

Types of iconicity and combinatorial strategies distinguish semantic categories in silent gesture across cultures

Ortega, Gerardo; Özyürek, Asli

DOI:

[10.1017/langcog.2019.28](https://doi.org/10.1017/langcog.2019.28)

License:

Other (please specify with Rights Statement)

Document Version

Peer reviewed version

Citation for published version (Harvard):

Ortega, G & Özyürek, A 2019, 'Types of iconicity and combinatorial strategies distinguish semantic categories in silent gesture across cultures', *Language and Cognition*. <https://doi.org/10.1017/langcog.2019.28>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

Checked for eligibility: 01/10/2019

This article has been published in a revised form in *Language and Cognition* [<https://doi.org/10.1017/langcog.2019.28>]. This version is free to view and download for private research and study only. Not for re-distribution or re-use. © UK Cognitive Linguistics Association 2019.

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

**Types of iconicity and combinatorial strategies
distinguish semantic categories in silent gesture
across cultures**

GERARDO ORTEGA
University of Birmingham
and

ASLI ÖZYÜREK
Radboud University and Max Planck Institute for Psycholinguistics

Abstract

In this study we explore whether different types of iconic gestures (i.e., *acting*, *drawing*, *representing*) and their combinations are used systematically to distinguish between different semantic categories in production and comprehension. In Study 1, we elicited silent gestures from Mexican and Dutch participants to represent concepts from three semantic categories: actions, manipulable objects, and non-manipulable objects. Both groups favoured the *acting* strategy to represent actions and manipulable objects; while non-manipulable objects were represented through the *drawing* strategy. Actions elicited primarily single gestures whereas objects elicited combinations of different types of iconic gestures as well as pointing. In Study 2, a different group of participants were shown gestures from Study 1 and were asked to guess their meaning. Single-gesture depictions for actions were more accurately guessed than for objects. Objects represented through two-gesture combinations (e.g., *acting* + *drawing*) were more accurately guessed than objects represented with a single gesture. We suggest iconicity is exploited to make direct links with a referent, but when it lends itself to ambiguity, individuals resort to combinatorial structures to clarify the intended referent. Iconicity and the need to communicate a clear signal shape the structure of silent gestures and this in turn supports comprehension.

keywords: silent gesture, iconicity, language emergence, combinatorial structure, emerging sign language.

Introduction

Iconicity, understood as the motivated mappings between form and meaning, is said to be a core property of language (Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015; Perniss, Thompson, & Vigliocco, 2010), and may be exploited to create novel communication systems in the absence of linguistic conventions (Gibson et al., 2013; Goldin-Meadow, So, Ozyürek, & Mylander, 2008a; Perlman & Lupyan, 2018; Sulik, 2018; Verhoef, Kirby, & de Boer, 2015; Vigliocco, Perniss, & Vinson, 2014). Within the scope of research on language emergence and evolution, psychologists and cognitive scientists have shown that silent gesture is a powerful embodied tool which bootstraps a communicative system in the absence of linguistic means (Fay, Arbib, & Garrod, 2013). While some have argued that iconicity influences the sequencing of events in silent gesture (Christensen, Fusaroli, & Tylén, 2016), research has rarely investigated whether and how different types of iconicity (i.e., modes of representation; Müller, 2013, 2016) and their combinations contribute towards the creation of novel communicative signals. Furthermore, it is well established in the literature that humans adapt the form of a message to the needs of their interlocutor (i.e., audience design; Bell, 1984), and that the particular sequences of silent gesture is a response to produce a clear signal in an otherwise noisy channel (Gibson et al., 2013). Nonetheless, few have attempted to investigate directly if indeed interlocutors benefit from these different types of iconic strategies in silent gesture to accurately interpret their meaning.

In this study we investigate the preferred strategies to express concepts in silent gesture within and across semantic categories (i.e., production); and evaluate the communicative efficiency of these gestural signals (i.e., comprehension). In Study 1 (production), we explored whether and how different types of iconicity may be recruited in distinguishing semantic categories in silent gesture. We also investigated the implementation of combinatorial patterning in the manual modality, whereby the representation of a concept may be achieved through the combined meaning of different types of iconic gestures. In Study 2 (comprehension), we tested whether the gestures observed in Study 1 facilitate interpretation of the

referent by a different group of viewers. We argue that the human capacity to generate different types of iconic gestures, primarily through the representation of bodily actions, and the communicative pressure to reduce ambiguity for an interlocutor, shape the form of manual symbols, which in turn support comprehension. Critically, these biases should be observed regardless of gesturers' linguistic background (Goldin-Meadow et al., 2008a). In order to address this question, these tasks were performed by participants from two different countries, the Netherlands and Mexico, to investigate whether they show similar patterns in the production and interpretation of iconic depictions. The expected findings would suggest that silent gestures are exploited as building blocks from which sign languages begin to emerge.

Types of iconicity as strategy to differentiate semantic distinctions in the manual-visual modality

It has long been argued that iconic form-meaning mappings lie at the centre of the origins of language, ontogenetically and phylogenetically (Imai & Kita, 2014; Perniss & Vigliocco, 2014; Ramachandran & Hubbard, 2001). Only recently, however, that empirical research has begun to investigate the role of different types of iconicity in language learning and emergence. There is general consensus in the field of gesture studies that there are four types of iconic representations that can be exploited to depict a concept (Müller, 2013, 2016). *Acting* denotes how an object is manipulated; *representing* uses the hand to recreate the form of an object; *drawing* describes the outline of a referent; and *moulding*¹ depicts the three-dimensional characteristics of an object (Figure 1). Interestingly, linguists have also documented remarkably similar iconic strategies in conventionalised sign languages albeit with different labels (i.e., handling, instrument, tracing, size and shape specifiers) (Klima & Bellugi, 1979; Mandel, 1977; Nyst, 2016; Padden et al., 2013).

¹ We did not find instances of the *moulding* category in our data so in the remainder of the manuscript we refer only to the *drawing* category.

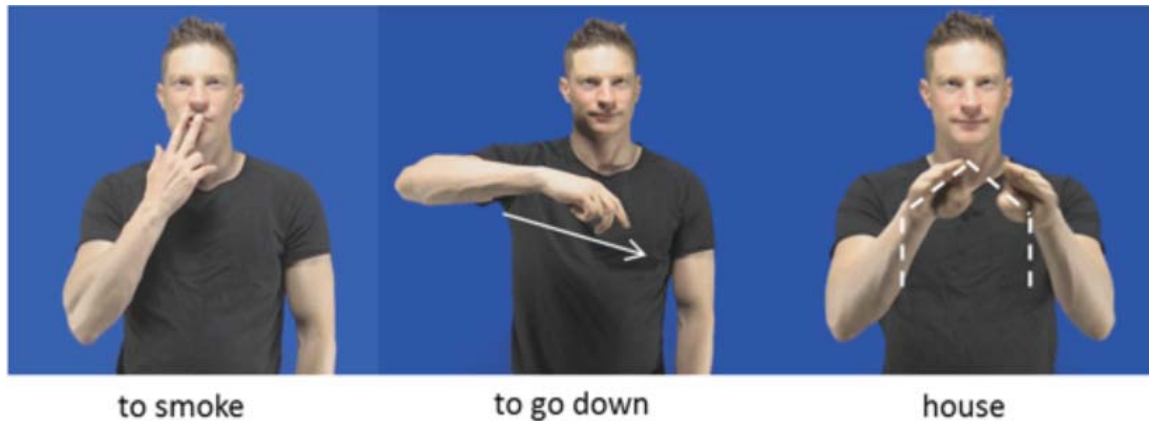


Figure 1. Different types of iconic representations in gesture. (A) *Acting*: shows a person pretending to hold a cigarette with two fingers near the mouth for the action of smoking. (B) *Representing*: the gesturer depicts the concept of descending by representing two legs with index and middle fingers in wiggling motion. (C) *Drawing*: a ‘house’ is represented with the hands tracing the outline of the building.

A growing body of evidence shows that the form of different types of gestures is not as idiosyncratic and heterogeneous as previously thought but rather exhibit consistent patterns motivated by the body as main articulator and the form of the referent itself (Chu & Kita, 2016). Regarding silent gesture, Van Nispen et al. (2017) found that there was a strong preference for the *acting* strategy when people were asked to express objects with the hands and without speaking. For co-speech gesture, Masson-Carro et al. (2015) found that objects with high manual affordances (i.e., they could be manipulated with the hands) were primarily represented through an *acting* strategy, while objects with low affordances were described using the *drawing* strategy. These patterns show that to a certain extent the human body and the physical attributes of the referent (i.e., manipulability) makes gesturers to align different types of iconic representations with specific referents.

Different types of iconic strategies also seem to be exploited to mark distinctions between actions and objects. Padden, Hwang, Lepic, and Seegers (2015) found that when gesturers are asked to express actions (elicited with video vignettes) and hand-held tools (elicited with pictures), they tend to favour the *acting* strategy around 90% of the times. However, the authors report a slight but significant trend to

implement the *acting* strategy for actions and *representing* for tools. The notion of *patterned iconicity* posits that these subtle differences are at the core of noun-verb marking in emerging and established sign languages with the *acting* strategy used to express verbs and the *representing* strategy to express nouns (Brentari, Renzo, Keane, & Volterra, 2015; Padden et al., 2015, 2013). Typological investigations of sign languages lend further support to this claim by showing that a large number of sign languages exhibit clear preferences to associate specific types of iconicity to different semantic category (Kimmelman, Klezovich, & Moroz, 2018).

