

## **UC Merced**

# **Proceedings of the Annual Meeting of the Cognitive Science Society**

### **Title**

Systematicity and arbitrariness in novel communication systems

### **Permalink**

<https://escholarship.org/uc/item/6fr703z6>

### **Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 31(31)

### **ISSN**

1069-7977

### **Authors**

Kirby, Simon  
Oberlander, Jon  
Theisen, Carrie Ann

### **Publication Date**

2009

Peer reviewed

# Systematicity and Arbitrariness in Novel Communication Systems

**Carrie Ann Theisen (C.A.Theisen@sms.ed.ac.uk)**

School of Informatics, University of Edinburgh  
Informatics Forum, 10 Crichton Street, Edinburgh EH8 9AB, UK

**Jon Oberlander (jon@inf.ed.ac.uk)**

School of Informatics, University of Edinburgh  
Informatics Forum, 10 Crichton Street, Edinburgh EH8 9AB, UK

**Simon Kirby (simon@ling.ed.ac.uk)**

Language Evolution and Computation Research Unit, University of Edinburgh  
Dugald Stewart Building, 3 Charles Street, Edinburgh EH8 9AD, UK

## Abstract

Human languages include vast numbers of learned, arbitrary signal-meaning mappings but also many complex signal-meaning mappings that are systematically related to each other (i.e. *not* arbitrary). Although arbitrariness and systematicity are clearly related, the development of the two in communication systems has been explored independently. We present an experiment in which participants invent signs from scratch to refer to a set of real concepts that share semantic features. Through interaction, the systematic re-use of arbitrary elements emerges.

**Keywords:** arbitrariness; systematicity; signs; language evolution; emergent communication

## Introduction

Two of language's most fascinating properties, arbitrariness and systematicity, characterize the nature of the mappings between signals and meanings. A sign is *arbitrary* when there is no inherent relationship between the signal and its meaning. For example, the sounds in the word "house" have nothing to do with what the word means. In contrast, some subsets of signs in a language are *systematic*, in that signals for similar meanings share an element. The referring expressions "big house", "red house", "big apple", and "red apple" are an example. In language, words are often arbitrary while multi-word phrases are systematic. How does this property, the systematic re-use of arbitrary elements, emerge in communication systems?

Recent experimental work has shown that people are able to successfully communicate in the absence of conventional communication systems, often by creating novel signs. (de Ruiter et al., 2007; Galantucci, 2005; Garrod et al., 2007; Healey et al., 2002; Scott-Phillips, 2009). The first signs people produce in these situations are often not arbitrary, but rather iconic or motivated in some other way. (Galantucci, 2005; Garrod et al., 2007)

Psycholinguistic work has demonstrated how referring expressions can change during dialogue. (Garrod & Doherty, 1994; Pickering & Garrod, 2004). In particular, conversational partners collaborate to establish definite references, and allowing their referring expressions to shorten. (Clark & Wilkes-Gibbs, 1986). This simplification

causes iconic signs in novel communication systems to become more arbitrary. (Garrod et al., 2007)

Kirby (2001) demonstrated how, given a set of arbitrary signs, systematicity might evolve. Simple artificial agents learn sets of signs and detect chance regularities in them (e.g., that the words for two red items both contain the syllable "ka"). Over many generations of agents producing signals for new meanings (meanings they didn't learn signals for) according to the regularities they observed, a set of signs can become systematic. Kirby et al. (2008) confirmed the result in human experimental participants.

Taking these two lines of research together, we have one route to the systematic re-use of arbitrary elements: people generate signs that are non-arbitrary, those signs become arbitrary as they simplify, by chance there are a few signal-meaning regularities, generations of people propagate these regularities, and the language becomes systematic. It's this longer history of a communication system, from the birth of the first sign to a set of signs which systematically re-uses arbitrary elements, that the current work aims to explore.

