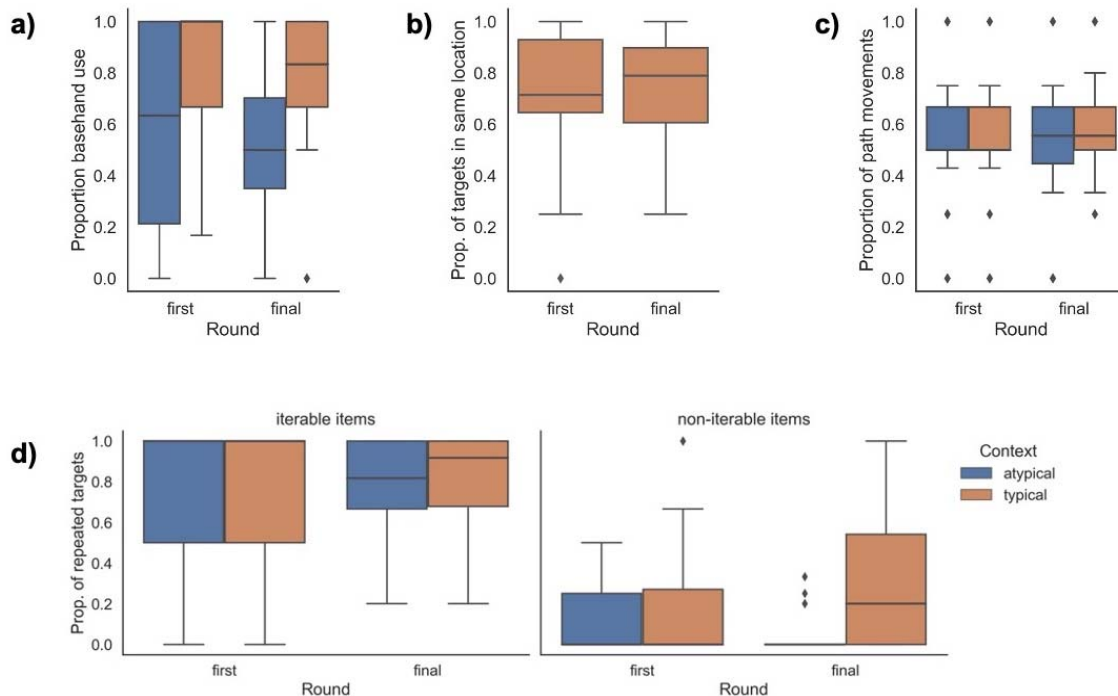


Figure 6. Gesture form analyses for experiment 1, showing base hand use (a), proportion of target action gestures produced in the same location for typical and atypical contexts (b), gesture size (c) and repetitions shown for iterable and non-iterable target actions (d).

Location. We used a logistic mixed effects model to analyse whether at each trial participants gesture target actions in the *same* location across typical and atypical contexts (see Figure 6b). Model comparison indicated that including round did not improve fit compared to the null model ($\chi^2 = 0.49$, $p = 0.48$). The model revealed a significant intercept ($\beta = 1.39$, $SE = 0.44$, $z = 3.17$, $p = 0.002$), suggesting that, on average, participants gesture TAs in the same location across contexts.

Size. We analyse gesture size as how often participants produce target action gestures with path movements (shown in Figure 6c), using a logistic mixed effects model. Models including only context ($\chi^2 = 0.25$, $p = 0.61$) and only round ($\chi^2 = 0.48$, $p = 0.49$) did not improve fit over a null model. The grand mean from the model intercept did not suggest a reliable preference for path movements overall ($\beta = 1.28$, $SE = 1.81$, $z = 0.71$, $p = 0.48$).

Repetitions. We analyse how often target actions are repeated in gestures across typical and atypical contexts using a logistic mixed effects model, adding an additional deviation-coded predictor (including all interactions) of iterability. Some of the events can elicit target actions that can be, and typically are, iterated (e.g., a hammering gesture); other events typically achieve their goal with one movement and thus are not usually iterated actions (e.g., putting on a ring). Our findings are illustrated in Figure 6d. A model including all 3 main effects without interaction terms suggested improved fit over a reduced model without round ($\chi^2 = 5.97$, $p = 0.01$). We found a significant main effect of scene type, such that gestures for typical scenes were produced more often with repetitions than gestures for atypical scenes ($\beta = 0.92$, $SE = 0.35$, $z = 2.65$, $p = 0.008$). We also found a main effect of iterability, with non-iterable items demonstrating fewer repetitions ($\beta = -4.23$, $SE = 0.77$, $z = -5.47$, $p < 0.001$).

Convergence. Finally, we analyse the extent to which communication between partners has affected the gestures they produce between the first and final production rounds.

We compare gestures produced across pairs of participants (paired in the interaction stage) with pseudo-pairs, matching pairs of non-interacting participants, to assess the specific role communication has in shaping the systems participants produce. We compared the pairs on the four formal properties (base hand, location, size, and repetitions) for target action gestures, and calculated the proportion of those properties that pairs converge on for each target scene (illustrated in Figure 7). We analyse the proportion of form parameters that are the same for paired participants using a logistic mixed effects model, with the proportions weighted by the number of parameters, including fixed effects of round and pair type (both deviation-coded). We include by-pair and by-item random intercepts with a random slope of round for the by-item intercept (including a random slope with the by-pair intercept led to singular fit). The model including the interaction term did not improve fit over the model without ($\chi^2 = 0.24$, $p = 0.63$). Inspection of the model indicated main effects of round ($\beta = 0.10$, $SE = 0.04$, $z = 2.41$, $p = 0.02$) and pair type ($\beta = 0.19$, $SE = 0.06$, $z = 3.13$, $p = 0.002$)—participants produce more similar gestures to other participants in the final round than in the first but, importantly, similarity is greater for interacting pairs than for pseudo-pairs.

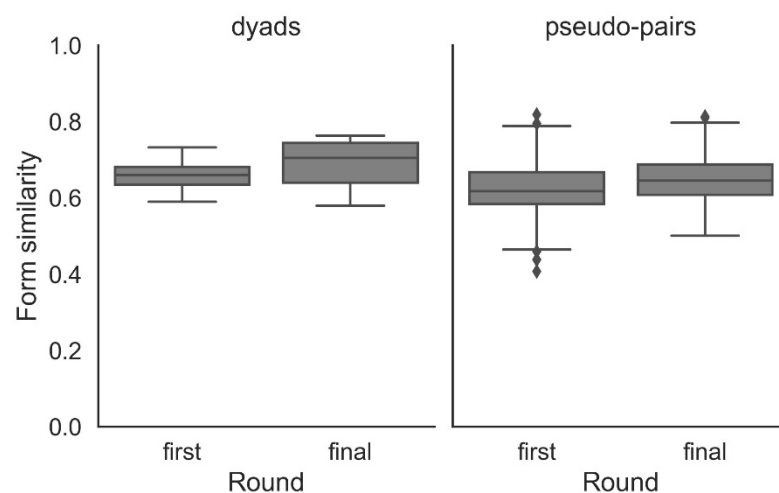


Figure 7. Similarity in form parameters across rounds for real-paired dyads and pseudo-dyads (i.e., who did not interact during the experiment).

2.3. Experiment 1 Summary

In experiment 1, we examined gestural production in contexts aiming to elicit noun-like and verb-like gestures for target objects, investigating how participants' improvised gestures change after interaction with a partner. Our findings suggest that, even in improvised gestures, participants make distinctions between descriptions of targets designed to elicit nouns and targets designed to elicit verbs. Gesture sequences describing typical scenes tended to be shorter than those describing atypical scenes. Gestures for target actions were primarily produced in final position for typical (i.e., verb-eliciting) contexts, but were rarely produced in final position for atypical (i.e., noun-eliciting) contexts. We also found that target action gestures for typical targets were more frequently produced with a base hand gesture than target action gestures for atypical targets, and typical targets were more frequently repeated than atypical targets. Our findings for target position and base hand use reflect distinctions found in ASL, NSL and those made by Nicaraguan homesigners, as reported by Abner et al. (2019). These patterns suggest that some features distinguishing nominal and predicate forms can emerge even in the earliest stages of a communication system. However, we do not find distinctions based on the features of gesture location, gesture size, nor do we see further systematisation of the distinctions following communication. Analysing the convergence between interacting dyads and pseudo-pairs of participants reveals the role interaction plays. We find some patterns of convergence across pseudo-pairs, highlighting general pressures (i.e., iconicity) that

may affect gesture similarity. However, interacting participants produce gestures that are more similar to each other's than pseudo-pairs of participants, suggesting that similarities between the gestures produced by interacting participants cannot be attributed solely to iconic representations that would be similar across all participants.