A shortcoming of these studies, however, is that they have focused exclusively on two categories such as manipulable vs. non-manipulable objects in co-speech gesture (Masson-Carro et al., 2015) or actions vs. tools in silent gestures (Padden et al., 2015, 2013). An interesting but untested question is whether we observe systematic patterns in the iconic strategies used in silent gesture across different semantic categories. Based on previous research, we predict that *acting* will be chief amongst other strategies, and particularly important to represent actions (Padden et al., 2015, 2013) and objects that can be manipulated with the hands (Masson-Carro et al., 2015). Further, previous research has not investigated whether the combinations of different iconic strategies are implemented to distinguish different semantic categories. We predict that the combination of different types of iconic gestures might be an advantageous strategy to communicate concepts more accurately for an interlocutor, and when type of iconicity alone does not succeed in marking subtle differences.

Beyond single iconic representations: combinatorial patterning in iconic strategies in the manual visual modality

Language is said to be an efficient system partly because it makes use of a finite set of words and combines them to coin new expressions instead of creating new labels for every new concept (Hockett, 1960). A novel linguistic label for every new meaning would clutter the signal space and would result in a system with multiple similar items that could be potentially confusing to an interlocutor. The presence

of combinatorial structures solves this issue in that the linguistic system can employ existing elements and reduce the number of items within the signal space (Verhoef et al., 2015). The undeniable advantage of combinatorial structure allows language not only to be easier to compute but also makes it more predictable, more learnable, and easier to transmit (Micklos, 2016; Smith & Kirby, 2008; Verhoef et al., 2015).

Emerging sign languages also exploit the combinatorial potential of individual signs to mark distinctions across semantic categories. In Al-Sayyid Bedouin Sign Language (ABSL), an emerging sign language in Israel, multi-sign combinations are a productive strategy widely used to refer to concepts lacking a conventionalised linguistic label (Meir, Aronoff, Sandler, & Padden, 2010. See Haviland, 2013 for similar claims). For example, for the noun ‘lipstick’, ABSL users produce a sign representing the action of applying lipstick (i.e., *acting*) followed by a sign tracing its shape and size (i.e., *drawing*). The verb consists of a single sign depicting the action of applying lipstick (i.e., *acting*) (Tkachman & Sandler, 2013). Emerging sign languages thus recombine existing signs to generate new meanings and critically, they resort to two-sign combinations to express objects (Haviland, 2013; Tkachman & Sandler, 2013).

The aforementioned studies are unique in that they investigate the presence of combinatorial patterning in deaf users of emerging sign languages. But, where does this combinatorial strategy stem from? Simulations of language emergence in the lab have provided convincing evidence that through social interaction and cultural transmission combinatorial patterns emerge (Galantucci & Garrod, 2011; Micklos, 2016; Verhoef et al., 2015). It remains an empirical question whether people will exploit the combinatorial potential of different types of iconic gestures to mark semantic distinctions during spontaneous production of silent gesture even before processes of social transmission begin. Given the prevalence of multi-sign combinations at the earliest stages of sign language emergence around the world (Haviland, 2013; Tkachman & Sandler, 2013), we predict that this combinatorial mechanism can be found

outside a signed linguistic system, that is, in the silent gestures of speakers of typologically different languages. Importantly, we expect that perceivers of these different strategies will benefit from them and will be able to identify the referent accurately.

To sum up, previous studies in co-speech and silent gesture have focused separately on the iconic strategies employed to represent concepts within-category (i.e., manipulable vs. non-manipulable objects) (Masson-Carro et al., 2015) or across-category representations (i.e., actions vs. objects) (Padden et al., 2015, 2013). An experimental design including all three categories is likely to exert additional pressure to the task because participants will have to make distinctions amongst concepts that vary semantically to

different degrees (e.g., ‘hammering’ vs. ‘lighter’ vs. ‘pyramid’). The question is whether different types of iconic representations align systematically to specific semantic categories and if so, how. It also remains to be investigated whether individuals implement combinatorial patterning systematically in silent gestures, and what semiotic resources are recruited to create them.

Further, it has been argued that people design the structure of silent gesture to reduce ambiguity for an interlocutor (Gibson et al., 2013), but few have tested these claims directly (for an exception see Hall, Ahn, Mayberry, & Ferreira, 2015). By carrying out a comprehension study of silent gestures, it will be possible to understand to what extent the strategies implemented in production are effective for accurate comprehension of different semantic categories for a different group of perceivers.

Finally, we test the specificity of these expected patterns and ask whether they generalise across different populations. Linguistic and cultural conventions are important determinants that shape the form of many gestures including emblems, co-speech gestures (Kendon, 2004; Kita & Özyürek, 2003), descriptions of object sizes (Nyst, 2016), and motion events (Kita & Özyürek, 2003; Özçalışkan, Lucero, & Goldin-Meadow, 2016). However, as stated earlier, individuals from different linguistic backgrounds

converge in the strategies to represent events in silent gesture (Goldin-Meadow, So, Ozyürek, & Mylander, 2008b; Özçalışkan et al., 2016). Looking at the silent gestures across unrelated linguistic groups will be an initial step to establish whether iconic gestural depictions and combinatorial strategies are generalisable skills to communicate concepts effectively in the absence of linguistic conventions.

The Present Study

In this study we investigate the production (Study 1) and comprehension (Study 2) of iconic silent gestures by Mexican and Dutch adults with no knowledge of any sign language. We ask: 1) Do individuals use different types of iconic representations for different semantic categories?; 2) Are additional combinatorial strategies implemented to mark distinctions across semantic categories?; 3) How effective are the different strategies to express the meaning of the intended referent to a viewer?; and 4) Are there similarities in production and perception of silent gestures across different cultures?

In Study 1 (production), we elicited silent gestures for a list of words, and then described the types of iconicity used for each semantic category, as well as any additional strategies participants implemented to differentiate concepts across semantic category (i.e., combinatorial patterning). First, we expected to see different types of iconic depictions aligning with specific semantic categories with *acting* being a prominent strategy (Van Nispen et al., 2017). Secondly, we hypothesised that participants will feel the communicative pressure to express a differentiated semantic category and as a result will generate combinations of different types of gestures to reduce ambiguity in their signal.

The gestures from Study 1 served as stimulus materials for Study 2 (comprehension), in which we asked a different group of participants in each culture to guess their meaning. The *acting* strategy has a direct correspondence with actions and thus we expect that this mapping will be the most accurately guessed for the representation of actions. We also expect that the combination of multiple gestures and

types of iconicity will aid participants in interpreting the meaning of some concepts, in particular objects, as has been observed in emerging sign languages (Haviland, 2013; Tkachman & Sandler, 2013).

Finally, we expect to find similarities between both cultures. The Mexican and Dutch groups make an interesting subject of study, not only because they are two unrelated cultural groups but also because the typological features of Spanish and Dutch may lead to significant differences in gestural production. If we find differences in their manual representations for the same concepts it could be argued that gestural depictions are grounded on linguistic representations or social conventions. However, if they converge in their strategies we could argue that their gestures are independent of the language-specific encodings and originate from shared cognitive biases and the tendency to produce an unambiguous gestural signal. This finding will add to the mounting evidence that silent gesture exhibits generalised forms across cultures and thus has limited influence of speech (Goldin-Meadow et al., 2008b; Özçalışkan et al., 2016).

Study 1

Methodology

Participants

Twenty native speakers of Dutch (10 females, age range: 21-46, mean: 27 years) living in the area of Nijmegen, the Netherlands; and 20 native speakers of Mexican Spanish (10 females, age range: 19-25, mean: 20 years) living in Mexico City took part in the study.

Stimuli

The stimuli consisted of two sets of thirty words (one for each cultural group) from three semantic categories: 10 actions with an object (verbs such as ‘to smoke’), 10 manipulable objects (nouns such as ‘telephone’), and 10 non-manipulable objects (nouns such as ‘pyramid’). In order to evaluate if indeed our categorisation of items into manipulable and non-manipulable object was accurate, two different

groups of Mexican and Dutch participants rated on a 7-point Likert scale how likely the objects could be hand-held (1 low and 7 high manipulability). None of the raters took part in the actual experiment. A mixed analysis of variance with ratings as dependent variable and type of object (manipulable vs. non-manipulable objects) and cultural group (Mexican vs. Dutch) as fixed factors revealed a main effect of type of object $F(1,476) = 580.94, p < 0.001, \eta^2 = 0.550$. Manipulable objects (mean: 6.11, SE = 0.106) had significantly higher scores than non-manipulable objects (mean: 2.50, SE = 0.106). That is, manipulable objects were regarded as more likely to be manipulated with the hands than words in the non-manipulable condition. There was no main effect of cultural group $F(1,476) = 0.315, p = 0.575, \eta^2 = 0.001$ or significant interaction $F(1,476) = 0.251, p = 0.617, \eta^2 = 0.001$ which indicates that both Mexican and Dutch participants produced the same manipulability ratings (see Appendix).

Procedure

Participants were tested individually in a quiet room with one or two cameras recording their gestures. They were instructed to generate a gesture that conveyed exactly the same meaning as the word displayed on the screen. They were explicitly told that they were not allowed to speak or say the target word; and they could not point at any object present in the room. They were only allowed to say the word 'pass' if they could not come up with a spontaneous gesture. They were also told that their videos were going to be shown to another group of participants who would have to guess the meaning of their gesture.

Words were presented in black font on a white background in a different randomised list for each participant. Each trial started with a fixation cross in the middle of the screen for 500 ms. This was followed by a word presented for 4000 ms, time in which participants had to come up with their gestural depictions. The next trial started as soon as the 4000 ms had run out. We limited the time for gesture production so as to force participants to produce their most intuitive responses. We avoided the use of pictures to reduce the possibility of prompting participants with visual cues in their gestural productions

and to encourage them to express a generic form of the target concept. Participants' renditions were video recorded and later annotated using the software ELAN (Sloetjes & Wittenburg, 2018).