Goldin-Meadow et al. (1995)'s study on the emergence of systematicity in homesign (gestures created by deaf, non-signing children for use with their caretakers) covers this range. They found that, in the early stages of the homesign systems, a particular value of a particular gesture component (such as a 1" distance between the thumb and index finger) was used in gestures for just one object. In the later stages, the homesigners apparently collapsed some distinctions between objects and applied some values of gesture components to more than one object, increasing the systematicity of his or her set of gestures. This work shows that systematicity doesn't require complete arbitrariness – the recurrences between signal components and meaning components weren't chance. Unfortunately, we cannot know whether homesigners systems would have been systematic from the earliest stages, given similar-enough objects.

Here we present an exploration of the emergence of the systematic re-use of arbitrary elements in one controlled experiment. In this way, we can probe the relationship between systematicity and arbitrariness as communication systems develop.

## Experiment

### Methods

**Participants** 32 University of Edinburgh students participated in exchange for £12. All were native British English speakers. Participants who played together didn't know each other.



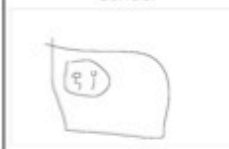













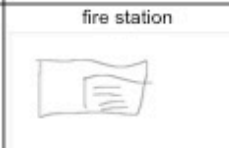
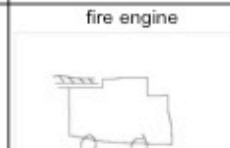



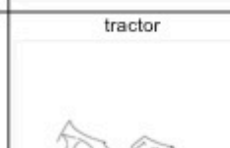


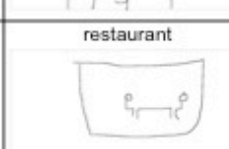

**Apparatus** Partners were seated in separate soundproof booths with computers. The experiment was run using the Pigeon software (Healey et al., 2002), which presented the item to draw each trial and provided a shared online whiteboard. Participants guessed and corrected their

partners' guesses in an MSN Messenger chat window.

**Stimuli** The items were chosen to share salient semantic features; each item can be thought of as one of five entity types (such as person or building) that relates to one of ten themes (such as education or agriculture). There were 26 core items, appearing with different frequencies. These are shown in Figure 1. Additionally, there were 14 filler items, occurring just once per game, intended to prevent participants from assuming that their set of items was closed. The items occurred in random order. Participants knew nothing about the items in advance. In particular, were never exposed to a list of the items.

teacher 	teaching 	school 	classroom 	school bus 
professor 	lecturing 	university 	lecture theatre 	
doctor 	medical emergency 	hospital 	operating room 	ambulance 
firefighter 	fire-fighting 	fire station 		fire engine 
farmer 	farming 	barn 		tractor 
chef 	cooking 	restaurant 	gourmet kitchen 	

**Figure 1.** One set of signs that emerged from the experiment. Each sign is the last occurrence of that item in the game. Signs are arranged according to the semantic features of the items, not by chronological order of the trials. Italics distinguish which participant was the Drawer that trial. The set is highly systematic, in that signs in many of the rows and columns share an element. Also notice how arbitrary the elements have become.

teacher 	teaching 	school 	classroom 	school bus 
professor 	lecturing 	university 	lecture theatre 	
doctor 	medical emergency 	hospital 	emergency room 	ambulance 
firefighter 	fire-fighting 	fire station 		fire engine 
farmer 	farming 	barn 		tractor 
chef 	cooking 	restaurant 	gourmet kitchen 	

**Figure 2.** One Mixed Last set of signs. Each drawing in a row or column is the last from a different game of the experiment (i.e. from different pairs of players). Notice how little systematicity the set has.

**Game** A team was allowed just one guess per trial. A team won 1 point for every correct guess but lost 1 point for any incorrect guess or drawing that included a symbol or convention. The goal was to win as many points as possible in the two hours of play. Participants from the three top-scoring teams were entered into a prize draw for an additional £20.