In experiment 2, we further explore how communicative constraints affect the distinctions between gestures produced to signal typical and atypical contexts. In experiment 1, we used a constrained model of communication, a reductionist operationalisation in which participants take set turns to produce and interpret gestures, and receive comprehensive feedback on their successes and errors. As discussed by [Kocab et al. \(2018\)](#), it is possible that some of the constraints in operationalisations of communicative behaviour do not always map well onto natural language use, and that currently, such operationalisations do not account for the full range of behaviours that comprise communication in situated, face-to-face interactions. Such interactions in the real world involve conventions related to turn-taking ([Stivers et al. 2009](#)), alignment ([Garrod and Pickering 2009](#)) and repair ([Dingemanse et al. 2015](#)) that are not possible to enact in the reduced operationalisation we use in experiment 1. In experiment 2, we investigate the same research questions using a more ecologically valid operationalisation of communication, in which turn-taking and feedback about communicative success or failure are under the control of the interacting participants themselves. Furthermore, we contrast the interactive scenario with a condition in which individual participants repeatedly improvise gestures for our event vignettes, without interacting with a partner.

3. Experiment 2

3.1. Methods

3.1.1. Participants

Forty participants took part in experiment 2 (24 female), recruited from the undergraduate population at the University of Chicago. In total, 20 participants took part in the individual condition, and 20 participants took part in a dyadic condition, in which they were paired with another individual (i.e., 10 dyads). Participants were reimbursed for taking part in the study with either a payment of USD 10 or one required research course credit. All participants were self-reported "native" speakers of English with no knowledge of ASL (10 participants reported speaking languages in addition to English). A session with one dyad (two participants) was excluded and re-recorded with a new pair of participants due to a technical error in video recording.

3.1.2. Materials

As in experiment 1, participants were asked to communicate about a set of events shown in vignettes selected from the stimuli used by [Abner et al. \(2019\)](#). A total of 32 vignettes were selected⁵, showing 16 unique items used in typical and atypical contexts. We also manipulated the iterability of the vignettes: half of the items were typically used in an iterable manner, where the action is repeated (e.g., *rocking the baby*); half were typically used in a non-iterable manner (e.g., *putting on a backpack*). The experiment took place in a private room and ran on an Apple Macbook Pro laptop using Microsoft Powerpoint. All vignettes were presented in a randomized order. Video recording was done using a Canon Vixia HF R800 camcorder mounted on a tripod. Participants were seated in the room with the laptop on a table beside them, facing the camera during the improvisation stage and facing their partner during the interaction stage.

3.1.3. Procedure

Participants took part in one of two conditions, a dyadic condition and an individual condition (shown in Figure 8). Throughout the experiment, participants were told not to speak or use mouthing, and were told to remain seated for the duration of the task.

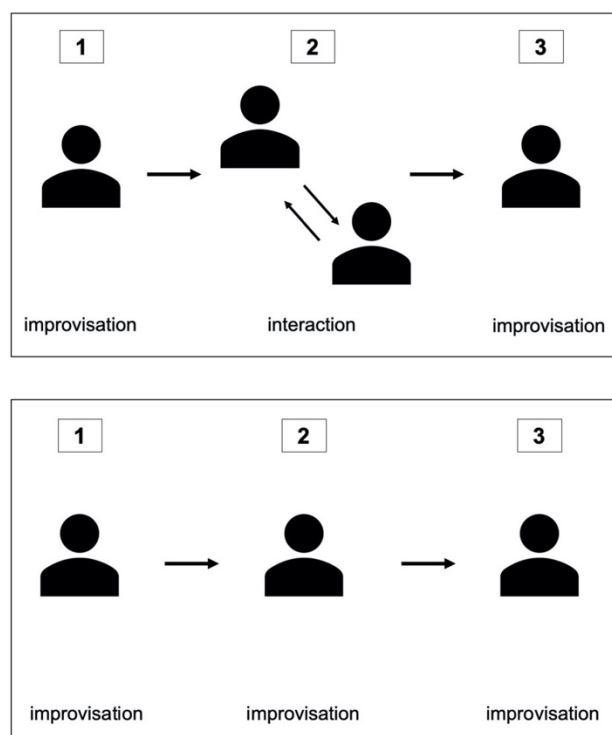


Figure 8. Stages in experiment 2, for the dyadic condition (**top**), identical to experiment 1, and the individual condition (**bottom**), where participants take part in 3 improvisation rounds.

All participants in both conditions were given 3 practice items before the main study began. Participants were shown a vignette and asked to describe it to the experimenter using gestures. Participants who did not produce an explicit gesture for the target item in the vignette (e.g., camera, baby) were instructed to do so by the experimenter. After successful completion of the practice items, the experimenter left the room. The participant controlled the progression of the experiment using the arrow keys to move from trial to trial. Participants were allowed to repeat each vignette as needed before responding. After the initial production stage, the experimenter re-entered the room only to set up the experiment for the following stage or to give instructions preceding the communication stage.

The dyadic condition largely replicated experiment 1 in structure. Participants first completed an initial improvisation stage before taking part in an interaction stage with a partner. However, in the interaction stage, participants were seated in the same room, facing each other, with the computer displaying the vignettes in sight. At each trial, both participants could see the target vignette playing on the computer. After they had watched the video, one participant, the designated gesturer, had to describe the vignette to their partner using only gestures. No feedback was required from their partner and no feedback was given from the experimenter. However, there were no other constraints on how participants could communicate with gestures during the session and paired participants were free to provide gestural feedback to each other, or enact repair strategies on their own productions. Note that this lack of constraints stands in contrast to experiment 1, where participants were physically separate and only interacted via webcam streams between computers. Once the gesturer had finished their gesture, the experiment proceeded to the next trial. Participants switched roles at each trial, producing a gesture on every other trial. Participants completed a total of 64 trials, each producing a gesture for all 32 items. Following the interaction round, participants completed a final improvisation round, identical to the first.

In the individual condition, participants completed three improvisation rounds, each using 3 different randomised sets of all 32 vignettes. Participants produced gestures to communicate each vignette to the camera without a partner, across all three rounds (i.e., no

communication took place). There was a brief break period in between rounds while the experimenter set up the next stimulus set.

3.1.4. Coding

Gesture coding for experiment 2 was identical to the coding carried out for experiment 1 and coding for both experiments was carried out concurrently. As for experiment 1, two coders completed coding for the data and a subset of 20% of the data was second-coded by KM to calculate the reliability. Cohen's Kappa (Cohen 1960) indicated high agreement across our variables of interest: first target action ($\kappa = 0.93$), base hand ($\kappa = 0.84$), gesture size ($\kappa = 0.92$), gesture location ($\kappa = 0.84$) and repetitions ($\kappa = 0.92$).

3.2. Results

Analysis for experiment 2 largely follows the analysis for experiment 1, with the additional inclusion of group (individual vs. dyad) as a deviation-coded fixed effect, along with context, round, and all interaction terms. Model selection follows the same procedure as experiment 1 and a full specification for each model can be found at <https://osf.io/qzgj1> (accessed on 21 March 2022).

Sequence length. The model including only scene type demonstrated an improved fit over the null model ($\chi^2 = 25.20$, $p < 0.001$); adding additional fixed effects did not improve model fit. Inspection of the model suggests a main effect of context ($\beta = -0.48$, $SE = 0.06$, $z = -7.64$, $p < 0.001$), with participants producing shorter sequences on average for typical contexts, compared to atypical contexts (illustrated in Figure 9a).