Coding and data analysis

Participants produced one gesture or sequences of gestures to depict the target word so we annotated individual meaningful gestural units. Each gesture would consist minimally of a preparation phase, a stroke and a (partial/full) retraction, and would often include a brief hold between gestural units. Once all the gestures were isolated, we classified them according to their type of iconic depiction. Adapting the taxonomy developed by Müller (2013), we categorised each gesture as follows. A gesture was coded as *acting* if the gesture represented how the referent is manipulated or if it depicted a bodily action associated with an object (i.e., the hand represents the hand). If the hands in any possible configuration recreated the form of an object it was coded as *representing* (i.e., hand as object). Finally, the category *drawing* was used if participants used their hands or fingers to describe the outline or the three-dimensional characteristics of an object. We also included the category *deictic* which consisted of pointing, showing and/or ostensive eye-gaze to elements of a gesture. Participants did adhere to the rule of not pointing to any object in the room but they used points and other ostensive cues to highlight elements of their gestures. Deictics are not a mode of representation *per se* but we decided to include this category given their high prevalence during the task.

After the categorisation of the gestures, a second coder independently classified 20% of the data ($n = 240$ descriptions out of 1200) into one of the different categories (i.e., *acting*, *representing*, *drawing*, and *deictic*) to verify for coding consistency. The interrater reliability was found to be strong ($\kappa = 0.895$, $p < 0.0011$, 95% confidence interval [CI] [0.853, 0.938]). For both groups and across semantic categories (actions with an object, manipulable objects, and non-manipulable objects), we calculated: 1) the number of gestures produced per concept per participant; 2) the proportion of different modes of representation

across all single-gesture renditions; 3) the proportion of deictics produced across semantic categories per participant; and 4) the proportion of gesture combinations to convey the meaning of a concept across semantic categories per participant.

Results

Number of gestures across semantic categories

We calculated the mean number of gestures produced per item per participant across the three semantic categories. If participants produced a multi-gesture stretch with two or more instances of the same gesture, we did not include the repetition in the count. In order to yield symmetric distribution of data, we performed arcsine transformations and then performed a mixed analysis of variance with mean number of gestures as dependent variable, and semantic category (actions with objects, manipulable, and non-manipulable objects) and cultural group (Dutch and Mexican) as fixed factors. The analysis revealed a main effect of semantic category $F(2,84) = 24.206, p < 0.001, \eta^2 = 0.366$. Post hoc comparisons at $p < 0.05$ revealed that actions with objects (mean = 1.14 gestures, SE = 0.05) elicited significantly fewer gestures than manipulable (mean = 1.68 gestures, SE = 0.05) and non-manipulable objects (mean = 1.45, SE=0.06). There was no main effect of cultural group $F(1,84) = 0.324, p = 0.571, \eta^2 = 0.004$ and no significant interaction $F(2,84) = 0.989, p = 0.376, \eta^2 = 0.023$ (Figure 2). The analysis shows that the vast majority of actions elicited a single gesture, which often depicted how the action is executed (Figure 2). Manipulable objects were predominantly depicted with more than one gesture. Non-manipulable objects have a split with an almost equal proportion of items being depicted with a single or multiple gestures. This pattern holds across cultures suggesting that both groups produced strongly overlapping mean values for the number of gestures for each category regardless of their languages.

It could be argued that the production of gestural combinations could be attributed to participants being aware that there are action-object pairs semantically related (e.g., ‘to smoke’ vs. ‘lighter’) and as a

result they developed a strategy to disambiguate between both concepts. In order scrutinise this alternative further, we removed these three word pairs from the categories actions and manipulable objects (i.e., ‘to smoke’ vs ‘lighter’; ‘to drink’ vs ‘mug’; ‘to phone’ vs ‘telephone’) and ran a second analysis on the number of gestures to see if differences still hold. An analysis of variance with number of gestures as dependent variable, semantic category as within-subjects variable, and cultural group as between-subjects variable revealed there was a significant main effect of semantic category $F(2,114) = 39.185, p < 0.001, \eta^2 = 0.407$. There was no main effect of group $F(2,114) = 4.530, p = 0.075, \eta^2 = 0.012$; and importantly, the interaction between semantic category and group was not significant $F(2,114) = 0.694, p = 0.502, \eta^2 = 0.012$. Pairwise comparisons after Bonferroni corrections revealed that actions (mean = 1.08, SE = 0.048) generated significantly fewer gestures than manipulable objects (mean = 1.68, SE = 0.48, $p < 0.0001$) and non-manipulable objects (mean = 1.43, SE = 0.48; $p < 0.0001$); the same pattern as the reported before the items were removed.

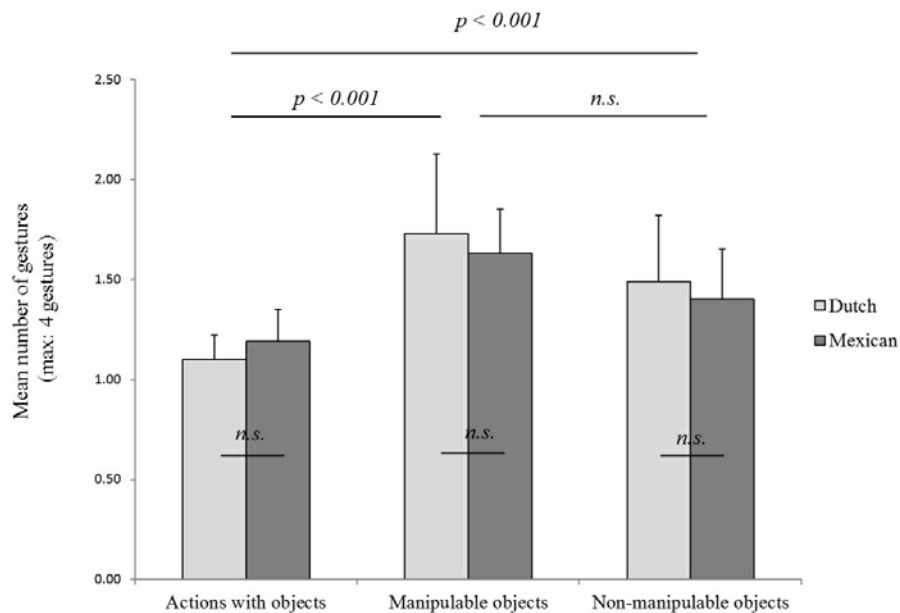


Figure 2. Mean number of gestures per condition across gestural groups (range: 1-4 gestures. Bars represent standard error.

Type of iconicity across semantic category

Here we zoom into single gesture depictions to investigate whether different types of iconic representations (*acting*, *representing*, or *drawing*) were used in specific semantic categories in both cultural groups (Figure 3). This consisted of 59.15% responses for the Mexican group (355 gestural depictions) and 56.66% for the Dutch group (340 gestural depictions). A chi-square on the Mexican dataset shows the *acting* strategy is strongly favoured for actions with objects and manipulable objects $\chi(4)^2 = 85.61$, $p < 0.0001$. There was a balanced distribution of type of depiction in non-manipulable objects. The Dutch group also showed a strong preference for *acting* depictions for actions with objects and manipulable objects, but for non-manipulable objects they favoured the *drawing* mode $\chi(4)^2 = 155.64$, $p < 0.0001$. The *representing* strategy was the least preferred strategy, and the Dutch and the Mexican groups did not differ in the proportion of depictions using it (Table 1).



Figure 3. Examples of the most frequent types of iconicity used for different semantic categories in both cultures. (A, D) The *acting* strategy is used to depict the concept ‘hammering’ by showing how a hammer is manipulated. (B, E) The concept ‘telephone’ was

depicted through the *representing* strategy with the hand adopting the shape of a receiver. (C, F) ‘Pyramid’ is represented through *drawing* because the hands trace its triangular outline.

Table 1. Percentage of gestures according to their type of iconic depiction across conditions in single gesture depictions in both cultural groups. N= instances in which a concept was depicted with a single gesture across participants (200 descriptions per condition in total).

	Mexican		
	Action with object (N=159)	Manipulable Objects (N=88)	Non-manipulable Objects (N=108)
Acting	0.72	0.80	0.33
Representing	0.28	0.18	0.33
Drawing	0.00	0.02	0.34
	1.00	1.00	1.00

	Dutch		
	Actions with object (N=181)	Manipulable Objects (N=70)	Non-manipulable Objects (N=89)
Acting	0.86	0.81	0.20
Representing	0.14	0.18	0.23
Drawing	0.00	0.01	0.57
	1.00	1.00	1.00

Combinatorial patterning in silent gesture

Participants also produced stretches of multiple gestures to convey the meaning of the intended referent, predominantly for objects. We looked at the different combinations of types of iconic depictions for all the stretches of multiple gestures (Table 2). For example, in order to convey the concept ‘pillow’ participants combined a tracing gesture of a square followed by an *acting* gesture of someone lying on a pillow (Figure 4). Depictions with more than two gestures or with combinations with the same type of representation (e.g., *acting* + *acting*) were included in the category ‘other’. The statistical analysis on the Mexican gestures shows that for actions, the favoured combination is *acting* + *representing*; whereas

drawing + *acting* is preferred for manipulable objects and non-manipulable objects $\chi(4)^2 = 82.32, p < 0.0001$. The same preference is observed for the Dutch gestures $\chi(4)^2 = 86.19, p < 0.0001$.



