How to Create Shared Symbols

Nicolas Fay[1], Bradley Walker[1], Nik Swoboda[2] & Simon Garrod[3]

[1]School of Psychological Science, University of Western Australia, 35 Stirling Highway, Crawley WA 6009, Australia

[2]Department of Artificial Intelligence, Universidad Politecnica de Madrid, 28040 Madrid, Spain

[3]Institute of Neuroscience and Psychology, University of Glasgow, Glasgow G12 8QB, United Kingdom

Running Head: How to Create Shared Symbols

Keywords: Interpersonal Communication, Interaction, Icon, Symbol, Observation, Observational Learning, Social Coordinative Learning, Cultural Evolution, Cumulative Cultural Evolution, Language Evolution

Corresponding author:

Nicolas Fay, School of Psychological Science, University of Western Australia

35 Stirling Highway, Crawley, WA 6009 Australia

Email: nicolas.fay@gmail.com; Tel: +61 (0)8 6488 2688; Fax: +61 (0)8 6488 1006

Word count (excluding title page, abstract and references): 9,332 words

Accepted for publication in *Cognitive Science* on 19 January 2018

Cognitive Science (2018) 1–30

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12600

Abstract

Human cognition and behaviour is dominated by symbol use. This paper examines the social learning strategies that give rise to symbolic communication. Experiment 1 contrasts an individual-level account, based on observational learning and cognitive bias, with an inter-individual account, based on social coordinative learning. Participants played a referential communication game in which they tried to communicate a range of recurring meanings to a partner by drawing, but without using their conventional language. Individual-level learning, via observation and cognitive bias, was sufficient to produce signs that became increasingly effective, efficient and shared over games. However, breaking a referential precedent eliminated these benefits. The most effective, most efficient and most shared signs arose when participants could directly interact with their partner, indicating that social coordinative learning is important to the creation of shared symbols. Experiment 2 investigated the contribution of two distinct aspects of social interaction: behaviour alignment and concurrent partner feedback. Each played a complementary role in the creation of shared symbols: behaviour alignment primarily drove communication effectiveness, and partner feedback primarily drove the efficiency of the evolved signs. In conclusion, inter-individual social coordinative learning is important to the evolution of effective, efficient and shared symbols.

1. Introduction

Humans are a symbolic species (Deacon, 1997). Human cognition and behaviour is dominated by symbol use, evident from our everyday use of numeric and linguistic systems. But where do these symbols come from? This question is presented by Harnad (1990) as the symbol grounding problem; how shared meanings can arise from arbitrary symbols in the absence of a pre-established shared symbol system. A solution to the symbol grounding problem was offered by Peirce (1931), who suggested that symbols evolved from motivated signs that share a non-arbitrary correspondence between the sign and its meaning, i.e., iconic signs that resemble their meaning (e.g., a portrait of van Gogh that brings the Dutch painter to mind), or indexical signs that share a natural association between the sign and its meaning (e.g., the smell of smoke is an index of fire). This paper examines the social learning strategies through which shared symbols might arise from motivated signs.

Human communication systems, such as language, are socially learned. We have a range of social learning strategies at our disposal, from individual-level strategies to more complex inter-individual strategies (Tomasello, Kruger, & Ratner, 1993). Social learning research has tended to focus on observational learning (an individual-level strategy), where an agent learns from observing the behaviour of a model (Bandura, 1977). To be successful, the agent must use perspective-taking to infer the observed model’s intentions. Experimental simulations of language evolution are often based on individual-level observation plus the cognitive biases that guide human inference (e.g., Kirby, Cornish, & Smith, 2008). Whereas individual-level strategies may be sufficient for simpler forms of social learning, inter-individual strategies may be important to more complex social learning (Morgan, Laland, & Harris, 2015). Social coordinative learning is an inter-individual strategy, as opposed to an individual-level strategy, because social learning arises when agents coordinate and integrate their perspectives. Contemporary theories of dialogue stress the importance of social coordinative processes to successful interpersonal communication (Clark, 1996; Pickering & Garrod, 2004).

Using an innovative experimental paradigm, Experiment 1 contrasts the contribution of observational learning (an individual-level strategy) with social coordinative learning (an inter-individual strategy) to the evolution of shared symbols. Experiment 1 demonstrates the importance of social coordinative learning to the evolution of effective, efficient and shared symbols. In Experiment 2 we identify two important aspects of social coordinative learning – behaviour alignment and concurrent partner feedback – and isolate the influence of each to examine their contribution to the evolution of shared symbols. The Experiment 2 results indicate that behaviour alignment improved communication success and concurrent partner feedback improved sign efficiency. Together, these complementary processes drove the interactive evolution of shared symbols.

We begin by reviewing the evidence suggesting that social interactive processes are important to the evolution of shared symbols. Next, we highlight some problems with the experimental paradigms used, and how they might limit the conclusions reached. We then explain the present experiments, report their findings, and discuss their significance.

1.1. Evidence that social interaction is important to the evolution of shared symbols

Naturalistic studies indicate that motivated signs are important to establishing shared sign-to-meaning mappings. For example, when sign language users lack a label for something they tend to use an iconic sign for it (Klima & Bellugi, 1979). However, communication systems tend not to remain iconic; whereas early sign languages and writing systems made extensive use of motivated signs, both have evolved in the direction of arbitrariness (Frishberg, 1975; Vaccari & Vaccari, 1961). Following Wescott (1971), we consider signs to lie on a continuum that ranges from absolutely motivated to absolutely arbitrary, with icons at one end and symbols at the other (with indices somewhere in-between) (see also Bronowski, 1967). We propose that social interaction is a key mechanism that drives the evolution of signs along this continuum, from (more) iconic to (more) symbolic.

Social interaction plays a key role in contemporary theories of dialogue (Clark, 1996; Pickering & Garrod, 2004). Actively participating in dialogue ensures that meanings are mutually agreed, or grounded, between pairs of interlocutors (Clark & Schaefer, 1987) and across laboratory 'generations' of interlocutors (Tan & Fay, 2011). Grounding is an opportunistic process, where interlocutors try to find commonalities that allow them to coordinate, or align, their perspectives. For example, if person B accepts person A's object description they can adopt that description, otherwise they can search out alternatives until a mutually acceptable description is identified. By contrast, if a passive observer does not understand person A's description the communication is likely to fail because there is no opportunity for the observer to negotiate a mutually acceptable alternative with person A.

Social interaction also plays an important role in experimental-semiotic simulations of sign evolution. Experimental-semiotic studies examine the creation of novel human communication systems under controlled laboratory conditions (for reviews see Fay, Ellison, & Garrod, 2014; Galantucci, 2017; Tamariz, 2017). They do this by using a paradigm in which human participants must communicate without using their existing shared language. Typically, participants communicate in a novel modality, for example, through drawing (Galantucci, 2005; Garrod, Fay, Lee, Oberlander, & MacLeod, 2007; Healy, Swoboda, Umata,

& King, 2007; Roberts, Lewandowski, & Galantucci, 2015) or by gesture (Christensen, Fusaroli, & Tylén, 2016; Fay, Arbib, & Garrod, 2013; Schouwstra & de Swart, 2014; Stolk, Verhagen, & Toni, 2016), and the experimenters study how the communication systems arise and evolve over repeated interactions between the participants.

A key finding is the importance of motivated signs and social interaction to the creation of shared symbols (Garrod et al., 2007). In Garrod et al. (2007), participants tried to communicate a set of recurring meanings to a partner by drawing on a shared whiteboard. Like the game Pictionary©, participants were not allowed to speak or use letters or numbers in their drawings. This procedure forced participants to create a novel communication system from scratch. When participants played the game with an interacting partner three things happened: their communication success improved as they repeatedly communicated the same meanings, the signs they used evolved from complex motivated signs to simpler, more symbolic signs, and over repeated interactions they increasingly used the same signs to communicate the same meanings (i.e., their behaviour aligned; see **Fig. 1**). This pattern, the creation of an effective inventory of shared symbols, has been widely replicated (Caldwell & Smith, 2012; Fay, Garrod, Roberts, & Swoboda, 2010; Garrod, Fay, Rogers, Walker, & Swoboda, 2010; Theisen, Oberlander, & Kirby, 2010). Analogous findings are observed in verbal referential communication experiments (Clark & Wilkes-Gibbs, 1986; Garrod & Anderson, 1987).