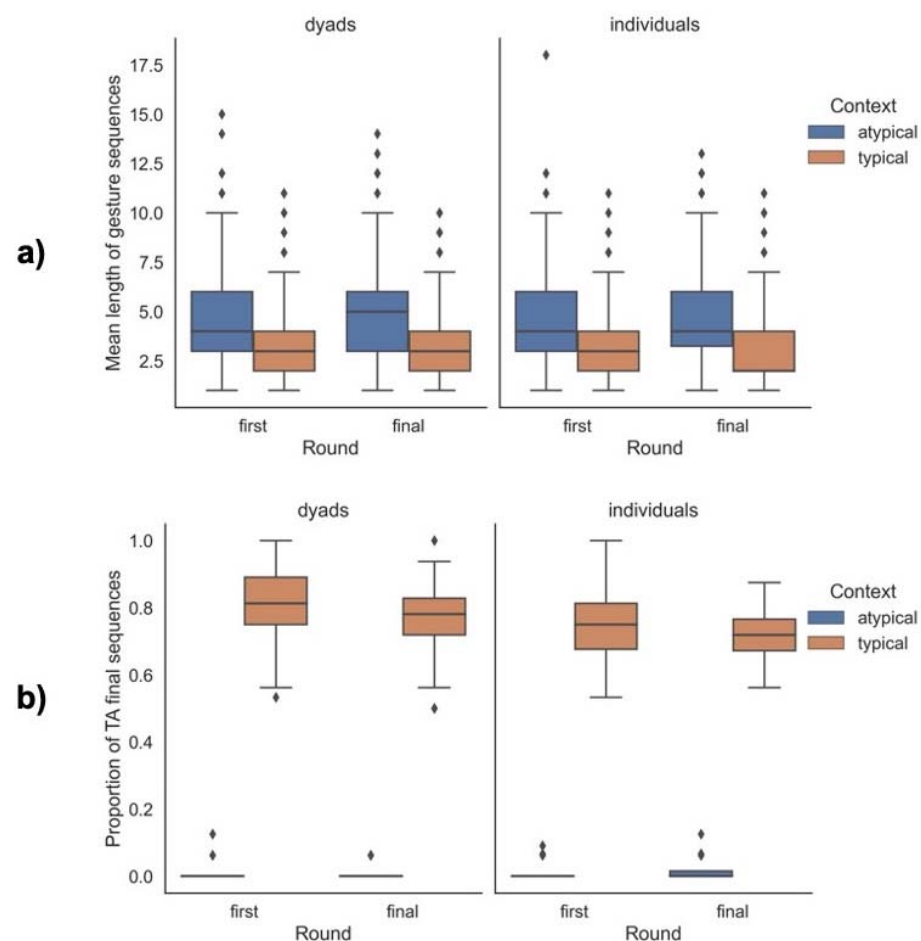


Figure 9. Mean gesture length (a) and Proportion of gesture sequences with target action (TA) in final position (b), shown for each round and each context (typical/atypical).

Target action position. Again, the model including only scene type (but no other additional fixed effects) showed improved fit over the null model ($\chi^2 = 41.62$, $p < 0.001$), with the model demonstrating preference for more target actions at the end of the sequence in typical contexts than in atypical contexts (Figure 9b; $\beta = 6.00$, $SE = 0.49$, $z = 12.23$, $p < 0.001$).

As in experiment 1, the remaining analyses focus on the first TA gesture found in each sequence, comparing matched TA gestures across typical and atypical trials.

Base hand use. We analysed the presence of base hand gestures (see Figure 10a) at each trial using a logistic mixed effects model. The model including all three main effects, as well as an interaction between round and group, showed improved fit over the model without the interaction term ($\chi^2 = 5.56$, $p = 0.02$). The model revealed a significant main effect of context, with base hand use more common in typical than atypical contexts ($\beta = 2.30$, $SE = 0.63$, $z = 3.65$, $p < 0.001$), as well as a significant interaction between round and group ($\beta = -1.12$, $SE = 0.46$, $z = -2.42$, $p = 0.02$), indicating an overall increase in base hand use between the first and final round for dyads only.

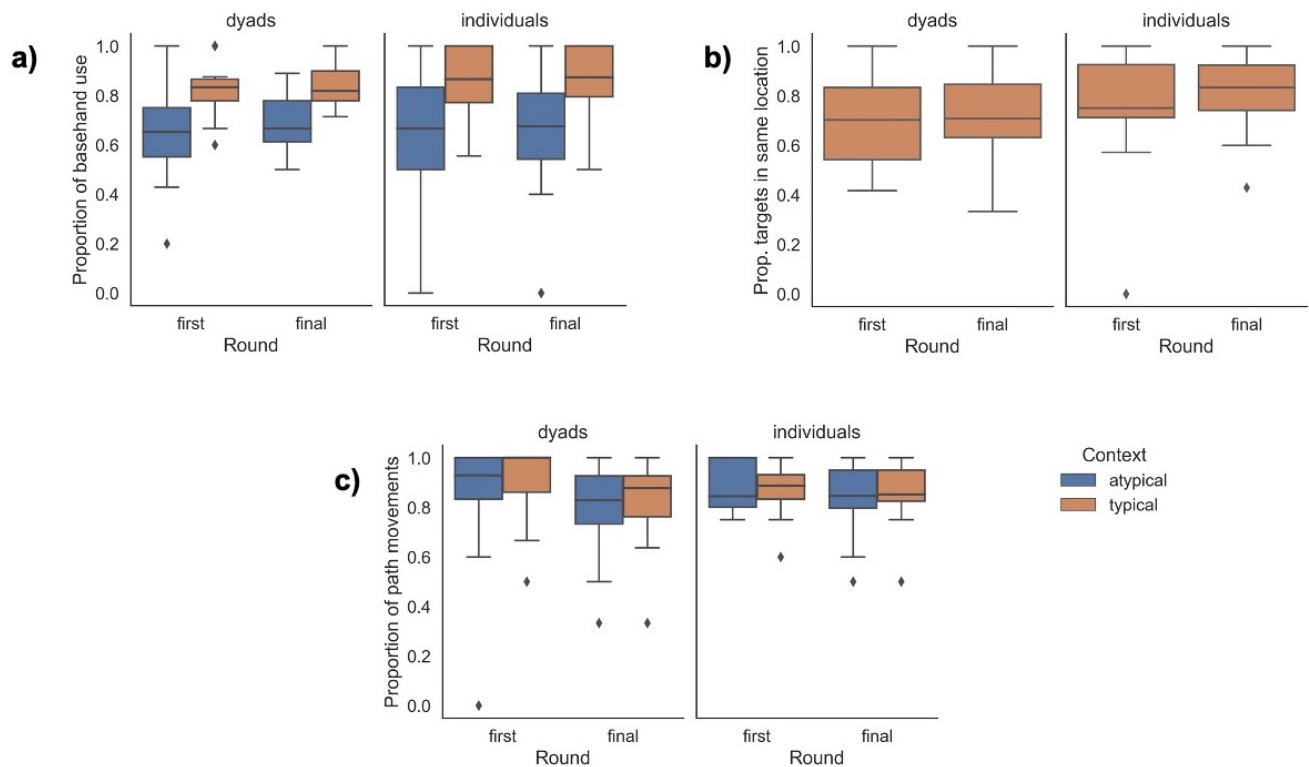


Figure 10. Gesture form analyses for experiment 2, showing base hand use (a), proportion of target action gestures produced in the same location for typical and atypical contexts (b), and gesture size (c) for both dyads and individuals.

Location. The proportion of typical and atypical targets gestured in the same location is shown in Figure 10b. Logistic mixed effects models including fixed effects of either group or round did not improve fit over the null model (group: $\chi^2 = 2.62$, $p = 0.11$, round: $\chi^2 = 0.01$, $p = 0.90$), and the model revealed a significant intercept ($\beta = 1.88$, $SE = 0.49$, $z = 3.81$, $p < 0.001$), indicating an overall preference across groups to place target action gestures in the same location in typical and atypical contexts.

Size. Figure 10c indicates that participants produce a high proportion of path gestures across rounds and contexts, for both dyads and individuals. Analysis using a logistic mixed effects model to predict path gesture production did not find an improved fit over the null model when including either context ($\chi^2 = 0.55$, $p = 0.46$), or context and round ($\chi^2 = 1.07$,

$p = 0.30$) as fixed effects, suggesting no reliable changes in the preference for path gestures across contexts and rounds.