Figure 4. Example of a multi-gesture depiction for the manipulable object ‘pillow’. Participants draw a square shape and then perform an *acting* gesture re-enacting the action of sleeping.

Table 2. Proportion of multi-gesture combinations. N=number of instances in which participants produced more than one gesture (200 descriptions per condition in total). Deictics are not included. The Other category includes sequences of different gestures using the same type of iconicity (e.g., *acting* + *acting*).

	Mexican		
	Actions with object (N=33)	Manipulable Objects (N=102)	Non-manipulable Objects (N=67)
DRAW+ACT	0.22	0.50	0.49
ACT+REP	0.58	0.15	0.15
DRAW+REP	0.00	0.13	0.18
Other	0.20	0.22	0.18

	1.00	1.00	1.00
	Dutch		
	Actions with object (N=18)	Manipulable Object (N=71)	Non-manipulable Object (N=75)
DRAW+ACT	0.18	0.56	0.51
ACT+REP	0.37	0.10	0.09
DRAW+REP	0.00	0.11	0.23
Other	0.45	0.23	0.17
	1.00	1.00	1.00

We also looked at the instances in which participants added an ostensive cue to refer to a specific feature of their previously produced gesture (i.e., pointing, showing, or direct eye-gaze). We looked at all attempts to represent a concept with both single and multiple gestures collapsed because some extensive cues could occur simultaneously with single gestures (e.g., showing or eye-gaze). We observed that out of the whole Mexican dataset, actions with objects (N = 192) and non-manipulable objects (N = 175) elicited the lowest proportion of deictics (0.04 and 0.09, respectively). Manipulable objects (N = 191), however, elicited significantly more (0.36). Dutch participants exhibited the same proportion of deictic productions across conditions: actions with objects (N = 199) and non-manipulable objects (N = 180) elicited a small proportion of deictics (0.04 and 0.09, respectively); but manipulable objects (N = 191) elicited the largest proportion (0.25). A chi-square revealed that there was no significant difference between the Mexican and Dutch groups because both produced significantly more deictics for manipulable objects than the other two categories $\chi(1)^2 = 86.19, p=0.733$. For instance, to represent ‘lighter’ or ‘toothbrush’ participants would perform the action of lighting a cigarette or brushing their teeth and then point or show an imaginary hand-held object (See Figure 5).



Figure 5. Examples of gestures incorporating a deictic gesture. For the concept ‘toothbrush’ participants across cultural groups often produced an *acting* gesture of brushing one’s teeth and then showed or pointed at an imaginary toothbrush (see upper panels). For ‘lighter’ and ‘cigarette’, participants in the Netherlands and Mexico re-enacted lighting a cigarette and then pointing to an imaginary object in their hands.

To sum up, we find supporting evidence that gesturers implement different iconic strategies and their combinations to express concepts in specific semantic categories in silent gesture. Actions with objects tend to be expressed with a single gesture using predominantly the *acting* mode of representation. Both types of objects (manipulable and non-manipulable), in contrast, were generally represented with more than one gesture. In cases when a single gesture was used for objects, manipulable objects elicited the *acting* strategy and non-manipulable objects elicited the *drawing* technique. Participants implemented two additional strategies complementary to the different types of iconic representations. First, participants produced two-gesture sequences -predominantly *acting* + *drawing*- to represent objects, but *acting* + *representing* for actions with objects. Second, manipulable objects elicited significantly more ostensive cues (i.e., pointing, showing) than actions and non-manipulable objects. Critically, these patterns were observed similarly in both the Dutch and the Mexican groups with the only difference being that the Dutch used more *drawing* than other types of depictions for non-manipulable objects.

These data provide strong evidence that the production of silent gestures is shaped by the natural affordances of the referent and the preferences to generate gestures depicting embodied actions. Regardless of cultural and linguistic background, participants showed a strong tendency to produce gestures implementing the *acting* strategy for actions and manipulable objects. However, when the referent did not allow manipulation with the hands, they resorted to an alternative strategy (i.e., *drawing*). Multi-gesture stretches were prevalent primarily in the depiction of objects. It is possible that a single gesture could lead to misinterpretation so participants added information to reduce the risk of ambiguity. For instance, in order to express the concept ‘toothbrush’ participants added a point to the *acting* gesture of brushing one’s teeth because an interlocutor might conclude that the referent is the action and not the intended object.

Humans are known to adapt the form and quality of their utterances to accommodate to their interlocutors’ needs (Bell, 1984; Clark & Murphy, 1982). People design their utterances taking into account the information a recipient may require to decipher a message. However, deciphering the iconic elements in gestures is a complex process (Hassemer & Winter, 2018). Here we argue that participants may be sensitive to the potential ambiguity of their gestural renditions and thus generated additional gestures to minimise the risk of misinterpretations. In order to test this claim, we investigated whether a different group of participants were sensitive to these different gestural strategies and correctly recognised the intended referent. To that end, we carried out a comprehension study in which we showed the silent gestures from the production task to two new groups of participants (one Mexican and one Dutch). We predicted that actions with objects, generally depicted with a single gesture, would be the most accurately guessed because their meaning will be disambiguated by the direct action-to-action mapping. However, both types of objects would be guessed less accurately than actions. We expected participants to be more accurate at interpreting objects that were represented through multi-gesture combinations than when they

were depicted with a single gesture because they will contain more relevant information about the intended referent.

Study 2

Methodology

Participants

Participants for this study were 13 Dutch (3 females, age range: 21-45, mean: 29 years) and 13 Mexican adults (6 females, age range: 25-36, mean: 28 years). None of them took part in the gesture generation task.

Stimuli

Each group of participants took part in an online experiment in which they were shown 30 video clips from the gestures generated in Study 1 (from their own cultural group) and were asked to guess the concept the gesturer intended to represent. Given that gesturers in Study 1 tended to express concepts with a single gesture (e.g., *drawing*) or with combination of gestures (e.g., *acting* + deictic), in the comprehension task we selected stimuli that reflected this preference. We included 10 videos clips of a single gesture where the *acting* strategy was implemented. Manipulable objects included five (single) gestures implementing the *acting* strategy and five gestures consisting of an *acting* + *deictic* combination. Similarly, non-manipulable objects included five (single) gestures implementing the *drawing* strategy and five gestures consisting of an *acting* + *drawing* combination. In order to minimise the risk of participants responding based on who performed the gestures instead of looking at the gesture itself, we ensured that at least one video of all gesturers were included but maximally two times. A summary of the stimulus materials are presented in Table 3.

Table 3. Summary of the stimuli and their type of representation (single vs. multiple gestures) across conditions

	Number of items	Type of representation
Actions with objects (N=10)	10	Acting
Manipulable Objects (N=10)	5 5	Acting Acting + deictic
Non-manipulable objects (N=10)	5 5	Drawing Acting + drawing
TOTAL	30	

Procedure

All videos clips were presented in randomised order on an online platform. Mexican and Dutch participants were instructed that they were going to see 30 videos (one at a time) and that their task was to type in the concept that was being represented by the gesturer. They could watch the video as many times as needed. Participants were encouraged to give a response but if they could not come up with an answer they could write ‘no idea’. Once participants had entered a response they could not go back to a previous trial, and they could not move forward unless they had completed the previous trials.

Coding and analysis

Accuracy in responses was quantified numerically following the Dutch version of the Boston Naming Task² (Roomer, Hoogerwerf, & Linn, 2011). Responses received a score of 3 if they were exactly the same word as the target; a score of 2 if the response was a semantically related to the target and belonged to the same part of speech; and a score of 1 point if responses were semantically related but

² This instrument quantifies accuracy to provide a lexical label for a given visual prompt. It is mainly used to assess linguistic competence during lexical retrieval for patients with aphasia and other language-related impairments.

corresponded to another part of speech. Participants did not get any points if their response was semantically distant from the intended meaning of the gesture.

Results

We calculated the mean value for each item across participants per condition (maximum score = 3). For the Mexican group, actions was the condition most accurately guessed (mean = 2.33; $SD = 0.44$) followed by manipulable objects (mean = 1.29; $SD = 0.34$), and then non-manipulable objects (mean = 0.66; $SD = 0.85$). A univariate ANOVA showed that indeed categories were guessed to different degrees $F(2,27) = 20.417$; $p < 0.001$; $\eta^2 = 0.602$. Pairwise comparisons after Bonferroni corrections revealed that actions were guessed significantly more accurately than manipulable (CI = 0.497-1.158; $p = 0.001$) and non-manipulable objects (CI = 0.996-2.343; $p = 0.001$) and these in turn did not differ from each other (CI = -1.304-0.427; $p = 0.072$).

The same pattern emerged in the Dutch group: actions with objects had the highest accuracy scores (mean = 2.31; $SD = 0.78$), followed by manipulable objects (mean = 1.23; $SD = 0.66$), and finally, non-manipulable objects (mean = 0.768; $SD = 0.973$). A univariate ANOVA confirmed that the three different conditions were guessed differentially $F(2,27) = 19.553$; $p < 0.001$; $\eta^2 = 0.457$. Pairwise comparisons after Bonferroni corrections shows that there is a significant difference between actions and manipulable objects (CI = 0.228-1.925; $p = 0.010$) and non-manipulable objects (CI = 0.698-2.395; $p < 0.001$); but there was no difference between the two types of objects (CI = -0.378-1.318; $p = 0.507$). See Figure 6.