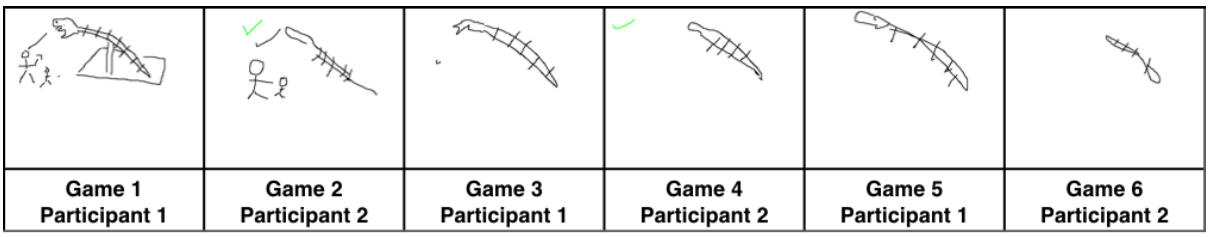

Fig. 1. Sign refinement and alignment for the meaning 'Museum' over 6-games between a pair of participants in the Interaction condition from Experiment 1 of the present study. Participants alternated directing and matching roles from game to game. At Game 1 Museum was communicated using a complex motivated sign that included a dinosaur, an exhibit space and two viewers. By Game 6 the sign has lost much of its initial motivation, evolving into a simpler, more symbolic representation, communicated by only the dinosaur's spine. In addition to this symbolization process, the interacting partners' signs became increasingly similar, or aligned, across games.

Experiments that manipulate the opportunity for interaction with a partner suggest that social coordinative processes are crucial to communication success and sign symbolization. Garrod et al. (2007) asked a group of passive observers to pick out the meaning associated with each of the signs produced by interacting pairs. Identification accuracy was lower among non-interacting observers compared to participants actively involved in the social coordinative process, indicating that social interaction was important to communication success. Furthermore, sign comprehension was lower among passive observers who were shown the later, simplified signs (games 4-6) compared to passive observers who were shown the more complex signs produced in the earlier games (see also Fay & Ellison, 2013; Fay, Garrod, & Roberts, 2008). This indicates that the signs became more abstract and symbolic, and their meaning became less accessible to observers, across repeated interactions. Note, the later signs were identified at higher than chance levels, indicating that they had not become absolute symbols, but they had become more symbolic, or less motivated, relative to the initial signs. Analogous results are returned by a verbal referential communication study (Schober & Clark, 1989). Furthermore, when interactive partner feedback was unavailable, participants' signs became more complex across repeated productions, as opposed to simpler and more symbolic. This pattern is seen in experimental-semiotic studies (Garrod et al., 2007, 2010) and verbal referential communication studies (Hupet & Chantraine, 1992; Krauss & Weinheimer, 1966).

1.2. Experimental paradigms limit the conclusions that can be drawn

The experimental studies reviewed indicate that social coordinative learning enhances communication success and sign symbolization. However, the evidence is inconclusive. This is because the interactive and the non-interactive conditions they are contrasted with are not comparable. The interactive conditions engaged (repeated) production and (repeated) comprehension processes, whereas the non-interactive conditions engaged either (repeated) production (e.g., Garrod et al., 2007) or repeated comprehension processes (e.g., Schober & Clark, 1989), but not both processes. It is therefore unclear if the benefits of social coordinative learning arise because social interaction engages production and comprehension processes, or because of the opportunity it affords for interactive grounding. An additional confound is that participants in the non-interactive conditions may have been less attentive, compared to interacting participants, given that they were not required to respond to the communicator.

These problems open the door to individual-level explanations of the observed phenomena. A simplicity bias captures the systematic preference to choose the simplest solution to a problem (Chater & Vitányi, 2003). It follows that a simplicity bias may drive sign simplification and, therefore, explain how the initially motivated signs became increasingly arbitrary and symbolic over repeated use. This individual-level explanation is consistent with a principle of least effort (Piantadosi, Tily, & Gibson, 2011; Zipf, 1949).

Without contrasting an interactive condition against a comparable non-interactive condition, it is unclear if sign symbolization arises through social coordinative learning or through observational learning that is guided by a simplicity bias.

Other research indicates that behaviour alignment can occur in the absence of direct social interaction. Verbal referential communication studies, in which participants describe events pictured on cards, show that interlocutors align their lexical choices and syntax, and this occurs with or without direct social interaction with a partner (Branigan, Pickering, McLean, & Cleland, 2007; Branigan, Pickering, Pearson, McLean, & Brown, 2011). These findings suggest that a cognitive bias toward behaviour alignment may be sufficient for the creation of a shared inventory of sign-to-meaning mappings. However, lexical priming can only occur when participants already share a lexicon, just as syntactic priming can only occur when participants already share a grammar. This is not the case in experimental-semiotic studies, where participants are tasked with creating a shared inventory of signs and combinatorial rules from scratch. Under these circumstances, social coordinative processes may be important to referential alignment.

Experiment 1 demonstrates the importance of social interaction to the creation of shared symbols. It does this by isolating the role of social coordinative learning from the role of observational learning and cognitive bias.

2. Experiment 1. How to create shared symbols: Social Interaction, observation and cognitive bias

Experiment 1 tests the contribution of social interaction to three outcomes: communication success (or cognitive alignment), sign symbolization (operationalized using an information theoretic measure of sign complexity) and behaviour alignment (human ratings of the extent to which interlocutors used the same signs to communicate the same meanings). These outcomes are important because any functional communication system should be effective, efficient and shared. Interacting pairs of participants were compared against participants allocated to a 'Pseudo-Interaction' condition that eliminated the opportunity for social coordinative learning.

The Interaction condition is similar to that used in other experimental-semiotic studies (Caldwell & Smith, 2012; Fay et al., 2010; Garrod et al., 2007, 2010; Theisen et al., 2010). Participants communicated by drawing a range of experimenter-specified meanings to a co-present partner across a virtual whiteboard tool (Healy, Swoboda, & King, 2002). Their partner tried to identify the intended meaning from a list of competitors, but could also interact graphically by drawing on the virtual whiteboard. Participants alternated directing and matching roles from game to game (1-6).

Two Pseudo-Interaction conditions were created that included (repeated) production and (repeated) comprehension processes but eliminated social interaction. In each Pseudo-Interaction condition participants believed they were directly interacting with a co-present partner, but they were not. Instead, the drawings produced by participants in the Interaction condition were played back to them across the virtual whiteboard tool. When it was their turn to communicate each meaning they were told their drawings would be sent to their partner, but they were not. Thus, in the Pseudo-Interaction conditions participants could be influenced by their partner but could not influence their partner, i.e., communication was one-way as opposed to two-way (as was the case in the Interaction condition).

Two Pseudo-Interaction conditions were tested: Pseudo-Interaction: Precedent and Pseudo-Interaction: Broken Precedent. In the Pseudo-Interaction: Precedent condition, at Game 1 participants tried to identify the meaning associated with each of the drawings produced by their partner (a participant from the Interaction condition). In this condition their partner set the referential precedent by producing the first drawing for each meaning at Game 1. In the Pseudo-Interaction: Broken Precedent condition the participant drew first (at Game 1) and therefore set the referential precedent. Because there are a variety of ways that participants can communicate the different meanings (**Fig. 7** illustrates four different ways that participants communicated the meaning 'Parliament'), it is likely that the referential precedent set by the participant in this condition will be broken by their partner. Because referential precedents (or conceptual pacts; Brennan & Clark, 1996) set an expectation that a particular sign will be consistently used to pick out a particular meaning, they reduce uncertainty and aid partner comprehension (Keysar & Barr, 2002; Kronmüller & Barr, 2015; see also Relevance Theory; Sperber & Wilson, 1987). So, breaking a referential precedent is likely to negatively impact interpersonal communication.

Comparing the Interaction condition to the Pseudo-Interaction conditions allowed us to determine the contribution of social coordinative processes above and beyond the contribution of observational learning and cognitive biases. In the context of the present study, a simplicity bias may be sufficient to drive sign symbolization, and an alignment bias may be sufficient for interlocutors to create a shared inventory of sign-to-meaning mappings. However, if communication success, sign symbolization and sign alignment are stronger in the Interaction condition this would support the view that inter-individual social coordinative learning is important to the creation of shared symbols.

We predict that communication success, sign efficiency and sign alignment will be lower in the Pseudo-Interaction: Broken Precedent condition compared to the Pseudo-Interaction: Precedent condition. Our key prediction is that social interaction is important to each of these outcomes. If correct, communication success, sign efficiency and sign alignment will be highest in the Interaction condition compared to the Pseudo-Interaction conditions.

3. Method

Experiments 1 and 2 received approval from the University of Western Australia Ethics Committee. Participants viewed an information sheet before giving written consent to take part in the study. The information sheet and consent form were both approved by the Ethics Committee. All methods were performed in accordance with the guidelines from the NHMRC/ARC/University Australia's National Statement on Ethical Conduct in Human Research.

3.1. Participants

A convenience sample of sixty undergraduate students (42 self-reported females and 18 self-reported males) participated in exchange for partial course credit or payment (A$10). The sample size was based upon prior studies using the same experimental paradigm (Fay et al., 2010; Garrod et al., 2007, 2010). No statistical analyses were run prior to collecting the full sample. Participants were tested in unacquainted pairs, or individually, in testing sessions lasting up to 1-hour. All participants reported being free of any uncorrected visual impairment.