Repetitions. We show the proportion of trials with repeated targets in Figure 11, and use a logistic mixed effects model including context, round, group, and iterability of the target action as fixed effects, along with their interactions. A model including all fixed effects and interaction terms showed improved fit over a reduced model ($\chi^2 = 26.32$, $p = 0.006$). Inspection of the model revealed a main effect of iterability ($\beta = -4.22$, $SE = 0.72$, $z = -5.87$, $p < 0.001$) and an interaction between round and iterability ($\beta = -1.39$, $SE = 0.36$, $z = -3.92$, $p < 0.001$). Participants in both groups produce more repetitions for iterable than non-iterable target actions. For iterable items, repetitions increase between the first and final round but decrease between the rounds for non-iterable items.

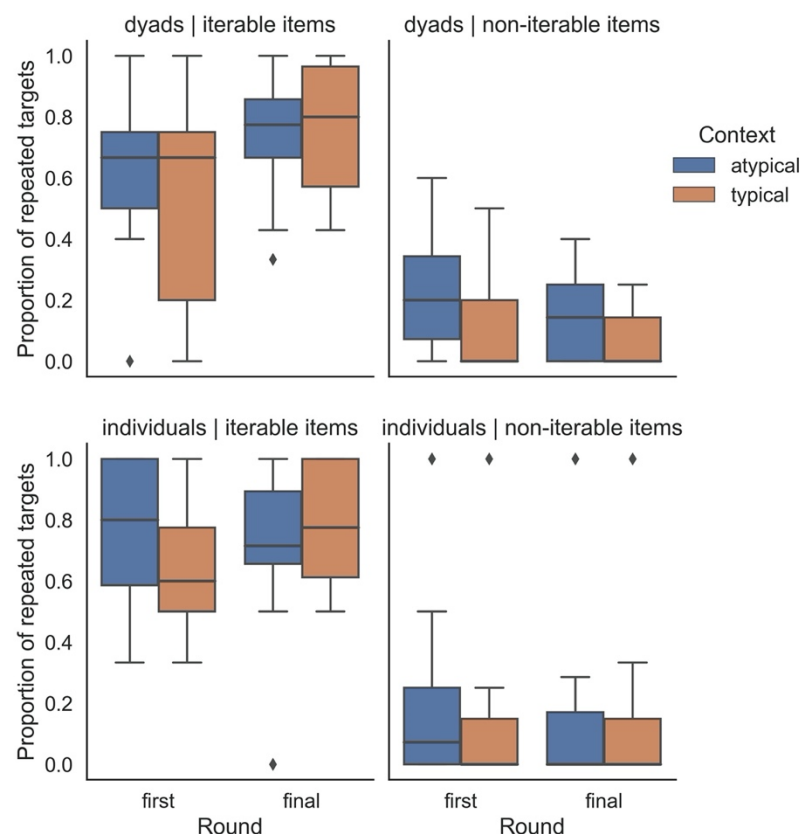


Figure 11. Proportion of repeated targets, shown for iterable and non-iterable target actions, and for participants in the dyadic and individual conditions.

Convergence. We measure convergence between pairs and pseudo-pairs on the different formal properties of gestures as in experiment 1. In addition, we include pseudo-pairs created from individual participants (who never communicate with a partner) matched with other participants in the same condition. Figure 12 shows the mean form similarity for each set of paired participants. We analyse form convergence using a logistic mixed effects model as described in experiment 1. We include round and pair type as fixed effects, with round deviation-coded. We include by-pair and by-item random intercepts with a random slope of rounds for both intercepts. Model comparison indicated that the model with the interaction between round and pair type did not improve fit over a reduced model without the interaction ($\chi^2 = 0.12$, $p = 0.94$). Inspection of the reduced model suggested a significant effect of round ($\beta = -0.13$, $SE = 0.06$, $z = -2.19$, $p = 0.03$), indicating that, across groups, we see a small reduction in the similarity between the first and final production rounds. We also find a significant effect of pair type for the pseudo-dyads ($\beta = -0.19$, $SE = 0.07$, $z = -2.78$, $p = 0.005$), but not for the pseudo-individuals ($\beta = -0.002$, $SE = 0.07$, $z = -0.04$, $p = 0.97$). Participants in the dyadic condition that did not interact with each other

demonstrated lower form similarity than participants who did interact with each other. Individuals who only produced gestures in isolation showed similar levels of convergence as participants who communicated together in dyads.

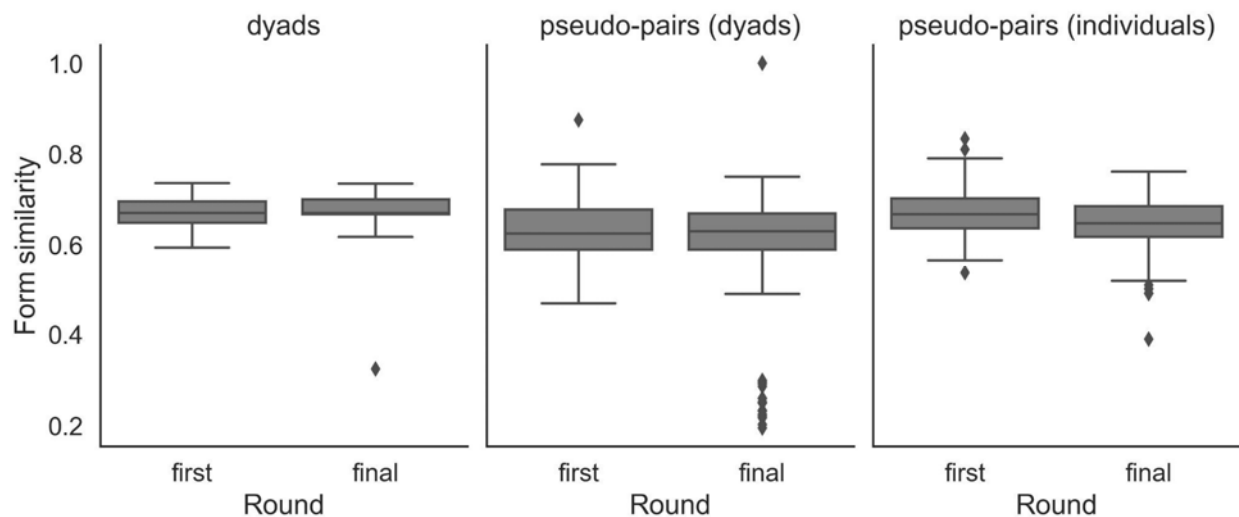


Figure 12. Similarity in form parameters across rounds for real-paired dyads (**left panel**) and pseudo-dyads, made up of participants in the dyadic condition paired with participants with whom they did not interact (**middle panel**) and participants in the individual condition (**right panel**), paired with other individuals with whom they did not interact.

3.3. Experiment 2 Summary

In experiment 2, we investigated the emergence of distinctions between gestures communicating noun-like and verb-like meanings during improvisation by individuals and following interaction between pairs of participants. We used an operationalisation of interaction that allowed for more unconstrained and organic turn-taking and repair strategies between participants than in experiment 1. We further compared productions by dyads before and after interaction with productions by individuals who repeatedly produced gestures over 3 rounds but without communicating the target scenes to a partner. We replicated findings from experiment 1. Participants produced shorter gesture sequences when describing targets in a typical context than in an atypical context. Participants were also more likely to place gestures for target actions in the final position of a sequence, and to use a base hand gesture, when describing typical (i.e., verb-like) contexts than atypical (i.e., noun-like) contexts. Finally, we found that the frequency of repetitions maps onto the iconicity of the event, with iterable items gestured with more repetitions than non-iterable items. Notably, our findings from individuals (not in dyads) align in key ways with those from dyads and from experiment 1, suggesting that, while communication allows participants in pairs to converge on a shared system, the distinctions that do emerge are not driven by communication but can emerge through improvisation alone.