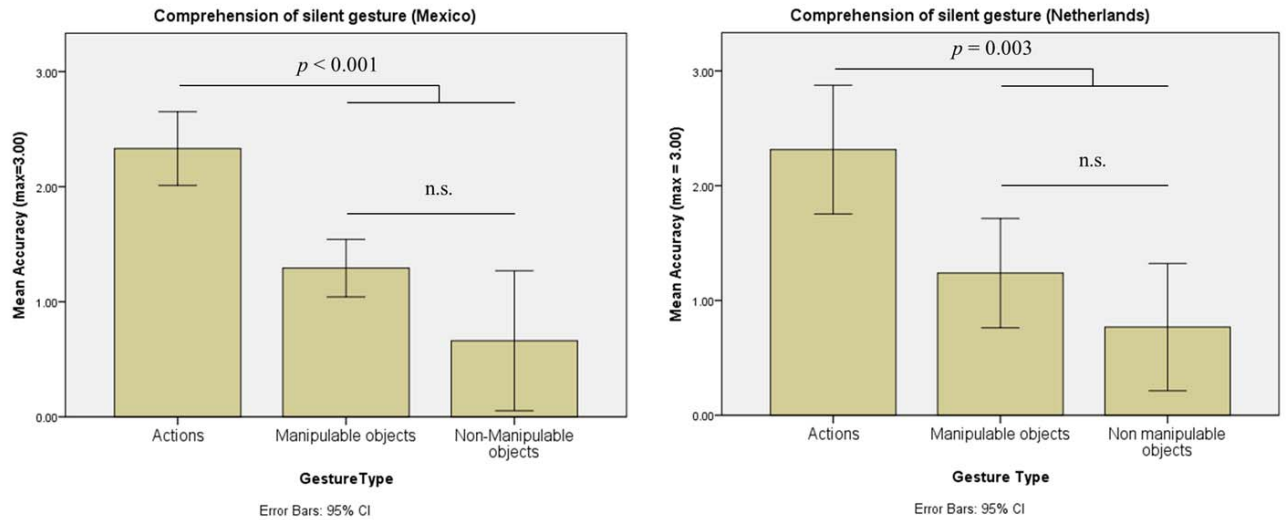


Figure 6. Comprehension mean accuracy scores across semantic categories across cultural groups (maximum score: 3.0). Bars represent standard deviation.

In order to evaluate the contribution of single vs multi-gesture strings in the comprehension of manipulable and non manipulable objects, we separated items that were depicted with one or more gestures for both cultural groups. In the Mexican group, manipulable objects with a single gesture (mean=1.14, $SD = 0.18$) yielded less accuracy than gestures with multiple units (mean=1.44, $SD = 0.42$). The same pattern was observed in non-manipulable objects in that concepts with a single gesture (mean=0.58, $SD = 1.09$) reached lower accuracy than concepts expressed with multiple gestures (mean=0.73, $SD = 0.64$). The Dutch group exhibits the same trend. In manipulable objects, single gestures elicited lower accuracy (mean=0.83, $SD = 0.31$) than multiple gestures (mean=1.51, $SD = 0.73$); and in non-manipulable objects, single gestures also elicited lower accuracy (mean=0.33, $SD = 0.31$) than multiple gestures (mean=1.21, $SD = 0.91$). This shows that multiple gestures are a more effective strategy to represent objects than single gestures, because, we argue, single gestures are an ambiguous signal that is harder to interpret.

We also explored whether single gestures or their combination were interpreted as nouns, verbs, or another category (e.g., adjective) and compared their prevalence across conditions. We found that out of 390 possible responses, the Mexican group produced a total of 285 verbs, 90 nouns, and 13 adjectives (Table 4). A chi-square revealed that there was no significant difference between actions and manipulable objects, which indicates a dominant interpretation of gestures as actions (verbs) instead of objects (nouns) for these two categories $\chi(1)^2 = 2.6006$, $p = 0.272$. In contrast, there was a significant difference in the distribution of nouns, verbs, and adjectives between manipulable and non-manipulable objects $\chi(1)^2 = 14.606$, $p < 0.0001$, and between actions and non-manipulable objects $\chi(1)^2 = 27.1256$, $p < 0.0001$. Looking only at manipulable and non-manipulable objects, it was found that there was higher prevalence of noun responses when the referent was represented with multiple gestures than with a single one $\chi(1)^2 = 6.254$, $p = 0.043$. For non-manipulable objects, there were more noun responses when the referent was depicted with a single (*drawing*) gestures than with multiple gestures $\chi(1)^2 = 25.917$, $p < 0.0001$.

Table 4. Proportion of the distribution of responses according to word type across condition for Mexican and Dutch participants. In the manipulable and non-manipulable conditions there were five items represented with a single gesture and 5 items represented with multiple.

Mexican					
	Actions	Manipulable Objects		Non-manipulable Objects	
	Single <i>acting</i> (N=130)	Single <i>acting</i> (N=65)	Combinations (N=65)	Single <i>drawing</i> (N=65)	Combinations (N=65)
Verbs	0.85	0.86	0.55	0.34	0.78
Nouns	0.14	0.12	0.31	0.49	0.20
Adjectives	0.01	0.02	0.03	0.14	0.02
	1.00	1.00	1.00	1.00	1.00
Dutch					
	Actions	Manipulable Objects		Non-manipulable Objects	
	Single <i>acting</i> (N=130)	Single <i>acting</i> (N=65)	Combinations (N=65)	Single <i>drawing</i> (N=65)	Combinations (N=65)

Verbs	0.92	0.80	0.43	0.17	0.49
Nouns	0.08	0.15	0.57	0.78	0.46
Adjectives	0.00	0.00	0.00	0.00	0.00
	1.00	1.00	1.00	1.00	1.00

Regarding the Dutch data, out of 390 possible responses, participants produced 242 verbs and 138 nouns (no adjectives). A chi-square revealed that there was no significant difference between actions and manipulable objects with more action (verb) responses for both categories $\chi(1)^2 = 2.460, p < 0.116$. There was a significant difference in the distribution of nouns and verbs between manipulable and non-manipulable objects $\chi(1)^2 = 20.761, p < 0.0001$, and between actions and non-manipulable objects $\chi(1)^2 = 40.30, p < 0.0001$. Looking only at both types of objects, we found that for manipulable objects there was higher prevalence of noun responses when the referent was represented with multiple gestures than with a single one $\chi(1)^2 = 22.652, p < 0.0001$. For non-manipulable objects, there were more noun responses when the referent was depicted with a single (*drawing*) gestures than with multiple gestures $\chi(1)^2 = 15.701, p < 0.0001$ (Table 4).

The results of the comprehension study can be summarised as follows: gestures representing actions, which were expressed primarily with a single *acting* gesture, were the most accurately guessed followed by manipulable objects and non-manipulable objects. Both types of objects represented through multiple gestures were guessed more accurately than those expressed with a single gesture. Further, we found that overall there was an overwhelming preference to interpret gestures as actions (verbs) instead of objects (nouns). Looking in more detail into participants' responses, we found that in both groups, when the dominant type of representation is *acting* (actions and manipulable objects) they assume that the referent is an action (verb). This is an accurate assumption for the category actions with objects but not for manipulable objects. However, the presence of deictics or multiple gestures in manipulable objects drives participants away from interpreting gestures as action and instead they interpret them as an object

associated with such action. When the preferred type of representation is *drawing* (non-manipulable objects), participants are significantly more inclined to interpret gestures as objects, and interestingly, the presence of multiple gestures does not add much to their accurate interpretation. This suggests that the *drawing* type of iconic depiction on its own has a heavy semantic load that skews participants to assume that the referent is an object. Importantly, these results hold for both Mexican and Dutch participants. These data allow us to confirm that the strategies implemented in gesture production facilitate identification of the intended target and that these patterns can be generalised across two cultures.

Discussion

In two studies we have shown that when speakers of different languages are asked to communicate concepts in silent gesture they align different types of iconic representations with specific semantic categories. The representation of actions is most frequently and easily achieved in a single gesture when they are expressed through *acting* strategy because there is a direct correspondence between gesture and referent. However, the representation of objects in a single gesture might result in a vague signal so people resort to generating combinations of gestures to reduce ambiguity. By highlighting elements of an iconic gesture through pointing, or through the combination of different types of iconic representations, gesturers reduce the number of possible interpretations of this otherwise ambiguous manual depiction (see Figure 8 for a summary of results). The results from the comprehension task show that gestures employing the *acting* strategy to represent actions yield the highest accuracy because of the direct correspondences with the referent. Objects in general are interpreted more accurately when they are represented through the combination of two gestural units than when they are depicted with a single gesture. Mexican and Dutch participants exhibit the same behaviours in both production and comprehension. These results support claims that the communicative pressure to reduce ambiguity (Bell, 1984; Gibson et al., 2013), and the physical constraints of the body as iconic device (Hall, Mayberry, & Ferreira, 2013) shape the form of

spontaneous silent gesture to distinguish concepts across semantic categories. These strategies are effectively exploited by viewers of these gestures to interpret accurately the intended referent. The striking similarities between our results and the strategies to distinguish actions and objects in emerging sign languages (Haviland, 2013; Meir et al., 2010) suggest that these strategies are part of our cognitive system and constitute the raw materials from which new communicative systems begin to emerge.