3.2. Task and procedure

The goal for each participant was to graphically communicate 16 confusable meanings (e.g., ‘Arnold Schwarzenegger’, ‘Brad Pitt’, ‘Russell Crowe’) in such a way that their partner could identify their intended meaning. Like the game Pictionary©, participants were prohibited from using letters or numbers in their drawings. A review of the drawings produced by participants indicated they had followed the experimental instructions. The Director drew each meaning from their ordered list (16 targets plus 4 distractors; see **Table 1** for a complete listing) and their partner, the Matcher, tried to identify each meaning from their randomly ordered list of the same meanings.

The task was administered using a virtual whiteboard tool (Healy et al., 2002), which recorded all drawing activity. This tool has been used in a range of graphical communication studies (Fay et al., 2010; Garrod et al., 2010; Healy et al., 2007; Theisen et al., 2010). Each participant sat at a computer terminal where drawing input and meaning selection was made via a standard mouse. For the Director, each to-be-depicted meaning was highlighted in white text on a dark background at the top of the interface. Holding down the left mouse button initiated drawing. Director drawing was restricted to black ink and Matcher drawing was restricted to green ink (to distinguish between participants). By clicking an erase button on the interface participants were able to erase parts of their own drawing and their partner’s drawing. All drawing and erasing activity was displayed simultaneously on the Director and Matcher’s shared virtual whiteboards. When the matcher believed they had identified the director’s intended meaning they clicked the relevant button at the top of their interface, where there was a list of buttons corresponding to the competing meanings. Meaning selection brought the current trial to an end and initiated the next trial. No time limit was imposed, and participants were given no explicit feedback with regard to their communication success. Having participants communicate remotely across networked computers meant they were unaware of their partner's identity.

3.3. Conditions

Participants were randomly allocated to one of three conditions: Interaction (N= 30, or 15 interacting dyads), Pseudo-Interaction: Precedent (N= 15 individuals) or Pseudo-Interaction: Broken Precedent (N= 15 individuals). In the Interaction condition pairs of participants played 6 consecutive games of the task with the same partner, using the same meaning set on each game. For the Director, the first 16 meanings were always the target meanings (presented in a different random order on each game). The final 4 meanings were always the distractor meanings (presented in a different random order on each game). The 4 distractor meanings were the same on each game and for each pair of participants. The distractor meanings were never communicated. Distractor meanings were included to ensure that Matchers could not use a process of elimination to identify the final target meaning. However, over the course of the experiment participants may have realized the distractor meanings were never communicated, and may have used a process of elimination to identify the final target meaning on the later games. For the Matchers, all 20 meanings were presented in a different random order on each game. In the Interaction condition participants alternated between directing and matching roles from game to game (i.e., Participant 1 was the Director on games 1, 3 and 5 and the Matcher on games 2, 4 and 6, and Participant 2 was the Director on games 2, 4 and 6 and the Matcher on games 1, 3 and 5). Irrespective of directing or matching role, participants were able to graphically interact within a trial. Thus, a Matcher might provide graphical feedback to the Director by annotating part of their drawing or by offering a graphical alternative. This occurred on 11.60% of trials (23.33%, 11.25%, 13.75%, 8.33%, 7.50%, 5.42% of trials at game 1 to game 6 respectively).

The drawings produced by participants in the Interaction condition seeded the Pseudo-Interaction conditions. The Pseudo-Interaction conditions provided participants with exactly the same informational experience as participants in the Interaction condition, but without the opportunity for social interaction. Participants in each Pseudo-Interaction condition were told they would observe the interaction between two people playing the Pictionary-type task (the Director and Matcher from the Interaction condition) and would interact with one of them (the Director) when it was their turn to communicate the target meanings. Because the virtual whiteboard tool (Healy et al., 2002) makes pixel-by-pixel recordings of participants' drawings, we were able to dynamically play back the drawings from the Interaction condition to participants in the Pseudo-Interaction conditions exactly as they were produced.

Table 1. The set of meanings that Directors communicated to Matchers (distractor meanings given in italic). Target and distractor meanings were fixed across conditions and throughout the experiment.

Places	**People**	**Entertainment**	**Objects**	**Abstract**
Art Gallery	Arnold Schwarzenegger	Cartoon	Computer Monitor	Homesick
Parliament	Brad Pitt	Drama	Microwave	Loud
Museum	*Hugh Grant*	*Sci-Fi*	*Refrigerator*	Poverty
Theatre	Russell Crowe	Soap Opera	Television	*Sadness*

In the Pseudo-Interaction: Precedent condition the participant acted as the Matcher on games 1, 3 and 5 and the Director on games 2, 4 and 6 (see **Fig. 2**). In this condition, participants received the drawings produced by Participant 1 (Director trials) from the Interaction condition, plus any associated Matcher feedback. Matcher feedback was included in the playback from the Interaction condition to ensure that participants in the Pseudo-Interaction conditions received the same information as the Matchers in the Interaction condition. Participant 1 from each interacting dyad (15 in total) seeded a Pseudo-Interaction: Precedent participant.

In the Pseudo-Interaction: Broken Precedent condition the participant acted as the Director on games 1, 3 and 5 and the Matcher on games 2, 4 and 6 (see **Fig. 2**). In this condition, participants received the drawings produced by Participant 2 (Director trials) from the Interaction condition, plus any associated Matcher feedback. Participant 2 from each interacting dyad seeded a Pseudo-Interaction: Broken Precedent participant. While observing the drawing playback from their partner, participants in the Pseudo-Interaction conditions were not permitted to produce graphical feedback. Whereas the drawing activity of Directors in the Interaction condition ended when the Matcher selected a meaning, the drawing activity of Directors in the Pseudo-Interaction conditions ended when they clicked a send button. They were told that doing so sent their drawing to their partner, who would then try to pick out their intended meaning.

Unlike the non-interactive conditions of prior studies, that contained either repeated production-only processes (Garrod et al., 2007; Hupet & Chantraine, 1992; Krauss & Weinheimer, 1964) or repeated comprehension-only processes (Garrod et al., 2007; Schober & Clark, 1989), the Pseudo-Interaction conditions in the present study included both processes. Like the Interaction condition, communication in the Pseudo-Interaction conditions involved regularly interchanging between production and comprehension processes (see **Fig. 2**).

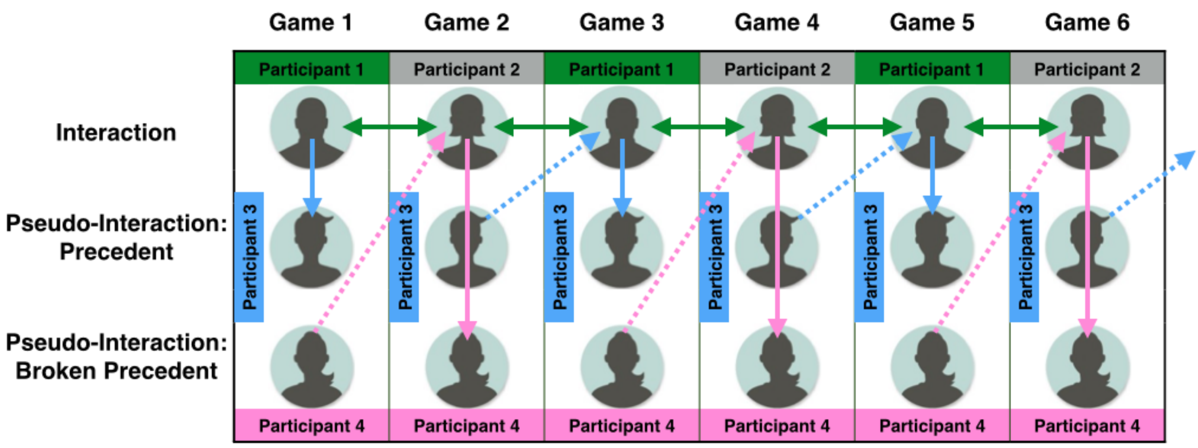

Fig. 2. Experiment 1 design. Pairs of participants in the Interaction condition took turns directing and matching across Games 1-6. In this condition both participants (Director, Matcher) could communicate during a trial, hence the bidirectional green arrows. Pseudo-Interaction: Precedent participants tried to identify the drawings produced by Participant 1 (Interaction condition) at Games 1, 3 and 5 (solid blue arrow) and communicated each meaning by drawing at Games 2, 4 and 6 (dashed blue arrow). Pseudo-Interaction: Broken Precedent participants drew each meaning for Participant 2 (Interaction condition) at Games 1, 3 and 5 (dashed pink arrow) and tried to identify the drawings produced by Participant 2 at Games 2, 4 and 6 (solid pink arrow). In the Pseudo-Interaction conditions only the Director could communicate during a trial, hence the unidirectional arrows.