**Procedure** Each trial, one participant was the Drawer and other was the Guesser. The Drawer saw an item (such as *professor*) on his screen and was allowed to draw immediately. The Drawer drew with a mouse, had only black ink, and could not erase anything. The Guesser saw everything the Drawer drew immediately, on her screen. The Guesser did not see the Drawer's mouse movements when he was not drawing, and could not draw herself. When she was ready, the Guesser guessed by typing into a chat window. The Drawer stopped drawing immediately

and either confirmed or corrected the guess in the chat window. Players advanced themselves to the next trial. Every six trials, the participants switched Drawer and Guesser roles. The participants played for two hours.

### Results: Systematicity

Figure 1 shows one of the systems that emerged from this game. Notice how systematic it is: the drawings in many of the rows and columns share an element. For example, the drawings for items relating to university education (in the second row) each have a filled-in diamond. As another example, four of the drawings for activities/situations (in the second column) have rows of squiggly lines.

To enable analysis of systematicity, each set of drawings was printed on a page in a table, organized so that rows and columns contain drawings for similar items (as in Figure 1). A single coder examined each row and each column for any element shared among two or more drawings. If there was a

shared element, the coder marked which of the drawings in that row or column included it.

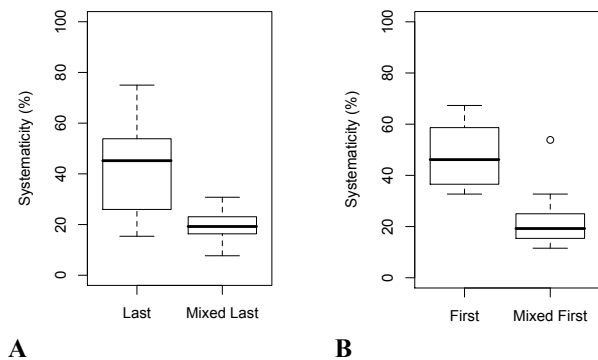
There are 26 drawings and each drawing is inspected twice – once as a member of its row and once as a member of its column. Thus, each set of drawings can receive a total score of 52. The total score divided by 52 is our systematicity score (a percentage).

For each of the 12 games from the experiment, we coded the set of First drawings (the first drawing of each item from that pair of participants) and the set of Last drawings (the last drawing of each item). To put their systematicity scores in context, we constructed 12 sets each of two kinds of comparison sets: Mixed First and Mixed Last. The Mixed First (or Last) sets were each composed of the First (or Last) drawings from *different* games of the experiment (i.e. from different pairs of players). For each Mixed set, for each item (e.g., *teacher*), we choose at random which of the games the drawing would be from, with the restriction that the drawings in each row and each column would be from different games. Figure 2 shows one Mixed Last set.

The coder marked these 48 sets in random order and blind.

A different coder marked three randomly chosen sets of each category independently. Her scores were strongly correlated with those of the original coder (Spearman’s  $\rho = 0.82$ ,  $p = 0.001$ ).

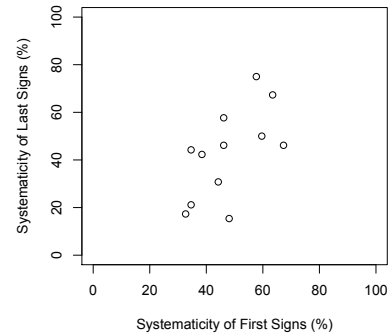
Figure 3A shows the mean systematicity for the Last and Mixed Last sets of signs. ( $M_{Last} = 42.79$ ,  $SD = 18.95$ ;  $M_{Mixed\ Last} = 19.39$ ,  $SD = 6.32$ ) Last sets of signs are more systematic than Mixed Last sets of signs. That is, signs drawn at the end of the games re-use elements more than can be attributed to the tendency across pairs of players to draw these items the same way (roughly, iconicity) – they are truly systematic. A Mann–Whitney U Test confirmed this ( $p < 0.005$ ).