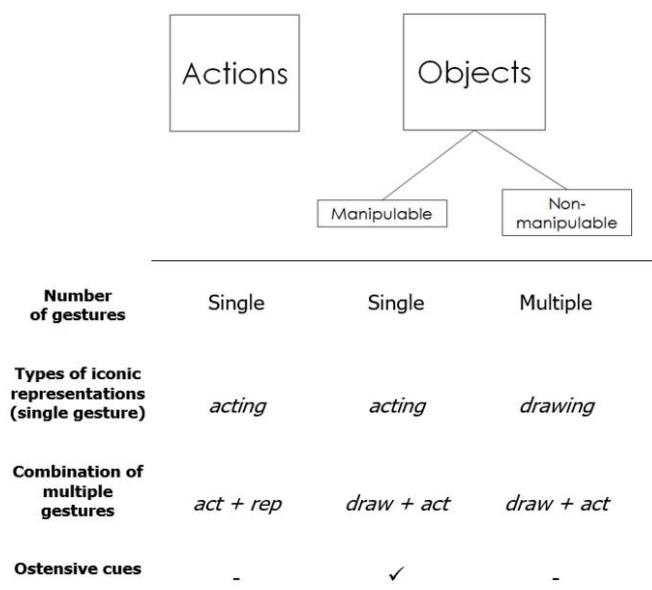


Figure 8. Summary of the results for each semantic category and the different strategies deployed in silent gesture. *Acting* was the most common strategy in single and multi-gesture combinations.

Types of iconic depictions

We replicated earlier studies (Padden et al., 2015, 2013; van Nispen, van de Sandt-Koenderman, Mol, & Krahmer, 2014) in that we found that participants showed a strong preference for the *acting* mode of representation in gestures depicting actions (and manipulable objects) which in turn were the most accurately guessed. In order to generate a gesture, participants had to retrieve a mental image of a concept and this simulated a motor plan associated with it. This is the reason why actions and manipulable objects

had a strong preference for the *acting* mode of representation, which is also a common iconic unit in multi-gestural strings (see Figure 7). In comprehension, the *acting* strategy tapped into participants' shared motor schemas and boosted performance in interpreting the intended referent.

Research has shown that gestures depicting actions are the first to be produced by children from different cultures (Pettenati, Sekine, Congestri, & Volterra, 2012; Pettenati, Stefanini, & Volterra, 2010), they are the first to be understood by toddlers (Tolar, Lederberg, Gokhale, & Tomasello, 2008); and deaf children learning a sign language have also shown a strong preference for signs that express the actions associated with a referent (Ortega, Sümer, & Ozyürek, 2017). The current study gives yet further evidence of the prevalence of the *acting* strategy in iconic manual depictions both in production and comprehension. This bias can be explained by our conceptual knowledge being grounded in our motor experiences (Barsalou, 1999, 2008) and gestures being manifestations of the actions with which we relate to the world (Cook & Tanenhaus, 2009; Hostetter & Alibali, 2008).

We also found that when an object does not lend itself to a clear form of manipulation or the affordances of the object does not permit the use of an *acting* representation, individuals turn to an alternative strategy. Similar to what has been reported for co-speech gestures (Masson-Carro et al., 2015), non-manipulable objects were more likely to be depicted through *drawing* (Dutch) or through an even distribution of other depicting strategies (Mexicans) (see Table 1). While speakers have a strong bias to depict a concept through the re-enactment of human-object interactions, when the referent has limited action affordances people implement an alternative iconic form of representation (i.e., *drawing*). These data speak in favour of claims that the natural properties of a referent and their structural iconicity will shape the form and sequencing of silent gestures (Christensen et al., 2016).

It is worth noting that gesturers did *not* show a clear distinction in aligning *acting* depictions for actions and the *representing* strategy for objects as has been reported for established and emerging sign

languages (Padden et al., 2015). In fact, the *representing* strategy was the least commonly attested strategy in both cultural groups. It may be possible that this type of distinction emerges within a deaf community who interact on a daily basis and is the result of conventionalisation. However, it is important to highlight that this is not the only means to mark action-object distinctions. Sign languages can also resort to different hand movements (Supalla & Newport, 1986), mouth patterns (Johnston, 2001), in addition to different types of iconic representations (Padden et al., 2013). Critically, all these possibilities often co-exist within the same linguistic system. Future investigations should consider the factors that push an emerging sign language to mark semantic distinctions with one strategy over another while bearing in mind that more than one mechanism is possible.

Combinatorial patterning in silent gesture

In order to express different types of objects, and when a single type of iconic depiction was insufficient, individuals across cultures used multi-gesture stretches and ostensive cues to depict manipulable and non-manipulable objects. Sequences such as *drawing* + *acting* (e.g., ‘pillow’) were an efficient strategy because they restricted the number of potential referents. When manipulable objects were represented with a single *acting* gesture participants wrongly assumed that the referent was an action instead of an object (e.g., target: spoon, answer: eating) or because they guessed a wrong word with the same shape as the tracing gesture (e.g., target: pyramid, answer: church). Similarly, ostensive cues, which often consisted of an *acting* gesture followed by a pointing or showing deictic (e.g., pointing at an imaginary toothbrush), were another strategy implemented to make semantic distinctions, primarily for manipulable objects. In order to avoid giving the impression that they were referring to an action, participants highlighted elements of their gesture to make the differentiation. Data from the comprehension study support our claim given that manipulable objects were more accurately guessed when they were depicted through multiple gestures than with a single one.

The sequencing of two meaningful units with different types of iconic depictions is reported to be a powerful strategy in emerging sign languages such as Al-Sayyid Bedouin Sign Language (ABSL) (Meir, Sandler, Padden, & Aronoff, 2012) and Yucatec Mayan Sign Language (YMSL) (Safar & Petatillo, in preparation). Interestingly, these young sign languages have been observed to produce two-sign sequences to refer to hand-held objects (Meir et al., 2010; Tkachman & Sandler, 2013) and single signs to refer to actions associated with those objects (Safar & Petatillo, in preparation). In an elicitation task with YMSL users, for example, the concept of ‘chopping with a machete’ is depicted through a single sign showing the rapid cutting movement of the blade. In contrast, when referring to the ‘machete’ itself, YMSL signers produce this same sign followed by a sign describing its length (size and shape specifiers –SASS- in the sign language literature, analogous to *drawing* gestures) (see Fig. 9). In our study, gesturers, who lack a conventionalised manual lexicon, are also capable of making subtle semantic distinctions through the combination of gestures with different iconic modes of representation. Critically, this level of combinatorial patterning is in place at the first instance of spontaneous gestural production and without repeated iterations of cultural transmission, as has been reported in other studies (Micklos, 2016; Motamedi, Schouwstra, Culbertson, Smith, & Kirby, 2018).



Figure 9. A Yucatec Maya Sign Language user describes the action of ‘chopping with a machete’ with a single sign, but to refer to the ‘machete’ he adds a sign describing the length of the tool (Safar & Petatillo, in preparation).

Interestingly, the combination of two gestural units is observed in two cultures that are geographically distant and that have typological distinct linguistic features. There is abundant research showing that despite the fact that gestures are present in every culture, there are important differences in gestural forms because they are shaped by the linguistic encoding of each language (Kita & Özyürek, 2003) or cultural conventions (Nyst, 2016). What is remarkable is that when these two unrelated cultures are prompted to produce silent gestures they follow remarkably similar strategies in the implementation of different iconic strategies and their combinations to discriminate semantic categories in the manual modality. In line with earlier research (Goldin-Meadow, McNeill, & Singleton, 1996), we find that production of silent gestures does not seem to be heavily mediated by language but rather originate from shared conceptual representations.

Conclusion

Iconicity is a key tool to convey meaning in silent gesture with the *acting* strategy occupying a chief place given its prevalence in most types of depictions. Single iconic gestures with a certain type of iconicity can do just so much to distinguish between categories. Gesturers employ additional strategies such as combining gestures with different types of iconicity or highlighting certain elements of a gesture with deictics that together go above and beyond the communicative power of individual gestures. Humans have at their disposal a powerful communicative system that can not only depict concrete referents but also, discriminate across different semantic categories. We argue that the two cultures under study show striking similarities because their iconic depictions originate from their capacity to iconically depict a referent with the visible bodily articulators and the need to produce unambiguous signals. Of huge

significance is the finding that these strategies support comprehension in different degrees, which suggests that they could be exploited effectively for communicative purposes in the absence of a linguistic system.

Silent gestures have been argued to be a unique form of manual communication that adopts linguistic properties akin to those of signs due to their discrete nature (Goldin-Meadow & Brentari, 2017; Goldin-Meadow et al., 1996) and because they reflect at least partially the structures observed in emerging sign languages (Meir, et al., 2017; Meir, Sandler, Padden, & Aronoff, 2012). Our study lends support to these claims and goes beyond by arguing that they also exhibit combinatorial patterning before iterated learning and transmission. Our conception of the world, the physical attributes of referents, the constraints of the body as a channel of communication, and the human capacity to produce an informative signal, operate in tandem to shape the structure of this mode of communication. Our study adds to the mounting body of evidence showing that systematic conventions emerge instantaneously in light of the common ground shared amongst interlocutors and the inferences they make to produce a clear, unambiguous signal (Bell, 1984; Clark & Murphy, 1982; Gibson et al., 2013; Misyak, Noguchi, & Chater, 2016). This study further supports claims that iconicity lies at the core of language phylogeny (Vigliocco et al., 2014) by bootstrapping an effective communicative system in the manual modality in the absence of linguistic means.