3.4. Measures

Communication Success was measured by determining if the Matcher correctly identified the Director's intended meaning on each trial. Correct guesses were given a score of 1 and incorrect guesses a score of 0.

Sign Symbolization. Following Garrod et al. (2007), less complex signs were considered to be more symbolic. Sign complexity was measured using Pelli et al.'s (2006) information theoretic measure of perimetric complexity [Perimetric complexity = (inside + outside perimeter)^2/ink area]. Previous work indicates this to be an effective scale-free measure of drawing complexity (Fay et al., 2010; Garrod et al., 2007; Tamariz & Kirby, 2014).

Behaviour Alignment. To measure behaviour alignment, pairs of drawings from each dyad (at Game 1-2, 2-3, 3-4, 4-5 or 5-6) were presented side-by-side on a computer screen and were rated for similarity (author BW). The drawings were rated on a Likert scale from 0-9, where 0= very dissimilar and 9= very similar. In total 3600 pairs of drawings were rated for similarity (16 meanings x 5 pairs of adjacent games x 15 dyads x 3 conditions). A subset of drawings (240 pairs of drawings; 80 randomly sampled from each condition) were rated for similarity by a second judge (author NF). The raters were blind to the condition the drawings were sampled from. Comparison of the two sets of ratings showed strong inter-coder agreement (r= .834, p< .001).

4. Experiment 1 results

The data was analysed using logistic and linear mixed effects modelling, with crossed random effects for dyads and for items. All the analyses were performed and all the figures were created in R (R Core Team, 2013). Statistical models were estimated using the glmer() and lmer() function of lme4 (Bates, Maechler, Bolker, & Walker, 2013). We tested all effects using model comparison, comparing models with identical random effects, but with the fixed effect(s) of interest removed from one of the models. The maximal random effects structure justified by the experiment design was specified where possible (Barr, Levy, Scheepers, & Tily, 2013).

4.1. Communication success

We first compared communication success in the Interaction condition to the Pseudo-Interaction: Precedent condition at Games 1, 3 and 5. The data was analyzed using a logistic mixed effects model. Condition and Game were entered as fixed effects with interaction. Both fixed effects were centered. The random effects structure included by-Dyad and by-Item random intercepts, as well as by-Item random slopes for Condition. This was the maximal random effects structure that would converge. The best fitting model specified Condition and Game as fixed effects with interaction ($\chi^2(1) = 12.751,\ p < .001$). In both conditions communication success improved over games, but the improvement was stronger in the Interaction condition (β= 1.196, *SE*= 0.281, $\chi^2(1) = 18.789,\ p < .001$) compared to the Pseudo-Interaction: Precedent condition (β= 0.413, *SE*= 0.151, $\chi^2(1) = 8.962,\ p = .003$).

Next, the Interaction condition was compared to the Pseudo-Interaction: Broken Precedent Condition at Games 2, 4 and 6. Again, the best fitting model specified Condition and Game with interaction ($\chi^2(1) = 17.846,\ p < .001$). Whereas communication success improved across games in the Interaction condition (β= 1.827, *SE*= 0.579, $\chi^2(1) = 16.612$, $p < .001$), there was no statistical evidence of an improvement in communication success across games in the Pseudo-Interaction: Broken Precedent condition (β= 0.174, *SE*= 0.143, $\chi^2(1) = 1.489,\ p = .222$). See **Fig. 3** for data visualisation.

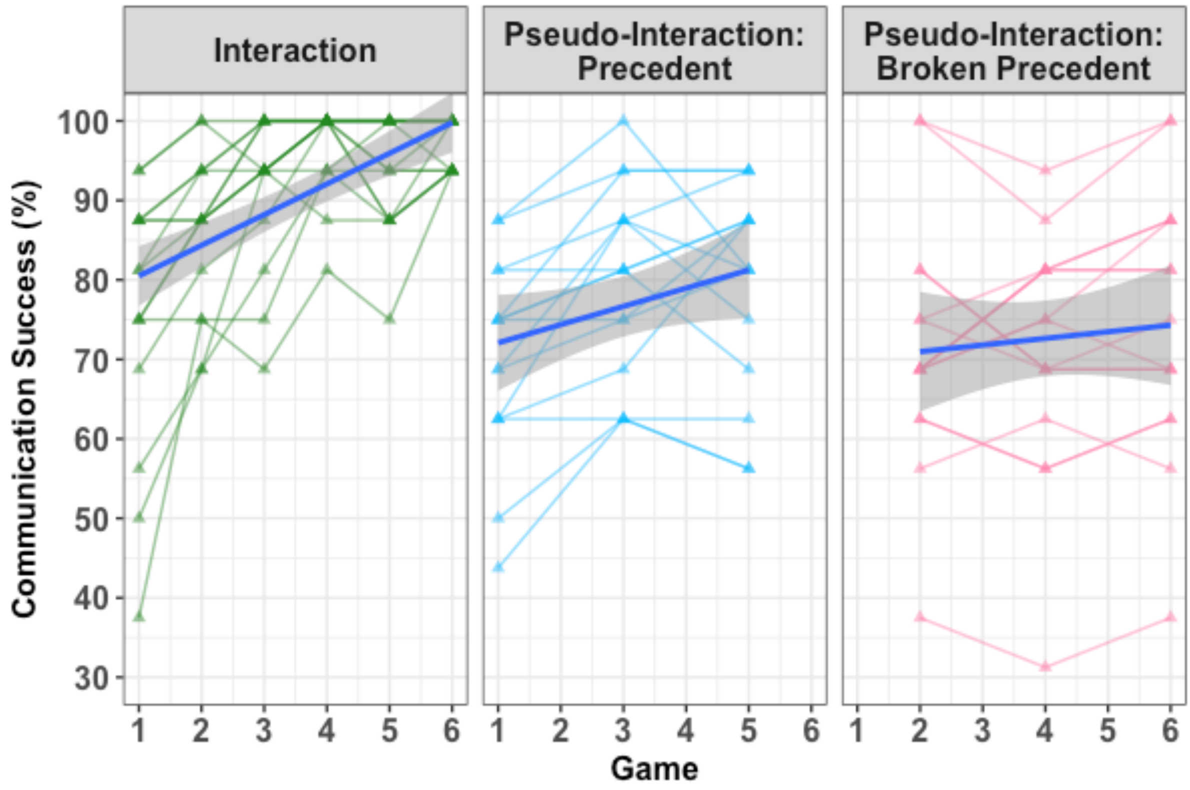

Fig. 3. Change in communication success (plotted for each dyad) across Games 1-6 in the Interaction condition and each of the Pseudo-Interaction conditions. The blue straight line is the linear model fit and the grey shaded area is the 95% confidence interval.

4.2. Sign symbolization

Examples of sign symbolization and sign alignment from the different experimental conditions are given in **Fig. 4**. Strong sign symbolization is observed in the Interaction condition. At game 1 the sign used to communicate 'Museum' is highly motivated; it uses icons (of a dinosaur, an exhibit space and two viewers) that structure-map (i.e., visually resemble) features of the objects that are typically seen in a museum. By Game 6 a much simpler sign is used, where the structure-mapping between the sign and its meaning is mostly absent; only the dinosaur's spine is retained from the earlier game 1 sign. At game 6 the mapping between the sign and its meaning has become more arbitrary, and therefore more symbolic. Strong behaviour alignment is also observed in this condition: over games, members of the interacting dyad increasingly used the same sign to communicate the same meaning. Sign symbolization and sign alignment are observed in the Pseudo-Interaction: Precedent condition, but they are weaker compared to the Interaction condition. By contrast, sign symbolization is minimal in the Pseudo-Interaction: Broken Precedent condition, and sign alignment is absent.