4. General Discussion

The categories of nouns and verbs are among the basic elements of human language (Bickerton 1990; Hockett 1977; Jackendoff 2002). Here, we asked whether systematic formal distinctions between noun- and verb-like forms emerge in improvised gestures, and whether those distinctions further conventionalise over time and through interactions. In particular, our work closely follows that reported by Abner et al. (2019), tracking how similar features (base hand, size of movement, and repetition) distinguish noun and verb signs in ASL, NSL and Nicaraguan homesigners. Table 1 provides a summary of our findings in comparison to those reported by Abner et al. (2019).

Table 1. Summary of findings from current studies and those reported by [Abner et al. \(2019\)](#) for ASL, NSL and Nicaraguan homesigners.

	Silent Gesture		Abner et al. (2019), Natural Languages		
	Experiment 1	Experiment 2	ASL	NSL	Nicaraguan Homesigners
Utterance final position	Higher proportion of utterance-final targets for typical scenes	Higher proportion of utterance-final targets for typical scenes	Higher proportion of utterance-final verb targets	Higher proportion of utterance-final verb targets	Higher proportion of utterance-final verb targets
Base hand	Higher proportion of base hand use with targets for typical scenes	Higher proportion of base hand use with targets for typical scenes	Higher proportion of base hand use with verb targets	Higher proportion of base hand use with verb targets for more recent cohort signers	No reliable trends found.
Location	Same location used for both typical and atypical targets	Same location used for both typical and atypical targets	Not studied	Not studied	Not studied
Size of movement	No reliable trends found	No reliable trends found	Higher proportion of proximal movement with verb targets	Higher proportion of proximal movement with verb targets	Higher proportion of proximal movement with verb targets
Repetitions	Higher proportion of repetitions for iterable targets and for targets in typical scenes	Higher proportion of repetitions for iterable targets.	Higher proportion of repetitions with noun targets for iterable and non-iterable targets	Higher proportion of repetitions for iterable targets; higher proportion of repetitions for noun targets for later cohorts	Higher proportion of repetitions for iterable targets. Tendency to produce more repetitions for noun targets.

Across both experiments we report, participants make distinctions between gestures they produce for targets appearing in typical contexts (designed to elicit verb-like gestures) and atypical contexts (designed to elicit noun-like gestures). Gesture sequences for typical contexts are shorter than gesture sequences for atypical contexts. This difference in length is largely driven by the additional verb gesture used to describe the action in atypical contexts (e.g., dig, drop). The target object and target action can be conflated and articulated simultaneously for typical contexts (e.g., a *taking a photo* gesture contains information about the object, camera, and the action, taking a photo with a camera). In contrast, the atypical action must be specified separately from the target object (e.g., digging with a camera requires a *dig* gesture and also a *camera* gesture). The conflation of object and action in descriptions of typical contexts is not inevitable and, indeed, there are some examples of participants who produce gesture sequences where they specify object information in one gesture (e.g., tracing a rectangular shape to indicate the camera) before producing a target action gesture.

However, since producing object and action information in a single gesture is sufficient to describe the typical contexts in this study, object-only information is often left out of descriptions of typical contexts, rendering those descriptions shorter than descriptions of atypical contexts.

When we focus only on gestures for target actions that capture the same property in both typical and atypical contexts (e.g., pushing the button on a camera for the *taking-a-picture* event and for the digging with a *camera* event), we find that target actions tend to appear in the final position of a gesture sequence for typical contexts, but not for atypical contexts. Previous silent gesture experiments have suggested that participants from different language backgrounds show a preference for verb-final sequences for non-reversible events (Goldin-Meadow et al. 2008; Hall et al. 2013; Meir et al. 2014; Schouwstra and de Swart 2014), and verb-final order (specifically, SOV) is considered grammatical across all documented sign languages (Napoli and Sutton-Spence 2014). Finally, our results dovetail with those reported by Abner et al. (2019), who found that signers across all three groups they studied (ASL signers, NSL signers, and Nicaraguan homesigners) produced verb (but not noun) targets in the utterance-final position. Our findings are therefore consistent with an interpretation that target action gestures act like verbs in typical contexts, but like nouns in atypical contexts.

We also find that participants across experiments and conditions produce more base hand gestures for target actions in typical contexts than in atypical contexts. Abner et al. (2019) reported findings for distinctions made using base hand articulation, though their findings are somewhat complex. Their results suggested that, for NSL signers, only those who had entered the signing community relatively late (when a language model had been established), used base hand articulation more often with verb targets than noun targets. There was a tendency for a similar pattern in Nicaraguan homesigners, but only in some of the individuals. Notably, Abner et al. (2019) found that ASL signers demonstrated very limited use of base hand gestures for both verb and noun targets, suggesting that the grammatical function and role of the base hand can vary cross-linguistically. Where they are used, Abner et al. (2019) suggest, base hand gestures iconically represent additional event arguments (such as the wall being hammered against), not properties inherent to an object, and therefore we might expect them to appear in verb-like productions more frequently than noun-like productions. Indeed, many of the strategies used to distinguish nouns and verbs cross-linguistically in sign languages reflect iconic features of objects and events. These features can then be systematised to distinguish grammatical categories (Wilbur 2008). For example, repetition can iconically represent event iterability, as our participants demonstrate: more repeated gestures are used when describing iterable events than non-iterable events. Findings from Nicaraguan homesigners and NSL cohort 1 signers indicate similar patterns—repetitions do not distinguish noun from verb targets, but do (not surprisingly) signal iterability. In contrast, ASL signers and NSL signers who entered the signing community later not only use more repetition overall for iterable items, but also use

repetition to distinguish noun and verb targets. Together, these findings suggest that the grammatical use of repetitions to distinguish word classes may develop over time. [Abner et al. \(2019\)](#) further suggest that using repetitions as a grammatical marker may emerge from the iconic use of repetitions. Some NSL signs for objects, which were associated with iterable actions, were repeated; as a result, repetition became associated with, and a marker for, nouns. In comparison, our finding from experiment 1 in which participants produce more repetitions for typical (verb) than atypical (noun) targets runs counter to this pattern, though the pattern we find is also attested in some sign languages ([Kubus 2008](#); [Schreurs 2006](#)). This finding suggests that the grammaticalisation of repetitions into word class markers, while possibly grounded in the iconic relation to iterability, may be flexible in how it is applied to distinguish noun and verb forms. Certainly, across both experiments 1 and 2, repetitions strongly (and iconically) distinguish iterable from non-iterable events.

We do not find that participants make any distinctions based on the two remaining form properties we analysed—the size of target action gestures, or the location of target gestures. In both cases, iconic representation of events would predict that distinctions could emerge based on either property. For example, [Kimmelman \(2009\)](#) suggests that verb forms may be derived from embodied enactments of events, which may rely on larger, iconic movements than on more economic, reduced forms. Similarly, locations inherent to an event may be preserved in a verb or action sign (such as holding a camera to the face to take a photo) but produced in a neutral space for an object sign (as the location is not intrinsically linked to the object alone). That we do not find distinctions based on these parameters is not surprising for a number of reasons. Firstly, though common strategies such as size, location and repetitions are used across sign languages to distinguish noun and verb forms, and have been hypothesised to have their bases in shared, iconic representations, not all languages mark grammatical categories across all parameters. Indeed, the use and perception of some distinctions such as the size of the signing space can vary depending on the signer's cultural or linguistic experience ([Emmorey and Pyers 2017](#); [McCaskill et al. 2011](#); [Mirus et al. 2001](#)). In addition, some representations may be more flexible in earlier stages of language emergence, as our experiment aims to model. For example, although natural word order preferences are widely documented in silent gestures ([Goldin-Meadow et al. 2008](#); [Hall et al. 2013](#); [Schouwstra and de Swart 2014](#)), and word order preferences appear early in emerging sign languages ([Napoli and Sutton-Spence 2014](#); [Sandler et al. 2005](#)), other properties may arise later through interaction with communities and transmission to new learners forming a linguistic community. In particular, we would expect spontaneous gestures, on the whole, to use a larger gesture space than conventionalised sign systems ([Flaherty et al. 2020](#); [Namboodiripad et al. 2016](#)), which may obfuscate more fine-grained gesture size distinctions used across scene types. That is, size distinctions may first require a reduction in the gesture/signing space to be discernible. Indeed, in experiment 2, we find that participants show a strong preference to produce larger path gestures, regardless of context—there is little variability here with which a distinction based on context could emerge.