References

- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577–609. <https://doi.org/10.1017/S0140525X99252144>
- Barsalou, L. W. (2008). Grounding Symbolic Operations in the Brain's Modal Systems. In G. Semin & E. Smith (Eds.), *Embodied Grounding. Social, Cognitive, Affective and Neuroscientific Approaches* (pp. 9–42). Cambridge: Cambridge University Press.
- Bell, A. (1984). Language style as audience design. *Language in Society*, 13, 240–250.
- Brentari, D., Renzo, A. Di, Keane, J., & Volterra, V. (2015). Cognitive, cultural, and linguistic sources of a handshape distinction expressing agentivity. *Topics in Cognitive Science*, 7(1), 95–123. <https://doi.org/10.1111/tops.12123>

- Christensen, P., Fusaroli, R., & Tylén, K. (2016). Environmental constraints shaping constituent order in emerging communication systems: Structural iconicity, interactive alignment and conventionalization. *Cognition*, 146, 67–80. <https://doi.org/10.1016/j.cognition.2015.09.004>
- Chu, M., & Kita, S. (2016). Co-thought and co-speech gestures are generated by the same action generation process. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 42(2), 257–270.
- Clark, H., & Murphy, G. (1982). Audience design in meaning and reference. In J. F. Le Ny & W. Kintsch (Eds.), *Language and Comprehension* (pp. 287–300). Amsterdam.
- Cook, S. W., & Tanenhaus, M. K. (2009). Embodied communication: speakers' gestures affect listeners' actions. *Cognition*, 113(1), 98–104. <https://doi.org/10.1016/j.cognition.2009.06.006>
- Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity and systematicity in language. *Trends in Cognitive Sciences*, 19(10), 603–615. <https://doi.org/10.1016/j.tics.2015.07.013>
- Fay, N., Arbib, M., & Garrod, S. (2013). How to bootstrap a human communication system. *Cognitive Science*, 37(7), 1356–1367. <https://doi.org/10.1111/cogs.12048>
- Galantucci, B., & Garrod, S. (2011). Experimental Semiotics: A Review. *Frontiers in Human Neuroscience*, 5(February), 1–15. <https://doi.org/10.3389/fnhum.2011.00011>
- Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A Noisy-Channel Account of Crosslinguistic Word-Order Variation. *Psychological Science*, 24(7), 1079–1088. <https://doi.org/10.1177/0956797612463705>
- Goldin-Meadow, S., & Brentari, D. (2017). Gesture, sign and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences*, (November), 1–59. <https://doi.org/10.1017/S0140525X15001247>
- Goldin-Meadow, S., McNeill, D., & Singleton, J. (1996). Silence is liberating: removing the handcuffs on grammatical expression in the manual modality. *Psychological Review*, 103(1), 34–55. <https://doi.org/10.1037/0033-295X.103.1.34>
- Goldin-Meadow, S., So, W. C., Ozyürek, A., & Mylander, C. (2008a). 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(27), 9163–9168. <https://doi.org/10.1073/pnas.0710060105>
- Goldin-Meadow, S., So, W. C., Ozyürek, A., & Mylander, C. (2008b). 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(27), 9163–9168. <https://doi.org/10.1073/pnas.0710060105>
- Hall, M. L., Ahn, Y. D., Mayberry, R. I., & Ferreira, V. S. (2015). Production and comprehension show divergent constituent order preferences: Evidence from elicited pantomime. *Journal of Memory and Language*, 81, 16–33. <https://doi.org/10.1016/j.jml.2014.12.003>
- Hall, M. L., Mayberry, R. I., & Ferreira, V. S. (2013). Cognitive constraints on constituent order: evidence

- from elicited pantomime. *Cognition*, 129(1), 1–17. <https://doi.org/10.1016/j.cognition.2013.05.004>
- Hassemer, J., & Winter, B. (2018). Decoding gestural iconicity. *Cognitive Science*, 1–16. <https://doi.org/10.1111/cogs.12680>
- Haviland, J. (2013). The emerging of nouns in a first generation sign language: Specification, iconicity, and syntax. *Gesture*, 13(3), 309–353.
- Hockett, C. (1960). The origin of speech. *Scientific American*, 203, 88–111.
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, 15(3), 495–514. <https://doi.org/10.3758/PBR.15.3.495>
- Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369(20130298). <https://doi.org/10.1098/rstb.2013.0298>
- Johnston, T. (2001). Nouns and verbs in Australian sign language: An open and shut case? *Journal of Deaf Studies and Deaf Education*, 6(4), 235–257. <https://doi.org/10.1093/deafed/6.4.235>
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
- Kimmelman, V., Klezovich, A., & Moroz, G. (2018). Iconicity patterns in sign languages. Retrieved from https://sl-iconicity.shinyapps.io/iconicity_patterns/
- Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, 48(1), 16–32. [https://doi.org/10.1016/S0749-596X\(02\)00505-3](https://doi.org/10.1016/S0749-596X(02)00505-3)
- Klima, E., & Bellugi, U. (1979). *The Signs of Language*. Harvard: Harvard University Press.
- Mandel, M. A. (1977). Iconic devices in American Sign Language. In A. Friedman (Ed.), *On the Other Hand: New Perspectives on American Sign Language* (pp. 57–107). New York: Academic Press.
- Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2015). Can you handle this? The impact of object affordances on how co-speech gestures are produced. *Language, Cognition and Neuroscience*, 1–11. <https://doi.org/10.1080/23273798.2015.1108448>
- Meir, I., Aronoff, M., Börstell, C., Hwang, S., Ilkbasaran, D., Kastner, I., ... Sandler, W. (2017). The origin of grammatical word order: insights from novel communication systems and young sign languages. *Cognition*, 158, 189–207. <https://doi.org/10.1016/j.cognition.2016.10.011>
- Meir, I., Aronoff, M., Sandler, W., & Padden, C. A. (2010). Sign languages and compounding. In S. Scalise & I. Vogel (Eds.), *Cross-disciplinary Issues in Compounding* (pp. 301–322). John Benjamins Publishing Company.
- Meir, I., Sandler, W., Padden, C., & Aronoff, M. (2012). Emerging Sign Languages. In M. Marschark & P. Spencer (Eds.), *The Oxford Handbook of Deaf Studies, Language, and Education* (Vol. 2, pp. 1–22). Oxford, UK. <https://doi.org/10.1093/oxfordhb/9780195390032.013.0018>
- Micklos, A. (2016). Interaction for facilitating conventionalization: Negotiating the silent gesture

communication of noun-verb pairs. In S. Roberts, C. Cuskley, L. McCrohon, L. Barcelo-Coblijn, O. Feher, & T. Verhoef (Eds.), *The Evolution of Language: Proceedings of the 11th International Conference (EVLANG11)*. New Orleans: EVOLANG. Retrieved from <http://evolang.org/neworleans/papers/143.html>

- Misyak, J., Noguchi, T., & Chater, N. (2016). Instantaneous Conventions : The Emergence of Flexible Communicative Signals. *Psychological Science*, 1–12. <https://doi.org/10.1177/0956797616661199>
- Motamedi, Y., Schouwstra, M., Culbertson, J., Smith, K., & Kirby, S. (2018). Evolving artificial sign languages in the lab: from improvised gesture to systematic sign. <https://doi.org/https://doi.org/10.31234/osf.io/be7qy>
- Müller, C. (2013). Gestural modes of representation as techniques of depiction. In C. Müller, A. Cienki, S. Ladewig, D. McNeill, & J. Bressemer (Eds.), *Body - Language - Communication: An International Handbook on Multimodality in Human Interaction* (pp. 1687–1701). Berlin: De Gruyter Mouton.
- Müller, C. (2016). From mimesis to meaning: A systematics of gestural mimesis for concrete and abstract referential gestures. In *Meaning, mind and communication: Explorations in cognitive semiotics* (Zlatev, J.). Frankfurt am Main: Peter Lang.
- Nyst, V. (2016). The depiction of size and shape in gestures accompanying object descriptions in Anyi (Côte d'Ivoire) and in Dutch (The Netherlands). *Gesture*, 15(2), 156–191. <https://doi.org/10.1075/gest.15.2.02nys>
- Ortega, G., Sümer, B., & Özyürek, A. (2017). Type of iconicity matters in the vocabulary development of signing children. *Developmental Psychology*, 53(1), 89–99.
- Ortega, G., Sümer, B., & Özyürek, A. (2014). Type of iconicity matters: Bias for action-based signs in sign language acquisition. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *36th Annual Conference of the Cognitive Science Society* (pp. 1114–1119). Austin, TX: Cognitive Science Society.
- Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2016). Does language shape silent gesture? *Cognition*, 148, 10–18. <https://doi.org/10.1016/j.cognition.2015.12.001>
- Padden, C., Hwang, S.-O., Lepic, R., & Seegers, S. (2015). Tools for Language: Patterned Iconicity in Sign Language Nouns and Verbs. *Topics in Cognitive Science*, 7(1), 81–94. <https://doi.org/10.1111/tops.12121>
- Padden, C., Meir, I., Hwang, S.-O., Lepic, R., Seegers, S., & Sampson, T. (2013). Patterned iconicity in sign language lexicons. *Gesture*, 13(3), 287–305.
- Perlman, M., & Lupyan, G. (2018). People Can Create Iconic Vocalizations to Communicate Various Meanings to Naïve Listeners. *Scientific Reports*, 8(1), 2634. <https://doi.org/10.1038/s41598-018-20961-6>
- Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology*, 1(227), 1664–1678. <https://doi.org/dx.doi.org/10.3389/fpsyg.2010.00227>
- Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity : from a world of experience to the experience

of language The bridge of iconicity : from a world of experience to the experience of language. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369(20130300). <https://doi.org/dx.doi.org/10.1098/rstb.2013.0300>