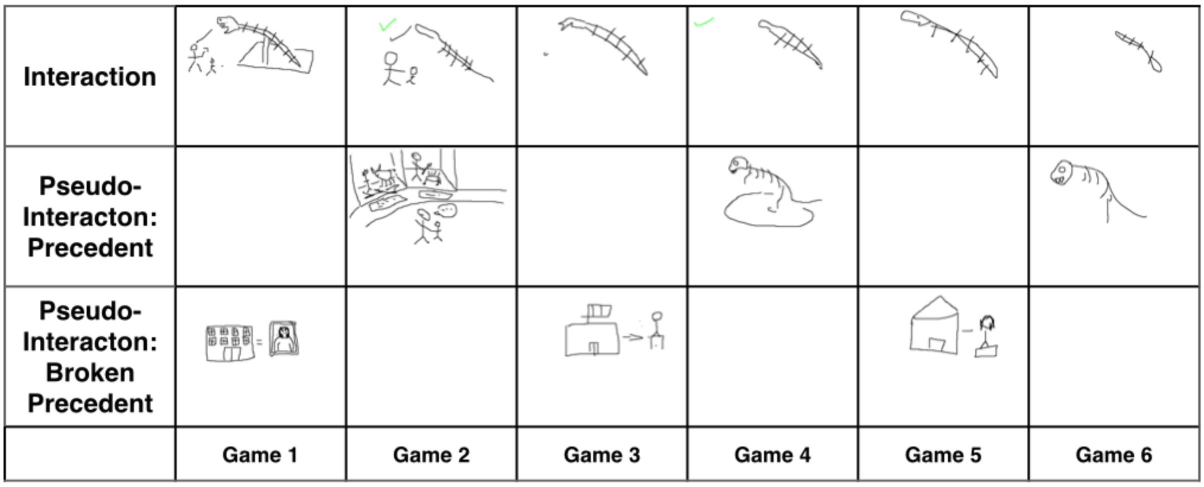

Fig. 4. Example drawings of the meaning 'Museum' from the different experimental conditions across Game 1-6 from Experiment 1.

First, sign complexity in the Interaction condition was compared to the Pseudo-Interaction: Precedent condition at Games 2, 4 and 6. The data was analyzed using a linear mixed effects model. Condition and Game were entered as fixed effects with interaction.

Both fixed effects were centered. The maximal random effects structure was specified. This included by-Dyad and by-Item random intercepts, as well as by-Dyad random slopes for Game and by-Item random slopes for the Condition by Game interaction. The best fitting model specified Condition and Game as fixed effects without interaction ($\chi^2(1) = 0.245$, $p = .621$). Sign complexity decreased over games in both conditions (β= -562.03, *SE*= 93.05, $\chi^2(1) = 23.433,\ p < .001$), but overall sign complexity was lower in the Interaction condition compared to the Pseudo-Interaction: Precedent condition (β= 2505.94, *SE*= 292.02, $\chi^2(1) = 37.426,\ p < .001$).

Next, sign complexity in the Interaction condition was compared to the Pseudo-Interaction: Broken Precedent condition at Games 1, 3 and 5 (same model). Here, the best fitting model specified Condition and Game as fixed effects with interaction ($\chi^2(1) = 20.023,\ p < .001$). Whereas sign complexity decreased across games in the Interaction condition (β= -1007.76, *SE*= 121.39, $\chi^2(1) = 29.448,\ p < .001$), there was no statistical evidence of a decrease in sign complexity in the Pseudo-Interaction: Broken Precedent condition (β= -67.92, *SE*= 131.93, $\chi^2(1) = 0.263,\ p = .608$). See **Fig. 5** for data visualisation.

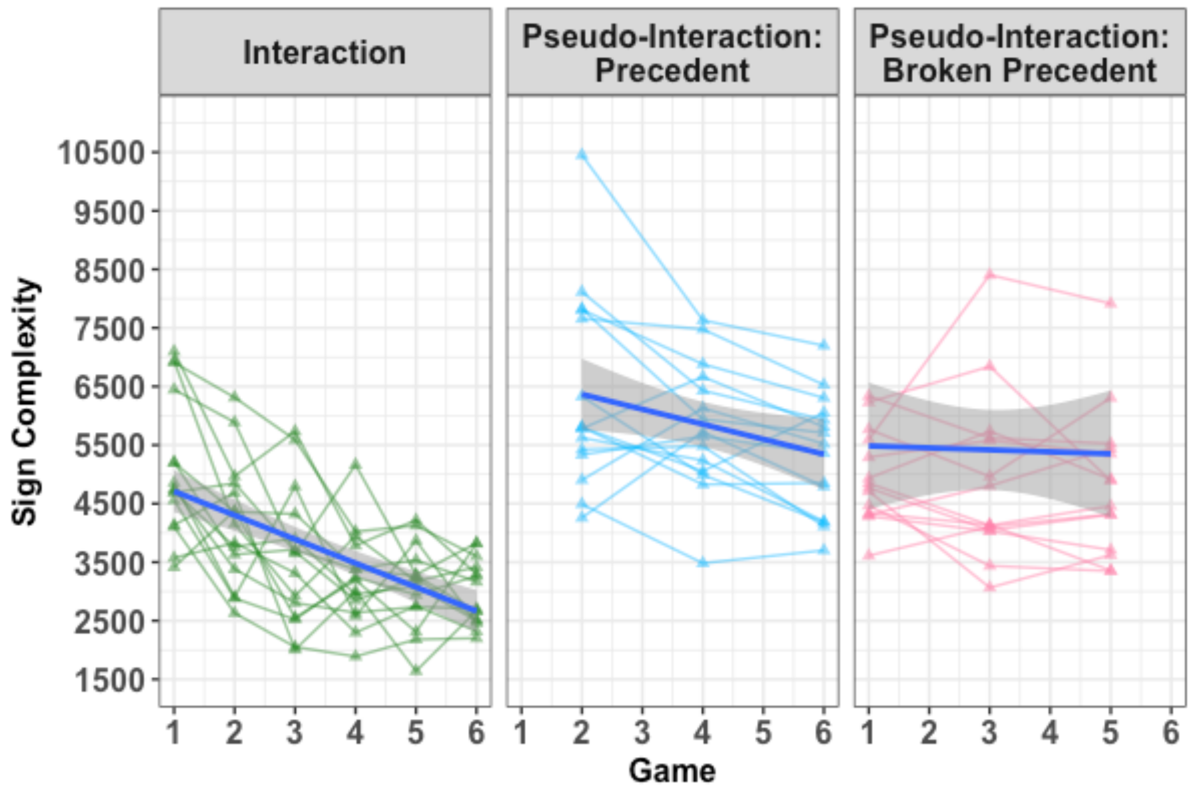

Fig. 5. Change in perimetric complexity of the signs (plotted for each dyad) across Games 1-6 in the Interaction condition and each of the Pseudo-Interaction conditions. The blue straight line is the linear model fit and the grey shaded area is the 95% confidence interval.

4.3. Behaviour alignment

The final analysis of Experiment 1 compared the change in behaviour alignment (operationalized as the extent to which drawings of the same experimental meaning became similar, i.e., sign alignment scores) over games in the different conditions. The sign similarity data was analyzed using a linear mixed effects model (same model used in sign complexity analysis, but with Condition factor coded). The best fitting model specified Condition and Game as fixed effects with interaction $(\chi^2(1) = 22.365, p < .001)$. The interaction effect is explained by the stronger increase in sign alignment scores across games in the Interaction condition (*β*= 0.455, *SE*= 0.043) compared to the Pseudo-

Interaction: Precedent condition (β= 0.071, *SE*= 0.073) and the Pseudo-Interaction: Broken Precedent condition (β= 0.121, *SE*= 0.043). Comparison of the Pseudo-Interaction conditions indicated that the sign alignment scores increased over games ($\chi^2(1) = 4.131, p = .042$), and that overall sign alignment was higher in the Pseudo-Interaction Precedent condition compared to the Pseudo-Interaction: Broken Precedent condition ($\chi^2(1) = 8.869,\ p = .002$). Note that sign alignment scores in both Pseudo-Interaction conditions are lower than neutral alignment, indicating that participants tended not to align their behaviour (see **Fig. 6** for data visualisation).