Across both experiments 1 and 2, we find that, although the distinctions participants produce may be grounded in iconic representations of events, participants who interact with each other converge on a shared system, producing gestures more similar to each other than would be expected if similarity was based on iconicity alone. In particular, interacting participants produce similar forms in both experiments 1 and 2 despite our two different operationalisations of communication, suggesting that the act of producing a communicative signal that is then interpreted by a partner is sufficient for conventionalised systems to emerge, regardless of the behaviours available in face-to-face interaction that might otherwise shape or facilitate the emerging communicative system ([Healey et al. 2007](#); [Roberts and Levinson 2017](#)). However, the distinctions between typical and atypical targets that emerge across participants do so at the earliest stage of improvisation. These distinctions map most closely onto the findings reported by [Abner et al. \(2019\)](#) for Nicaraguan home-signers, who produce distinctions between noun- and verb- targets that are still highly

variable across individuals, except for the strong preference (also found here) to place verb-like productions at the end of a sequence. Furthermore, our findings indicate that communication in itself is *not* sufficient for the further systematisation of these distinctions that we see in ASL and later cohorts of Nicaraguan Sign Language—communication in our case did *not* lead to substantial additional development of the gestures produced to signal typical vs. atypical targets. Consistent with these findings, previous work suggests that *both* using communicative signals in interaction and learning those signals by naive users of the system shape the emergence of categorical structure (Motamedi et al. 2019; Nölle et al. 2018; Raviv et al. 2019; Silvey et al. 2019). Moreover, it is the repetition of these processes over time that leads to the cultural evolution of systematic distinctions (Kirby et al. 2014; Mesoudi and Thornton 2018; Tamariz and Kirby 2016). Although communicative systems at all stages distinguish between noun-like and verb-like targets, manual communication systems evolve noun and verb *categories* marked by multiple features (see Goldin-Meadow et al. 1994, for evidence of noun–verb categories in a child homesigner in the United States). As such, future work is needed to test how preferences to distinguish noun and verb forms evolve through repeated interaction and iterated learning.

Finally, in experiments 1 and 2, we contrasted two experimental approaches to modelling communicative behaviour. In experiment 1, we operationalised interaction using a reduced director–matcher paradigm in which interacting participants took set turns to produce and interpret gestures (they selected one meaning from a restricted set of four possible interpretations), and all participants received feedback about whether their interpretation was successful. In experiment 2, the operationalisation of interaction was less restrictive, with participants free to negotiate turn-taking and repair strategies, and no limit was put on the meanings they could consider. Although there were small differences in the systems that participants produced (for example, participants in experiment 1 produced more repetitions for typical actions), our results from the two experiments closely align with each other, highlighting the robustness of the improvisation paradigm.

A final, important point is that we find similarities between the noun–verb distinctions created by participants in both experimental paradigms and the noun–verb distinctions found in the naturally emerging language studied in Nicaragua (Abner et al. 2019). For example, early distinctions were based on the order of gestures in a sequence and the use of base hand gestures to mark typical (verb-eliciting) contexts. Experimental models can rarely provide a perfect analogue of language emergence in the real world (Kocab et al. 2018), not least because the participants all know a language. Moreover, the experimental paradigm contains time- and task-related constraints that do not directly replicate language use in the real world. However, our experiments exemplify how such methods can be used alongside data from natural languages to test specific predictions about the processes and mechanisms that drive language evolution. A growing body of work uses these paradigms, informed by the available data from emerging sign languages, to explore key questions about how languages emerge (Hwang et al. 2016; Meir et al. 2014; Motamedi et al. 2019, 2021; Özyürek et al. 2015).

5. Conclusions

We investigated how participants distinguish between typical (verb-like) and atypical (noun-like) targets in novel manual communication systems across two experiments that examined the effect of communication on the emergence of the noun–verb distinction. We found that, across both experiments, clear distinctions emerged in the earliest improvisation stages. All of the participants placed gestures serving a verb role at the end of their utterances, and placed gestures serving a noun role earlier in the utterance. Participants also were biased to produce a base hand on gestures serving a verb function. The strategies used to distinguish between typical and atypical targets emerged early during improvisation, suggesting that the distinction between nouns and verbs is a basic feature of how we communicate, becoming conventionalised in languages over time. Although interacting participants converged on a shared communication system, we did not see further changes,

indicating that other processes (such as the transmission of the system to new learners) are involved in the conventionalisation of noun–verb distinctions. We suggest that using experimental methods to test these hypotheses alongside data from natural languages can help to build a robust picture of how systematic grammatical distinctions emerge.

Supplementary Materials: Files including the annotations made from video data (used for analysis) and all analysis scripts can be found at: <https://osf.io/qzgjt> (accessed on 21 March 2022). Video data from experiment 1 is available at: <https://datashare.ed.ac.uk/handle/10283/3195> (accessed on 21 March 2022).

Author Contributions: Conceptualization, Y.M., K.M., N.A., M.F., S.K. and S.G.-M.; Formal analysis, Y.M., K.M., N.A., M.F., S.K. and S.G.-M.; Funding acquisition, S.K. and S.G.-M.; Investigation, Y.M. and K.M.; Methodology, Y.M., K.M., S.K. and S.G.-M.; Validation, K.M., N.A. and M.F.; Visualization, Y.M.; Writing—original draft, Y.M.; Writing—review & editing, Y.M., K.M., N.A., M.F., S.K. and S.G.-M. All authors have read and agreed to the published version of the manuscript.

Funding: Funding for experiment 1 was awarded to YM as a Carnegie Caledonian Doctoral Scholarship from the Carnegie Trust for the Universities of Scotland (award number PHD060261). Funding for experiment 2 was provided by a grant from NSF BCS-1654154 to SGM, SK is a co-PI.

Institutional Review Board Statement: Ethical approval for study 1 was granted by the ethical review board in the Department of Linguistics and English Language at the University of Edinburgh. Ethical approval for study 2 was granted by the Institutional Review Board at the University of Chicago (IRB Study 97-074).

Informed Consent Statement: Informed consent was obtained from all participants prior to the start of the study.

Data Availability Statement: Video data from experiment 1, data files used for analysis and analysis scripts are publicly available (see Supplementary Materials).

Acknowledgments: We would also like to thank Kenny Smith, Marieke Schouwstra and Jennifer Culbertson for their contributions to the design of Experiment 1, and Destiny Gale, Nicola Lustig, Gayathri Rao and Gowri Rao, who coded the data for both experiments.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

- ¹ While accounts exist suggesting that some languages do not have clear noun–verb categories (Kaufman 2009; Kinkade 1983), these accounts have proved controversial with others providing analysis that shows while categories may not be overtly marked, nominals and predicates are distinguished at some level (Baker 2003; Koch and Matthewson 2009).
- ² The full list of items included can be found at OSF page. <https://osf.io/qzgjt> (accessed on 21 March 2022).
- ³ A list of target actions can be found at <https://osf.io/qzgjt> (accessed on 21 March 2022).
- ⁴ Plots throughout the manuscript were generated using the Python libraries Matplotlib and Seaborn (Hunter 2007; Waskom 2021).
- ⁵ A full description of vignettes can be found at <https://osf.io/qzgjt> (accessed on 21 March 2022).