- Pettenati, P., Sekine, K., Congestri, E., & Volterra, V. (2012). A Comparative Study on Representational Gestures in Italian and Japanese Children. *Journal of Nonverbal Behavior*, 36(2), 149–164. <https://doi.org/10.1007/s10919-011-0127-0>
- Pettenati, P., Stefanini, S., & Volterra, V. (2010). Motoric characteristics of representational gestures produced by young children in a naming task. *Journal of Child Language*, 37(4), 887–911. <https://doi.org/10.1017/S0305000909990092>
- Ramachandran, S., & Hubbard, E. M. (2001). Synaesthesia: a window to perception, thought and language. *Journal of Consciousness Studies*, 8(12), 3–34.
- Roomer, E. K., Hoogerwerf, A. C., & Linn, D. E. (2011). *Boston benoem taak 2011*. Utrecht.
- Safar, J., & Petatillo, R. (n.d.). Strategies of noun-verb distinction in Yucatec Maya Sign Languages. In O. Le Guen, J. Safar, & M. Coppola (Eds.), *Emerging Sign Languages of the Americas. Sign Language Typology Series*. Berlin: Mouton de Gruyter.
- Sloetjes, H., & Wittenburg, P. (2018). ELAN (version 5.2). Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from <https://tla.mpi.nl/tools/tla-tools/elan/>
- Smith, K., & Kirby, S. (2008). Cultural evolution: implications for understanding the human language faculty and its evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 363(1509), 3591–3603. <https://doi.org/10.1098/rstb.2008.0145>
- Sulik, J. (2018). Cognitive mechanisms for inferring the meaning of novel signals during symbolisation. *PLoS ONE*, 13(1), 1–27. <https://doi.org/10.1371/journal.pone.0189540>
- Supalla, T., & Newport, E. L. (1986). How many sits in a chair? The derivations of nouns and verbs in American Sign Language. In P. Siple (Ed.), *Understanding language through sign language research* (pp. 91–132). New York: Academic Press.
- Tkachman, O., & Sandler, W. (2013). The noun-verb distinction in two young sign languages. *Gesture*, 13(3), 253–286.
- Tolar, T. D., Lederberg, A. R., Gokhale, S., & Tomasello, M. (2008). The development of the ability to recognize the meaning of iconic signs. *Journal of Deaf Studies and Deaf Education*, 13(2), 225–240. <https://doi.org/dx.doi.org/10.1093/deafed/enm045>
- van Nispen, K., van de Sandt-Koenderman, M., Mol, L., & Krahmer, E. (2014). Pantomime Strategies: On Regularities in How People Translate Mental Representations into the Gesture Modality. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society (CogSci 2014)* (pp. 3020–3026). Austin, TX: Cognitive Science Society, Inc.
- Van Nispen, K., Van De Sandt-Koenderman, W. M. E., & Krahmer, E. (2017). Production and Comprehension of Pantomimes Used to Depict Objects. *Frontiers in Psychology*, 8(July), 1095. <https://doi.org/10.3389/FPSYG.2017.01095>

- Verhoef, T., Kirby, S., & de Boer, B. (2015). Iconicity and the Emergence of Combinatorial Structure in Language. *Cognitive Science*, 1–26. <https://doi.org/10.1111/cogs.12326>
- Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369(20130292). <https://doi.org/dx.doi.org/10.1098/rstb.2013.0292>

Appendix

Items shown to Mexican and Dutch participants along with their manipulability ratings (only for manipulable and non-manipulable objects)

1) Mexican stimuli

	English	Spanish	Category
1	to phone	llamar	Actions with objects
2	to drink	tomar	Actions with objects
3	to smoke	fumar	Actions with objects
4	to hammer	martillar	Actions with objects
5	to saw	serruchar	Actions with objects
6	to comb	peinar	Actions with objects
7	to cut (knife)	cortar con cuchillo	Actions with objects
8	to put lipstick on	poner labial	Actions with objects
9	to soap	enjabonar	Actions with objects
10	to iron	planchar	Actions with objects

	English	Spanish	Category	Rating
11	toothbrush	cepillo de dientes	Manipulable Objects	6.23
12	lightbulb	foco	Manipulable Objects	5.92
13	lighter	encendedor	Manipulable Objects	6.08
14	spoon	cuchara	Manipulable Objects	6.31
15	telephone	teléfono	Manipulable Objects	6.31
16	racket	raqueta	Manipulable Objects	6.23
17	cigarette	cigarro	Manipulable Objects	6.46
18	bottle	botella	Manipulable Objects	5.62
19	mug	taza	Manipulable Objects	6.15
20	pillow	cojín	Manipulable Objects	5.69
Mean:				6.10

	English	Spanish	Category	Rating
21	bed	cama	Non-manipulable objects	3.31
22	stairs	escaleras	Non-manipulable objects	2.54
23	train	tren	Non-manipulable objects	2.15
24	floor	piso	Non-manipulable objects	1.62
25	bridge	puente	Non-manipulable objects	1.31
26	wall	muro	Non-manipulable objects	1.85
27	sofa	sofá	Non-manipulable objects	2.92
28	pyramid	pirámide	Non-manipulable objects	1.15
29	table	mesa	Non-manipulable objects	3.92
30	car	coche	Non-manipulable objects	3.46
Mean:				2.42

2) Dutch stimuli

	English	Dutch	Category
1	to drink	drinken	Actions with objects
2	to comb	kammen	Actions with objects
3	to cut (knife)	snijden	Actions with objects
4	to smoke	roken	Actions with objects
5	to saw	zaggen	Actions with objects
6	to phone	telefoneren	Actions with objects
7	to put lipstick on	lippenstiften	Actions with objects
8	to pour	gieten	Actions with objects
9	to soap	inzeepen	Actions with objects
10	to hammer	hameren	Actions with objects

	English	Dutch	Category	Rating
11	toothbrush	tandenborstel	Manipulable Objects	7.00
12	clock	klok	Manipulable Objects	5.09
13	lightbulb	gloeilamp	Manipulable Objects	5.18
14	lighter	aansteker	Manipulable Objects	6.45
15	mug	mok	Manipulable Objects	6.55
16	telephone	telefoon	Manipulable Objects	6.18
17	pillow	het kussen	Manipulable Objects	5.45
18	racket	tennis racket	Manipulable Objects	6.09
19	soap	zeep	Manipulable Objects	6.27
20	spoon	lepel	Manipulable Objects	6.82
Mean:				6.11

	English	Dutch	Category	Rating
21	floor	verdieping	Non-manipulable objects	1.36
22	sun	zon	Non-manipulable objects	1.36
23	bed	bed	Non-manipulable objects	3.82
24	building	gebouw	Non-manipulable objects	1.45
25	table	tafel	Non-manipulable objects	4.27
26	wall	muur	Non-manipulable objects	2.09
27	pyramid	piramide	Non-manipulable objects	1.09
28	sofa	sofa	Non-manipulable objects	3.73
29	train	trein	Non-manipulable objects	1.82
30	bookshelf	boekenplank	Non-manipulable objects	4.82
Mean:				2.58

Captions

Figure 1. Different types of iconic representations in gesture. (A) *Acting*: shows a person pretending to hold a cigarette with two fingers near the mouth for ‘smoking’. (B) *Representing*: the gesturer depicts a ‘pair of scissors’ through extended index and middle fingers with open-close movement. (C) *Drawing*: a ‘window’ is represented with two index fingers tracing a square. (D) *Moulding*: open palms touching at the tip followed by downward movement represent the three-dimensional shape of ‘house’.

Figure 2. Mean number of gestures per condition across gestural groups (range: 1-4 gestures. Bars represent standard error.

Figure 3. Examples of the most frequent types of iconicity used for different semantic categories in both cultures. (A, D) The *acting* strategy is used to depict the concept ‘hammering’ by showing how a hammer is manipulated. (B, E) The concept ‘mug’ depicted through the *acting* strategy by re-enacting the action of drinking from a mug. (C, F) ‘Pyramid’ is represented through *drawing* because the hands trace its triangular outline.

Table 1. Percentage of gestures according to their type of iconic depiction across conditions in both cultural groups. N= instances in which a concept was depicted with a single gesture across participants (200 descriptions per condition in total).

Figure 4. Example of a multi-gesture depiction for the manipulable object ‘pillow’. Participants draw a square shape and then perform an acting gesture re-enacting the action of sleeping.

Table 2. Proportion of multi-gesture combinations. N=number of instances in which participants produced more than one gesture (200 descriptions per condition in total). Deictics are not included. The Other category includes sequences of different gestures using the same type of iconicity (e.g., acting + acting).

Figure 5. Examples of gestures incorporating a deictic gesture. For the concept ‘toothbrush’ participants across cultural groups often produced an acting gesture of brushing one’s teeth and then showed or pointed at an imaginary toothbrush (see upper panels). For ‘lighter’ and ‘cigarette’, participants in the Netherlands and Mexico re-enacted lighting a cigarette and then pointing to an imaginary object in their hands.

Table 3. Summary of the stimuli and their type of representation (single vs. multiple gestures) across conditions

Figure 6. Mean accuracy scores across semantic categories across cultural groups (maximum score: 3.0). Bars represent standard deviation.

Table 4. Distribution of responses according to word type across condition for Mexican participants. Maximum number of responses per category = 130 (13 participants x 10 items per condition)

Table 5. Distribution of responses according to word type across manipulable and non-manipulable objects for Mexican participants. There were five items represented with a single gesture and 5 items represented with multiple gestures. Maximum number of responses per condition = 65 (13 participants x 5 items per condition)

Table 6. Distribution of responses according to word type across condition for Dutch participants. Maximum number of responses per category = 130 (13 participants x 10 items per condition)

Table 7. Distribution of responses according to word type across manipulable and non-manipulable objects for Dutch participants. There were five items represented with a single gesture and 5 items represented with multiple gestures. Maximum number of responses per condition = 65 (13 participants x 5 items per condition)

Figure 7. Summary of the results for each semantic category and the different strategies deployed in silent gesture

Figure 8. A Yucatec Maya Sign Language user describes the action of ‘chopping with a machete’ with a single sign, but to refer to the ‘machete’ he adds a sign describing the length of the tool (Safar & Petatillo, in preparation).