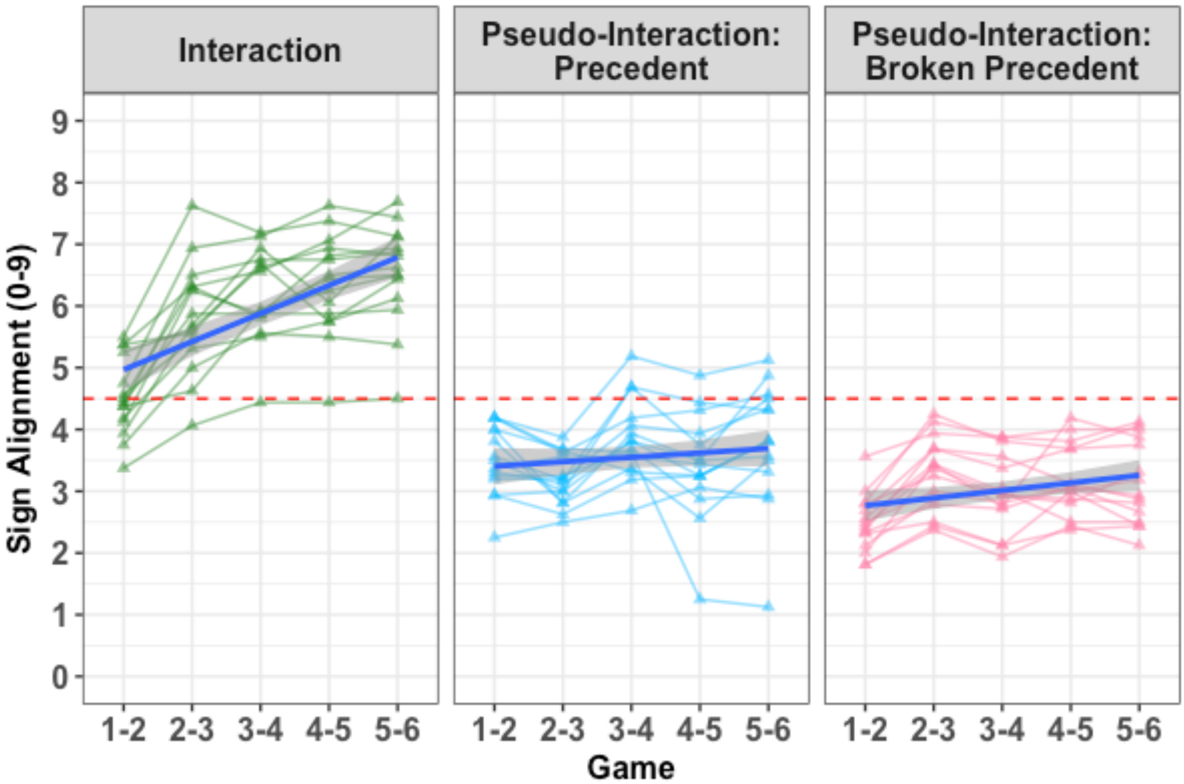

Fig. 6. Change in rated sign alignment (plotted for each dyad) over Games in the Interaction condition and each of the Pseudo-Interaction conditions. The horizontal dashed red line indicates neutral sign alignment. The blue straight line is the linear model fit and the grey shaded area is the 95% confidence interval.

5. Experiment 1 discussion

Experiment 1 tested the contribution of social interaction to the creation of shared symbols. Social interaction proved to be important to communication success. Participants in the Interaction condition showed the strongest improvement in communication success over games, replicating previous studies (Fay et al., 2010; Garrod et al., 2007, 2010; Schober & Clark, 1989). Crucially, the experimental paradigm ruled out alternative explanations of the enhanced communication success in the Interaction condition, such as the absence of interchanging production and comprehension processes in the non-interactive conditions, or lower attention among participants who passively observed the communication of active interlocutors (Garrod et al., 2007; Schober & Clark, 1989). Communication success improved over games among participants in the Pseudo-Interaction: Precedent condition. This indicates that observational learning is sufficient to improve communication success. In the Pseudo-Interaction: Broken Precedent condition there was no statistical evidence of an improvement in communication success. This highlights the importance of referential precedents to interpersonal communication.

The simplest, and most symbolic signs were produced by participants in the Interaction condition. This finding replicates previous studies (Fay et al., 2010; Garrod et al., 2007, 2010; Hupet & Chantraine, 1992; Krauss & Weinheimer, 1964), and supports an inter-individual principle of least collaborative effort (Clark, 1996; Clark & Wilkes-Gibbs, 1986). Sign simplification was observed in the Pseudo-Interaction: Precedent condition, indicating a role for a simplification bias in sign symbolization, and supporting the individual-level principle of least effort (Piantadosi et al., 2011; Zipf, 1949). In the Pseudo-Interaction:

Broken Precedent condition there was no evidence of a change in sign complexity. This again highlights the importance of referential precedents to interpersonal communication and demonstrates their effect on the expression of a simplification bias on sign symbolization.

Social interaction proved important to behaviour alignment. Participants in the Interaction condition showed a strong increase in their sign alignment scores across games, replicating previous studies (Branigan, Pickering, & Cleland, 2000; Fay et al., 2010; Garrod & Anderson, 1987; Garrod et al., 2007, 2010). Sign alignment was observed in the Pseudo-Interaction conditions, but was weaker compared to the Interaction condition. Consistent with prior studies (Branigan et al., 2007, 2011), and an alignment bias, sign alignment did not require social interaction, but it was stronger with it. In the Pseudo-Interaction: Broken Precedent condition sign alignment was lowest. This again highlights the importance of referential precedents to interpersonal communication, and demonstrates their effect on the expression of a behaviour alignment bias.

The Experiment 1 results supported our predictions. The most effective, most efficient and most shared communication systems were produced by participants in the Interaction condition. The results of the Pseudo-Interaction: Precedent condition suggest that observation and cognitive bias contributed to the evolution of shared symbols. In this condition communication success improved, the signs became more efficient and more aligned over games. This occurred despite the participants not being able to directly interact with their partner. However, breaking a referential precedent eliminated these effects (Pseudo-Interaction: Broken Precedent condition). When a referential precedent was broken, communication success did not improve, the signs did not become more efficient, and sign alignment was much lower. Participants in this condition may have interpreted their partner's behaviour as uncooperative, and this may have reduced their motivation to align their behaviour, indicating that inferential processes were guiding communication behaviour.

Taken together, the Experiment 1 results demonstrate that inter-individual social coordinative learning is important to the creation of shared symbols.

6. Experiment 2. How to create shared symbols: The complimentary roles of behaviour alignment and concurrent partner feedback

Experiment 1 demonstrated that inter-individual social-coordinative learning is important to the creation of shared symbols. Experiment 2 isolated two important aspects of social interaction – behaviour alignment and concurrent partner feedback – and investigated the contribution of each to the evolution of effective and efficient human communication systems.

Pickering and Garrod (2004) argue that interlocutors cognitively align by aligning their linguistic behaviour and this underlies successful communication (see Fusaroli & Tylén, 2016 for a discussion of other factors that influence successful interpersonal communication). While a correlation between referential alignment and cognitive alignment has been observed (Fay, Lister, Ellison, & Goldin-Meadow, 2014; Fusaroli et al., 2012; Reitter & Moore, 2014), the causal role of referential alignment on cognitive alignment is unclear. If referential alignment directly influences cognitive alignment, prohibiting interacting participants from imitating their partner's signs and aligning their behaviour will lower communication success. This was tested in Experiment 2.

Whereas referential alignment occurs across interaction episodes (i.e., as partners alternate directing and matching roles), concurrent partner feedback occurs within an interaction episode. Concurrent partner feedback can take a variety of forms. During conversation, listeners are co-narrators who provide verbal feedback (e.g., saying "mhm" while listening to a speaker) and visual feedback (e.g., frowning or nodding), that improves the flow of conversation (Bavelas, Coates, & Johnson, 2000; Clark & Krych, 2004; Mein, Fay, & Page, 2016). Like listeners in a conversation, Matchers in the present study can signal their attention and understanding by commenting on the Director's drawing, e.g., by adding a tick mark (see **Fig. 1**). During conversation listeners can indicate a communication breakdown and initiate a repair (e.g., by asking the speaker for clarification; Dingemanse et al., 2015; Schegloff, 2000). A similar repair mechanism was observed in Experiment 1 when a Matcher circled a part of the Director's drawing to request clarification. In addition to these information expansion requests, Matcher feedback can drive information contraction by bringing the trial to an end before the Director has completed their drawing (by clicking a meaning button, the equivalent of an interruption during conversation). So, we predict that Matcher feedback will contribute to communication success and to sign symbolization. This was tested in Experiment 2.

In addition to examining the effects of behaviour alignment and concurrent partner feedback, Experiment 2 also examined the interplay between behaviour alignment and concurrent partner feedback on communication success and sign symbolization to determine if they operate independently or if they interact. This was done by experimentally manipulating the opportunity for participants to imitate the signs produced by their partner, and the opportunity for participants to receive concurrent partner feedback, in a full factorial design. We then examined the effect of this on communication success and sign symbolization.

7. Method

7.1. Participants

A convenience sample of 120 undergraduate students (84 self-reported females and 36 self-reported males) participated in exchange for partial course credit or payment (A$10). The sample size was based upon prior studies using the same experimental paradigm (Fay et al., 2010; Garrod et al., 2007, 2010). No statistical analyses were run prior to collecting the full sample. Participants were tested in unacquainted pairs in testing sessions lasting 1 hour. All participants reported being free of any uncorrected visual impairment.

7.2. Task and procedure

The experimental paradigm is identical to that used in Experiment 1, including the meaning set (see **Table 1**). Like Experiment 1, Experiment 2 was administered using the virtual whiteboard tool developed by Healy et al. (2002).