**Figure 3.** Mean systematicity (%) and confidence intervals (confidence level = 95%) for First, Mixed First, Last, and Mixed Last sets of signs. Last sets are more systematic than Mixed Last. First sets are more systematic than Mixed First.

How did Last sets of signs get to be systematic? It turns out that First sets of signs are also truly systematic. Figure 2B shows the mean systematicity for First and Mixed First

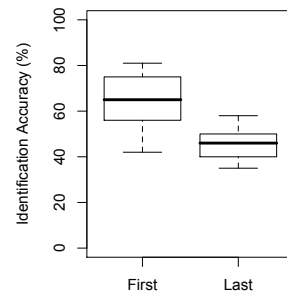
sets of signs. ( $M_{First} = 47.76$ ,  $SD = 11.85$ ;  $M_{Mixed\ First} = 22.28$ ,  $SD = 11.67$ ) A Mann–Whitney U Test confirmed that the First sets of signs have significantly higher systematicity scores than the Mixed First sets of signs ( $p < 0.001$ ). Further, as Figure 4 illustrates, there’s a strong correlation between the First and Last systematicity of the sets. (Spearman’s  $\rho = 0.62$ ,  $p < 0.05$ ).



**Figure 4.** Scatterplot of Last against First Systematicity (%). The ranking of the First sets of signs by systematicity correlates with that of the corresponding Last sets of signs.

**Results: Arbitrariness**

To measure the arbitrariness of the signs produced in the experiment, we followed Fay et al. (2008) and had new participants guess what they meant. 12 University of Edinburgh students, all native British English speakers, participated in exchange for one chance in a £25 prize draw for each correct guess. The experiment was run online, and lasted approximately 15 minutes. Participants learned about the original game, and that the drawings they’d see would be from different games and different points in the games, in random order. Each trial, a participant saw a screenshot of the whiteboard at the end of the trial in the original game. He guessed the meaning of it by clicking on one of 26 buttons, one for each possible item. Each participant was presented with the First drawings of each core item from one randomly-assigned original game and the Last drawings of each core item from a different randomly-assigned original game.



**Figure 5.** Mean identification accuracy (%) and confidence intervals (confidence level = 95%) for First and Last sets of signs. First sets of signs are more accurately identified than Last.

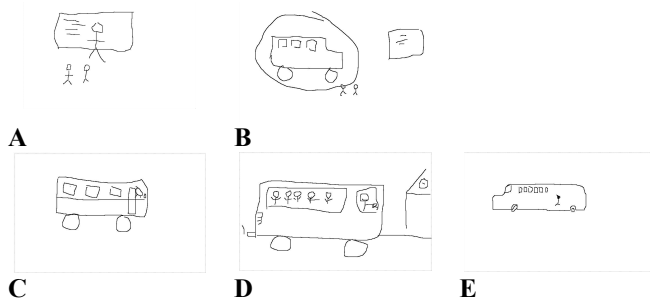
Figure 5 shows the mean identification rates (as proportions correct) for First and Last sets of signs. ( $M_{\text{First}} = 64.08$ ,  $SD = 12.37$ ;  $M_{\text{Last}} = 45.42$ ,  $SD = 6.86$ ) First sets of signs were more accurately identified than Last sets of signs. A Mann–Whitney U Test confirmed this ( $p < 0.001$ ). This suggests that the signs became more arbitrary over the course of the games.

### Discussion

We’ve presented an experiment in which the systematic re-use of arbitrary elements emerges. Last sets of signs are systematic, and becoming more arbitrary.

While previous work has explored the “evolution” of systematicity, this experiment has shown systematicity in the very first signs people use with each other. It appears to simply emerge, without explicit design on the part of the participants, as a natural part of dialogue.