References

- Abner, Natasha. 2017. What You See Is What You Get: Surface Transparency and Ambiguity of Nominalizing Reduplication in American Sign Language. *Syntax* 20: 317–52. [CrossRef]
- Abner, Natasha. 2021. Determiner phrases—Theoretical perspectives. In *Routledge Handbook of Theoretical and Experimental Sign Language Research*. Edited by Josep Quer, Roland Pfau and Annika Herrman. London: Routledge, pp. 213–31. [CrossRef]
- Abner, Natasha, Molly Flaherty, Katelyn Stangl, Marie Coppola, Diane Brentari, and Susan Goldin-Meadow. 2019. The noun–verb distinction in established and emergent sign systems. *Language* 95: 230–67. [CrossRef]
- Aronoff, Mark, Irit Meir, and Wendy Sandler. 2005. The Paradox of Sign Language Morphology. *Language* 81: 301–44. [CrossRef] [PubMed]
- Baker, Mark C. 2003. *Lexical Categories: Verbs, Nouns and Adjectives*. Cambridge: Cambridge University Press. [CrossRef]
- Bates, Douglas, Ben Bolker, Martin Machler, and Steven C. Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67: 1–48. [CrossRef]
- Beckner, Clay, Janet B. Pierrehumbert, and Jennifer Hay. 2017. The emergence of linguistic structure in an online iterated learning task. *Journal of Language Evolution* 2: 160–76. [CrossRef]

- Bickerton, Derek. 1990. *Language and Species*. Chicago: University of Chicago Press.
- Carrigan, Emily M., and Marie Coppola. 2017. Successful communication does not drive language development: Evidence from adult homesign. *Cognition* 158: 10–27. [CrossRef]
- Christensen, Peer, Riccardo Fusaroli, and Kristian Tylén. 2016. Environmental constraints shaping constituent order in emerging communication systems: Structural iconicity, interactive alignment and conventionalization. *Cognition* 146: 67–80. [CrossRef]
- Cohen, Jacob. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20: 37–46. [CrossRef]
- Coppola, Marie, and Elissa L. Newport. 2005. Grammatical subjects in homesign: Abstract linguistic structure in adult primary gesture systems without linguistic input. *Proceedings of the National Academy of Sciences* 102: 19259–53. [CrossRef]
- Dingemanse, Mark, Sean G. Roberts, Julia Baranova, Joe Blythe, Paul Drew, Simeon Floyd, Rosa S. Gisladdottir, Kobin H. Kendrick, Stephen C. Levinson, Elizabeth Manrique, and et al. 2015. Universal Principles in the Repair of Communication Problems. *PLoS ONE* 10: e0136100. [CrossRef]
- Emmorey, Karen, and Jennie Pyers. 2017. Cognitive biases in construing iconic mappings. Paper presented at the 11th International Symposium on Iconicity in Language and Literature, Brighton, UK, April 7; Available online: <https://osf.io/awyg2> (accessed on 21 March 2022).
- Fay, Nicolas, Michael Arbib, and Simon Garrod. 2013. How to bootstrap a human communication system. *Cognitive Science* 37: 1356–67. [CrossRef] [PubMed]
- Fay, Nicolas, Simon Garrod, Leo Roberts, and Nik Swoboda. 2010. The interactive evolution of human communication systems. *Cognitive Science* 34: 351–86. [CrossRef] [PubMed]
- Flaherty, Molly, Asha Sato, and Simon Kirby. 2020. An Emergent Language Becomes Smaller as It Evolves: New Evidence From Motion Tracking in Nicaraguan Sign Language. In *The Evolution of Language: Proceedings of the 13th International Conference (EvoLang13)*. Edited by Andrea C. Ravignani, Chiara Barbieri, Mauricio Martins, Molly Flaherty, Ella Lattenkamp, Hannah Little, Katie Mudd and Tessa Verhoef. Nijmegen: The Evolution of Language Conferences. [CrossRef]
- Garrod, Simon, and Martin J. Pickering. 2009. Joint Action, Interactive Alignment, and Dialog. *Topics in Cognitive Science* 1: 292–304. [CrossRef] [PubMed]
- Garrod, Simon, Nicolas Fay, John Lee, Jon Oberlander, and Tracy MacLeod. 2007. Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science* 31: 961–87. [CrossRef]
- Gershkoff-Stowe, Lisa, and Susan Goldin-Meadow. 2002. Is there a natural order for expressing semantic relations? *Cognitive Psychology* 45: 375–412. [CrossRef]
- Gibson, Edward, Steven T. Piantadosi, Kimberly Brink, Leon Bergen, Eunice Lim, and Rebecca Saxe. 2013. A Noisy-Channel Account of Crosslinguistic Word-Order Variation. *Psychological Science* 24: 1079–88. [CrossRef]
- Goldin-Meadow, Susan, Wing Chee So, Asli Özyürek, and Carolyn Mylander. 2008. The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences of the United States of America* 105: 9163–68. [CrossRef]
- Goldin-Meadow, Susan. 2003. *The Resilience of Language*. London: Taylor and Francis.
- Goldin-Meadow, Susan, and Heidi Feldman. 1977. The development of language-like communication without a language model. *Science* 197: 401–3. [CrossRef]
- Goldin-Meadow, Susan, Cynthia Butcher, Carolyn Mylander, and Mark Dodge. 1994. Nouns and Verbs in a Self-Styled Gesture System: What's in a name? *Cognitive Psychology* 27: 259–319. [CrossRef]
- Goldin-Meadow, Susan, Diane Brentari, Marie Coppola, Laura Horton, and Ann Senghas. 2014. Watching language grow in the manual modality: Nominals, predicates, and handshapes. *Cognition* 136: 381–95. [CrossRef]
- Hall, Matthew L., Rachel I. Mayberry, and Victor S. Ferreira. 2013. Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition* 129: 1–17. [CrossRef] [PubMed]
- Healey, Patrick G. T., Nik Swoboda, Ichiro Umata, and James King. 2007. Graphical language games: Interactional constraints on representational form. *Cognitive Science* 31: 285–309. [CrossRef] [PubMed]
- Heine, Bernd, and Tania Kuteva. 2007. *The Genesis of Grammar: A Reconstruction*. Oxford: Oxford University Press.
- Hockett, Charles F. 1977. *The View from Language: Selected Essays, 1948–1974*. Athens: The University of Georgia Press.
- Hopper, Paul J., and Sandra Thompson. 1985. The iconicity of the universal categories 'noun' and 'verb'. In *Iconicity in Syntax*. Edited by John Haiman. Amsterdam: John Benjamins.
- Hunger, Barbara. 2006. Noun/Verb Pairs in Austrian Sign Language (ÖGS). Available online: <https://doi.org/info:doi/10.1075/sll.9.1.06hun> (accessed on 21 March 2022).
- Hunter, John D. 2007. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* 9: 90–95.
- Hurford, James R. 2007. *The Origins of Meaning: Language in the Light of Evolution*. Oxford: Oxford University Press.
- Hwang, So-One, Nozomi Tomita, Hope Morgan, Rabia Ergin, Deniz Ilkbasaran, Sharon Seegers, Ryan Lepic, and Carol Padden. 2016. Of the body and the hands: Patterned iconicity for semantic categories. *Language and Cognition* 9: 573–602. [CrossRef]
- Jackendoff, Ray. 2002. *Foundations of Language*. Oxford: Oxford University Press.
- Johnston, Trevor. 2001. Nouns and verbs in Australian Sign Language: An open and shut case? *Journal of Deaf Studies and Deaf Education* 6: 235–57. [CrossRef]
- Kaufman, Daniel. 2009. Austronesian Nominalism and Its Consequences: A Tagalog Case Study. *Theoretical Linguistics* 35: 1–49. [CrossRef]