Experiment 2 examined the influence of behaviour alignment and concurrent partner feedback on communication success and sign symbolization. Participants were assigned to one of four conditions that represented a combination of the factors of interest: Imitation (Allow Imitation, Forbid Imitation) and Feedback (Allow Feedback, No Feedback). Thirty participants (15 dyads) were randomly assigned to each condition. In the Forbid Imitation conditions (Allow Feedback or No Feedback) participants were instructed not to imitate their partner's drawing for each meaning. They were told they would have to use a different sign to that used by their partner to communicate each meaning. So, in this condition, participants were unable to align their behaviour. In the No Feedback conditions (Allow Imitation or Forbid Imitation) participants were unable to produce within-trial feedback when acting as the Matcher. Specifically, they were unable to draw while acting as the Matcher (this functionality was removed from the virtual whiteboard tool). In this condition the Director clicked a send button when they had finished their drawing. Once done, the list of competing meanings became available for selection by the Matcher. Thus, Matchers were unable to interrupt the Director's communication and bring the trial to an end.

7.3. Measures

Like Experiment 1, Experiment 2 measured *Communication Success*, *Sign Symbolization* and *Behaviour Alignment*. Behaviour alignment was quantified by rating the similarity of pairs of drawings of the same meaning from each dyad (at Game 1-2, 2-3, 3-4, 4-5, 5-6) on a Likert scale from 0-9, where 0= very dissimilar and 9= very similar (author BW). In total 4800 pairs of drawings were rated for similarity (16 meanings x 5 pairs of adjacent games x 15 dyads x 4 conditions). A subset of drawings (1200 pairs of drawings; 300 randomly sampled from each condition) were rated for similarity by a second judge (author NF). The raters were blind to the condition the drawings were sampled from. Comparison of the two sets of ratings showed strong inter-coder agreement (r= .710, p< .001).

8. Experiment 2 results

Examples of sign symbolization and sign alignment from the different experimental conditions are given in **Fig. 7**. Participants who were instructed not to imitate their partner's sign for each meaning followed the instructions: one participant drew a building with a flag to communicate 'Parliament' and their partner drew a speaker at a podium (Forbid Imitation, Allow Feedback); another drew a parliamentary speaker with a hammer, and their partner drew a series of buildings (Forbid Imitation, No Feedback). When allowed to imitate their partner's signs, behaviour alignment was observed: onto a flag (Allow Imitation, Allow Feedback), or people seated around a table (Allow Imitation, No Feedback). These examples also highlight the diversity of signs used to communicate the same meaning in the present study. Concurrent partner feedback had a strong effect on sign symbolization: in the Allow Feedback conditions the signs were dramatically simplified across games, and in the No Feedback conditions they retained considerable sign complexity.

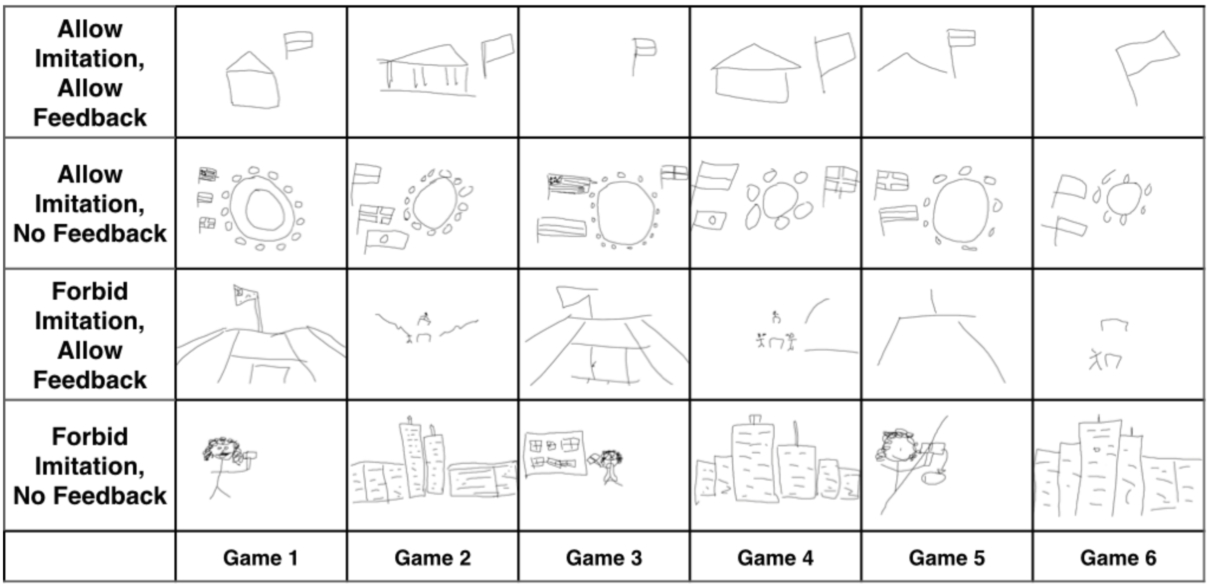

Fig. 7. Example drawings of the meaning 'Parliament' from the different experimental conditions across Game 1-6 from Experiment 2.

8.1. Manipulation check: Behaviour alignment

The first analysis tested whether participants who were forbidden from imitating the signs produced by their partner followed the instructions. The sign similarity data was analyzed using a linear mixed effects model. Imitation (Allow Imitation, Forbid Imitation), Feedback (Allow Feedback, No Feedback) and Game (1-6) were entered as fixed effects with interaction. All fixed effects were centered. The random effects structure included by-Dyad and by-Item random intercepts, as well as by-Dyad random slopes for Game and by-Item random slopes for the Imitation by Feedback interaction. This was the maximal random effects structure that would converge. The best fitting model specified Imitation and Game as fixed effects with interaction ($\chi^2(1) = 36.649, p < .001$). The interaction effect is explained by the increase in sign alignment scores over games in the Allow Imitation conditions (β= 0.407, *SE*= 0.055, $\chi^2(1) = 25.578, p < .001$) and the null effect of Game in the Forbid Imitation conditions (β= 0.035, *SE*= 0.039, $\chi^2(1) = 0.822,\ p = .365$). See **Fig. 8** for data visualisation.

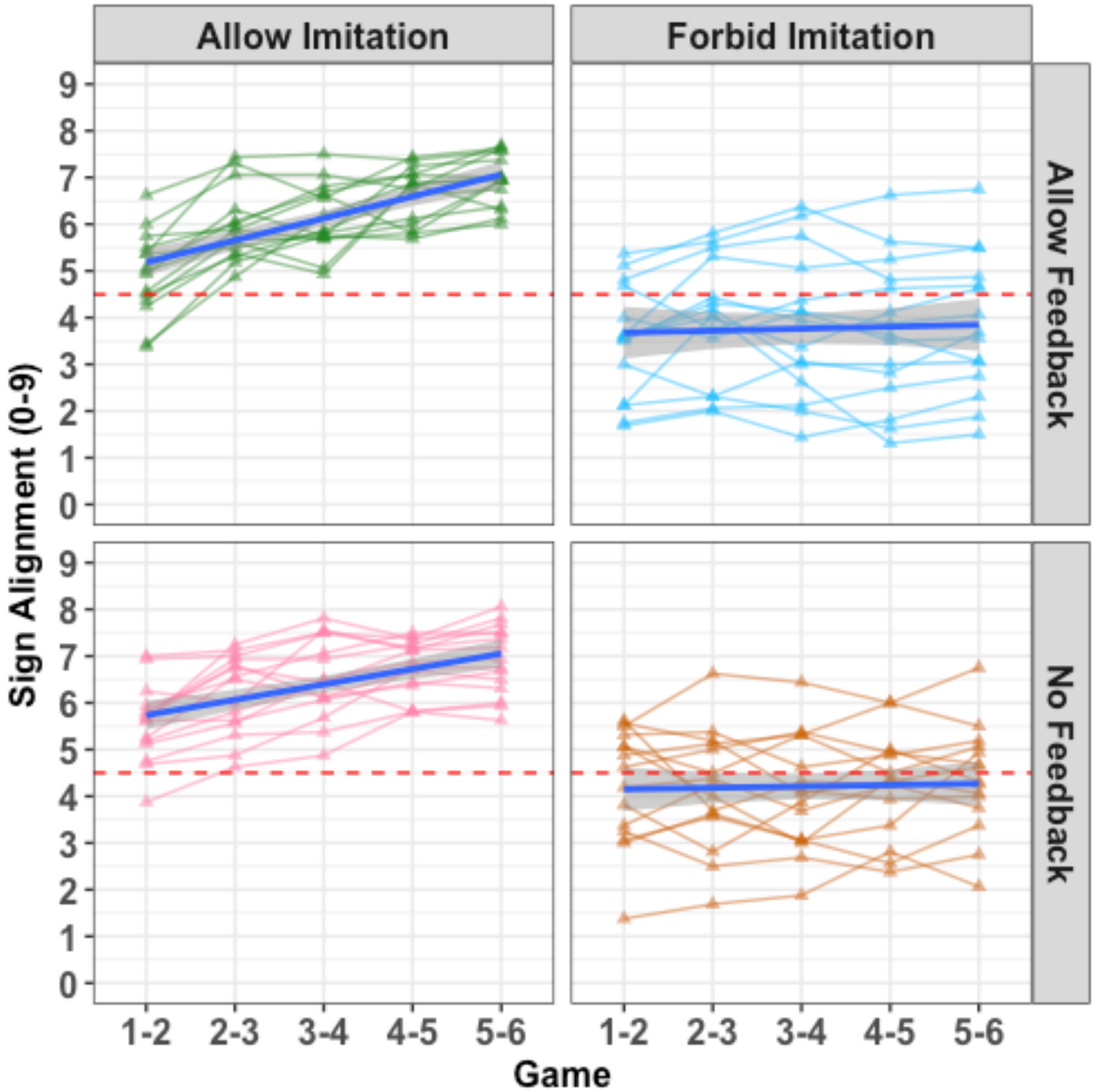

Fig. 8. Change in sign alignment scores (plotted for each dyad) for the different conditions over Games. The horizontal dashed red line indicates neutral sign alignment. The blue straight line is the linear model fit and the grey shaded area is the 95% confidence interval.