Where does this initial systematicity come from? One might expect that the first time a player draws a certain item (say, `school bus`) with his partner, he draws it no differently than if he were drawing with a new partner. But if this were the case, the First sets would have been no more systematic than the Mixed First sets. Instead, when drawing items for the first time, players seem to have referenced previous drawings of related items. Consider Figure 6, in which one pair’s first drawing of `school bus`, which occurred *after* another primary education item (`teacher`) had been drawn, is contrasted with three pairs’ first drawings of `school bus`, each of which occurred *before* any other primary education item had been drawn. The former drawing for `school bus` includes elements found in the previous drawing of `teacher`, viz. the chalkboard and two children - elements not found in the other pairs’ first drawings for `school bus`.



**Figure 5.** One pair’s first drawings for `teacher` (A) and then `school bus` (B), contrasted with three other pairs’ first drawings for `school bus` (C – E).

Thus, the systematicity results presented here apply not just to iconic reference to tangible objects, but to communication in general.

A common (albeit often implicit) assumption in the literature is that a novel communication system will first become arbitrary and then develop systematicity. For example, Garrod et al. (2007) say they offer an account of

the “evolution of sets of icons into sets of symbols, and of sets of symbols into symbol systems.” In contrast, the current work suggests that proper systematicity need not wait for arbitrariness.

Similarly, the current work shows that, as sets of signs become more arbitrary, they don’t necessarily become less systematic. Structure can be retained while the elements become arbitrary. Garrod et al. (2007) suggested this, but didn’t explore systematicity directly.

We’ve presented a paradigm that allows one to explore arbitrariness and systematicity in one experiment. Future work should explore the many issues surrounding the interaction of the two properties, as well as the transmission to others of communication systems which make systematic re-use of arbitrary elements.

### Acknowledgments

We are particularly grateful to Nic Fay for formative discussions about the paradigm and data to play with, Nik Swoboda for help with and modifications to Pigeon, and the Language Evolution and Computation Research Unit for feedback at every stage and hours of piloting. This work was supported by the European Union Integrated Project JAST (Joint Action Science and Technology), grant FP6-IST2-003747.

### References

Clark, H. H. & Wilkes-Gibb, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1-39.

De Ruiter, J.P., Noordzij, M., Newman-Norlund, S., Hagoort, P., & Toni, I. (2008). On the origin of intentions. *Attention and Performance XXII: Sensorimotor foundation of higher cognition* (pp. 593–610). Oxford: Oxford University Press.

Fay N., Garrod S., Roberts L. (2008). The fitness and functionality of culturally evolved communication systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1509): 3553-61.

Galantucci, B. (2005). An Experimental Study of the Emergence of Human Communication Systems. *Cognitive Science: A Multidisciplinary Journal*, 29(5), 737-767.

Garrod, S., & Doherty, G. (1994). Conversation, Coordination and Convention - an Empirical-Investigation of How Groups Establish Linguistic Conventions. *Cognition*, 53(3), 181-215.

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of Representation: Where Might Graphical Symbol Systems Come From? *Cognitive Science*, 31(6), 961-987.

Goldin-Meadow, S. (1995). The resilience of combinatorial structure at the word level: morphology in self-styled gesture systems. *Cognition*, 56(3), 195-262.

Healey, G. T., Swoboda, N., & King, J. (2002). A tool for performing and analysing experiments on graphical communication. *People and computers XVI: Proceedings*

- of HCI2002: *The 16th British HCI Group Annual Conference* (pp. 55-68). London: Springer-Verlag.
- Healey, P. G. T., Garrod, S., Fay, N., Lee, J., & Oberlander, J. (2002). Interactional context in graphical communication. *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 441–446). Mahwah, NJ: Laurence Erlbaum Associates, Inc.
- Kirby, S. (2001). Spontaneous evolution of linguistic structure: An iterated learning model of the emergence of regularity and irregularity. *IEEE Journal of Evolutionary Computation*, 5, 102–110.
- Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *PNAS*, 105(31), 10681-10686.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2), 169-226.
- Scott-Phillips, T. (2009) *The Social Evolution of Pragmatic Behaviour*. PhD Thesis, University of Edinburgh.