- Kimmelman, Vadim. 2009. Parts of speech in Russian Sign Language: The role of iconicity and economy. *Sign Language and Linguistics* 12: 161–86. [CrossRef]
- Kinkade, M. Dale. 1983. Salish evidence against the universality of ‘noun’ and ‘verb’. *Lingua* 60: 25–39. [CrossRef]
- Kirby, Simon. 2016. VideoBox: Video Recording, Streaming and Mirroring for Experiments. Available online: <http://edin.ac/2haREUz> (accessed on 21 March 2022).
- Kirby, Simon, Monica Tamariz, Hannah Cornish, and Kenny Smith. 2015. Compression and Communication in the Cultural Evolution of Linguistic Structure linguistic structure. *Cognition* 141: 87–102. [CrossRef]
- Kirby, Simon, Thomas L. Griffiths, and Kenny Smith. 2014. Iterated learning and the evolution of language. *Current Opinion in Neurobiology* 28: 108–14. [CrossRef]
- Kocab, Annemarie, Hannah Lam, and Jesse Snedeker. 2018. When Cars Hit Trucks and Girls Hug Boys: The Effect of Animacy on Word Order in Gestural Language Creation. *Cognitive Science* 42: 918–38. [CrossRef]
- Koch, Karsten, and Lisa Matthewson. 2009. The Lexical Category Debate in Salish and Its Relevance for Tagalog. *Theoretical Linguistics* 35: 125–37. [CrossRef]
- Kubus, Okan. 2008. An analysis of Turkish Sign Language (TİD) Phonology and Morphology. Master’s thesis, Middle East Technical University, Ankara, Turkey.
- McCaskill, Carolyn, Ceil Lucas, Robert Bayley, and Joseph Hill. 2011. *The Hidden Treasure of Black ASL: Its History and Structure*. Washington, DC: Gallaudet University Press.
- Meir, Irit, Mark Aronoff, Carl Börstell, So-One Hwang, Deniz Ilkbasaran, Itamar Kastner, Adi Lifshitz, Ben Basat, Carol Padden, and Wendy Sandler. 2014. The effect of being human and the basis of grammatical word order: Insights from novel communication systems and young sign languages. *Cognition* 158: 1–40. [CrossRef]
- Mesoudi, Alex, and Alex Thornton. 2018. What Is Cumulative Cultural Evolution? *Proceedings of the Royal Society B: Biological Sciences* 13: 285. [CrossRef]
- Mirus, Gene, Christian Rathmann, and Richard P. Meier. 2001. Proximalization and distalization of sign movement in adult learners. In *Signed Languages: Discoveries from International Research*. Edited by Valerie M. Dively, Melanie Metzger, Anne Marie Baer and Sarah Taub. Washington, DC: Gallaudet University Press, pp. 103–19.
- Morford, Jill P., and Susan Goldin-Meadow. 1997. From here and now to there and then: The development of displaced reference in homesign and English. *Child Development* 68: 420–35. [CrossRef]
- Motamedi, Yasamin, Kenny Smith, Marieke Schouwstra, Jennifer Culbertson, and Simon Kirby. 2021. The emergence of systematic argument distinctions in artificial sign languages. *Journal of Language Evolution* 6: 77–98. [CrossRef]
- Motamedi, Yasamin, Marieke Schouwstra, Kenny Smith, Jennifer Culbertson, and Simon Kirby. 2019. Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition* 192: 103964. [CrossRef]
- Namboodiripad, Savithry, Daniel Lenzen, Ryan Lepic, and Tessa Verhoef. 2016. Measuring Conventionalization in the Manual Modality. *Journal of Language Evolution* 1: 109–18. [CrossRef]
- Napoli, Donna Jo, and Rachel Sutton-Spence. 2014. Order of the major constituents in sign languages: Implications for all language. *Frontiers in Psychology* 5: 376. [CrossRef]
- Nölle, Jonas, Marlene Staib, Riccardo Fusaroli, and Kristian Tylén. 2018. The emergence of systematicity: How environmental and communicative factors shape a novel communication system. *Cognition* 181: 93–104. [CrossRef]
- Özçalışkan, Şeyda, Ché Lucero, and Susan Goldin-Meadow. 2016. Does language shape silent gesture? *Cognition* 148: 10–18. [CrossRef]
- Özyürek, Asli, Reyhan Furman, and Susan Goldin-Meadow. 2015. On the way to language: Event segmentation in homesign and gesture. *Journal of Child Language* 42: 64–94. [CrossRef]
- Peirce, Jonathan W. 2007. PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods* 162: 8–13. [CrossRef]
- Perlman, Marcus, Rick Dale, and Gary Lupyan. 2015. Iconicity can ground the creation of vocal symbols. *Royal Society Open Science* 2: 150152. [CrossRef]
- Pizzuto, Elena, and Serena Corazza. 1996. Noun morphology in Italian Sign Language (LIS). *Lingua* 98: 169–96. [CrossRef]
- R Core Team. 2013. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, Available online: <http://www.R-project.org/> (accessed on 21 March 2022).
- Raviv, Limor, Antje Meyer, and Shiri Lev-Ari. 2019. Compositional structure can emerge without generational transmission. *Cognition* 182: 151–64. [CrossRef]
- Roberts, Sean G., and Stephen C. Levinson. 2017. Conversation, cognition and cultural evolution: A model of the cultural evolution of word order through pressures imposed from turn taking in conversation. *Interaction Studies* 18: 402–42. [CrossRef]
- Sandler, Wendy, Irit Meir, Carol Padden, and Mark Aronoff. 2005. The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences of the United States of America* 102: 2661–65. [CrossRef]
- Schouwstra, Marieke, and Henrietta de Swart. 2014. The semantic origins of word order. *Cognition* 131: 431–36. [CrossRef]
- Schouwstra, Marieke, Kenny Smith, and Simon Kirby. 2016. From natural order to convention in silent gesture. *The Evolution of Language: Proceedings of the 11th International Conference (EVLANG11)*. Available online: <http://evolang.org/neworleans/papers/67.html> (accessed on 21 March 2022).
- Schreurs, Linda. 2006. The Distinction between Formally and Semantically Related Noun-Verb Pairs in Sign Language of the Netherlands (NGT). Master’s thesis, University of Amsterdam, Amsterdam, The Netherlands.

- Silvey, Catriona, Simon Kirby, and Kenny Smith. 2019. Communication increases category structure and alignment only when combined with cultural transmission. *Journal of Memory and Language* 109: 104051. [\[CrossRef\]](#)
- Singleton, Jenny L., Susan Goldin-Meadow, and David McNeill. 1995. The Cataclysmic Break Between Gesticulation and Sign: Evidence Against a Unified Continuum of Gestural Communication. In *Language, Gesture and Space*. Edited by Karen Emmorey and Judy S. Reilly. Mahwah: Lawrence Erlbaum Associates Publishers, pp. 287–311.
- Sloetjes, Han, and Peter Wittenburg. 2008. Annotation by category—ELAN and ISO DCR. Paper presented at the 6th International Conference on Language Resources and Evaluation, Marrakech, Morocco, May 28–30.
- Stivers, Tania, Nick J. Enfield, Penelope Brown, Christina Englert, Makoto Hayashi, Trine Heinemann, Gertie Hoymann, Federico Rossano, Jan Pieter de Ruiter, Kyung-Eun Yoon, and et al. 2009. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America* 106: 10587–92. [\[CrossRef\]](#)
- Sulik, Justin. 2018. Cognitive mechanisms for inferring the meaning of novel signals during symbolisation. *PLoS ONE* 13: e0189540. [\[CrossRef\]](#)
- Supalla, Ted, and Elissa L. Newport. 1978. How many seats in a chair? The derivation of nouns and verbs in American Sign Language. In *Understanding Language through Sign Language Research*. Edited by Patricia Siple. New York: Academic Press, pp. 91–131.
- Tamariz, Monica, and Simon Kirby. 2016. The cultural evolution of language. *Current Opinion in Psychology* 8: 37–43. [\[CrossRef\]](#)
- Theisen, Carrie A., Jon Oberlander, and Simon Kirby. 2010. Systematicity and arbitrariness in novel communication systems. *Interaction Studies* 11: 14–32. [\[CrossRef\]](#)
- Tkachman, Oksana, and Wendy Sandler. 2013. The noun–verb distinction in two young sign languages. *Gesture* 13: 253–86. [\[CrossRef\]](#)
- Verhoef, Tessa, Simon Kirby, and Bart de Boer. 2014. Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics* 43: 57–68. [\[CrossRef\]](#)
- Waskom, Michael L. 2021. Seaborn: Statistical data visualization. *Journal of Open Source Software* 6: 3021. [\[CrossRef\]](#)
- Wilbur, Ronnie B. 2003. Representations of telicity in ASL. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, vol. 39.
- Wilbur, Ronnie B. 2008. Complex predicates involving events, time and aspect: Is this why sign languages look so similar? In *Signs of the Time: Selected Papers from TISLR 2004*. Edited by Josep Quer. Seedorf: Signum Press.
- Wilcox, Sherman. 2004. Cognitive iconicity: Conceptual spaces, meaning, and gesture in signed language. *Cognitive Linguistics* 15: 119–47. [\[CrossRef\]](#)