8.2. Communication success

Next we examined the change in communication success across Games 1-6 in the Imitation (Allow Imitation, Forbid Imitation) and Feedback (Allow Feedback, No Feedback) conditions.

The data was analyzed using a logistic mixed effects model. Imitation, Feedback and Game were entered as fixed effects with interaction. All fixed effects were centered. Due to a technical error, item information was not recorded for the communication success data or the sign complexity data. The random effects structure included by-Dyad random intercepts. This was the maximal random effects structure that would converge. The best fitting model specified a three-way Imitation by Feedback by Game interaction ($\chi^2(1) = 6.919, p = .008$).

Comparison of the Allow Imitation conditions (Allow Feedback, No Feedback) indicated an improvement in communication success over games (*β*= 0.656, *SE*= 0.081, $\chi^2(1) = 41.398, p < .001$), but there was no statistical evidence that partner feedback affected communication success ($\chi^2(1) = 0.427,\ p = .513$). Comparison of the Forbid Imitation conditions (Allow Feedback, No Feedback) returned a Feedback by Game interaction ($\chi^2(1) = 3.699, p = .054$). This reflected the stronger improvement in communication success over games when partner feedback was allowed (*β*= 0.292, *SE*= 0.047, $\chi^2(1) = 41.150, p < .001$) compared to when participants were unable to provide partner feedback (*β*= 0.169, *SE*= 0.041, $\chi^2(1) = 17.509,\ p < .001$). We then compared the Allow Imitation conditions (collapsed) to each of the Forbid Imitation conditions (Allow Feedback, No Feedback). In each case, this returned a condition by Game interaction ($\chi^2(1) = 4.906, p = .027$ and $\chi^2(1) = 14.900,\ p < .001$). This is explained by the stronger improvement in communication success over games in the Allow Imitation conditions compared to each of the Forbid Imitation conditions. See **Fig. 9** for data visualisation.

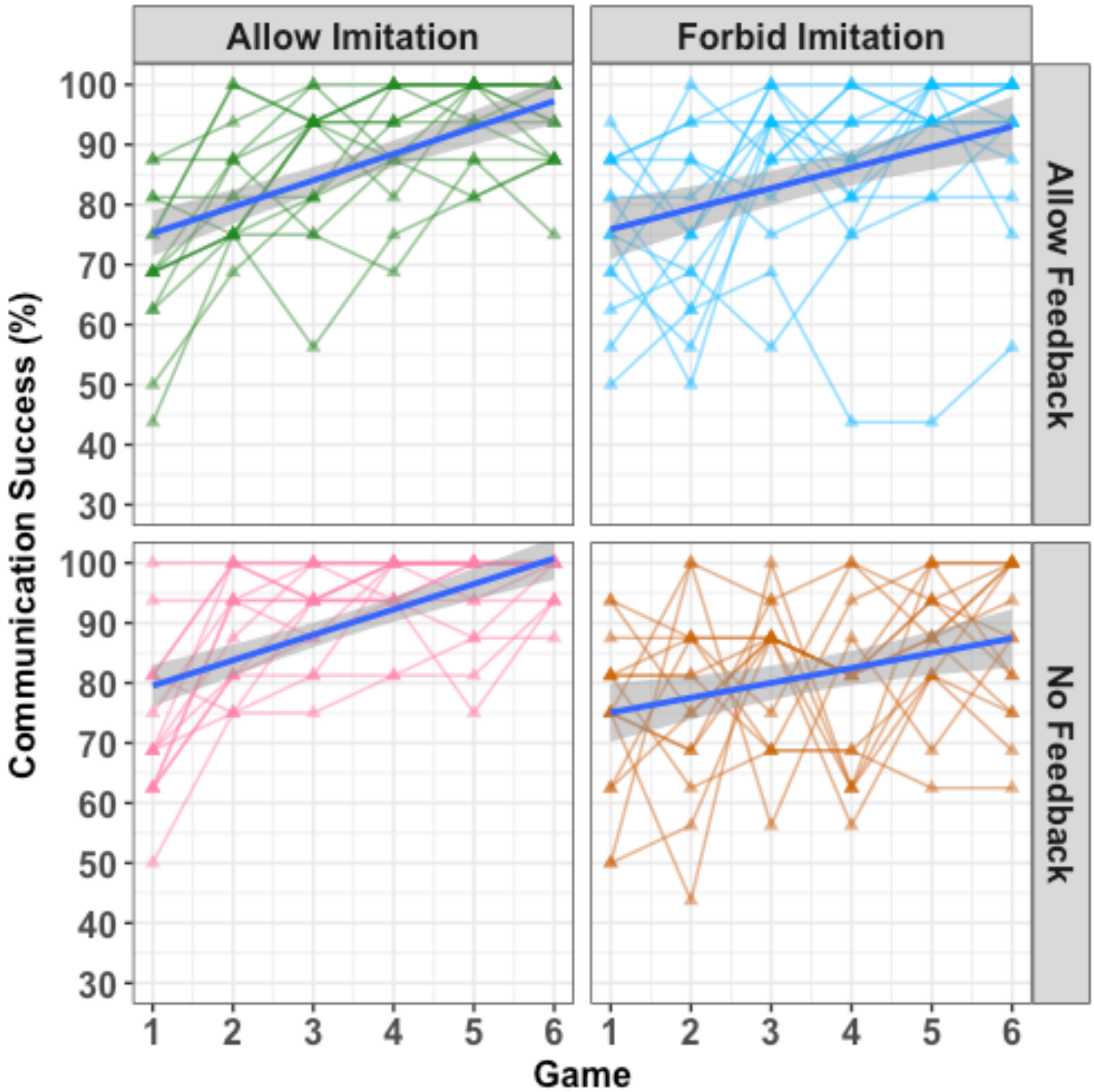

Fig. 9. Change in communication success (plotted for each dyad) for the different conditions across Games 1-6. The blue straight line is the linear model fit and the grey shaded area is the 95% confidence interval.

8.3. Sign symbolization

The final analysis compared the change in sign complexity over games in the different conditions. As before, less complex signs were considered to be more symbolic (see Garrod et al., 2007). Sign complexity was again measured using Pelli et al.'s (2006) information theoretic measure of perimetric complexity [Perimetric complexity = (inside + outside perimeter)^2/ink area]. The sign complexity data was analyzed using a linear mixed effects model. Imitation, Feedback and Game were entered as fixed effects with interaction. All fixed effects were centered. The random effects structure included by-Dyad random intercepts, as well as by-Dyad random slopes for Game. Like communication success, the best fitting model specified a three-way Imitation by Feedback by Game interaction ($\chi^2(1) = 4.926, p = .026$).

Comparison of the Allow Feedback conditions (Allow Imitation, Forbid Imitation) indicated a reduction in sign complexity over games (β= -591.57, *SE*= 68.68, $\chi^2(1) = 37.353, p < .001$), but there was no statistical evidence that imitation reduced sign complexity ($\chi^2(1) = 1.253,\ p = .263$). Comparison of the No Feedback conditions (Allow Imitation, Forbid Imitation) returned an Imitation by Game interaction ($\chi^2(1) = 6.649, p = .009$). This reflected the reduction in sign complexity over games when sign imitation was allowed (β= -754.70, *SE*= 129.6, $\chi^2(1) = 17.729, p < .001$), and a null effect of Game when sign imitation was forbidden (β= -220.1, *SE*= 147.00, $\chi^2(1) = 2.090,\ p = .148$). We then compared the Allow Feedback conditions (collapsed) to the Allow Imitation but No Feedback condition. The best fitting model included a main effect of Game ($\chi^2(1) = 53.245, p < .001$), indicating that sign complexity decreased over games, and a main effect of condition ($\chi^2(1) = 18.282,\ p < .001$), indicating that overall sign complexity was lower when participants could provide concurrent partner feedback. See **Fig. 10** for data visualisation.

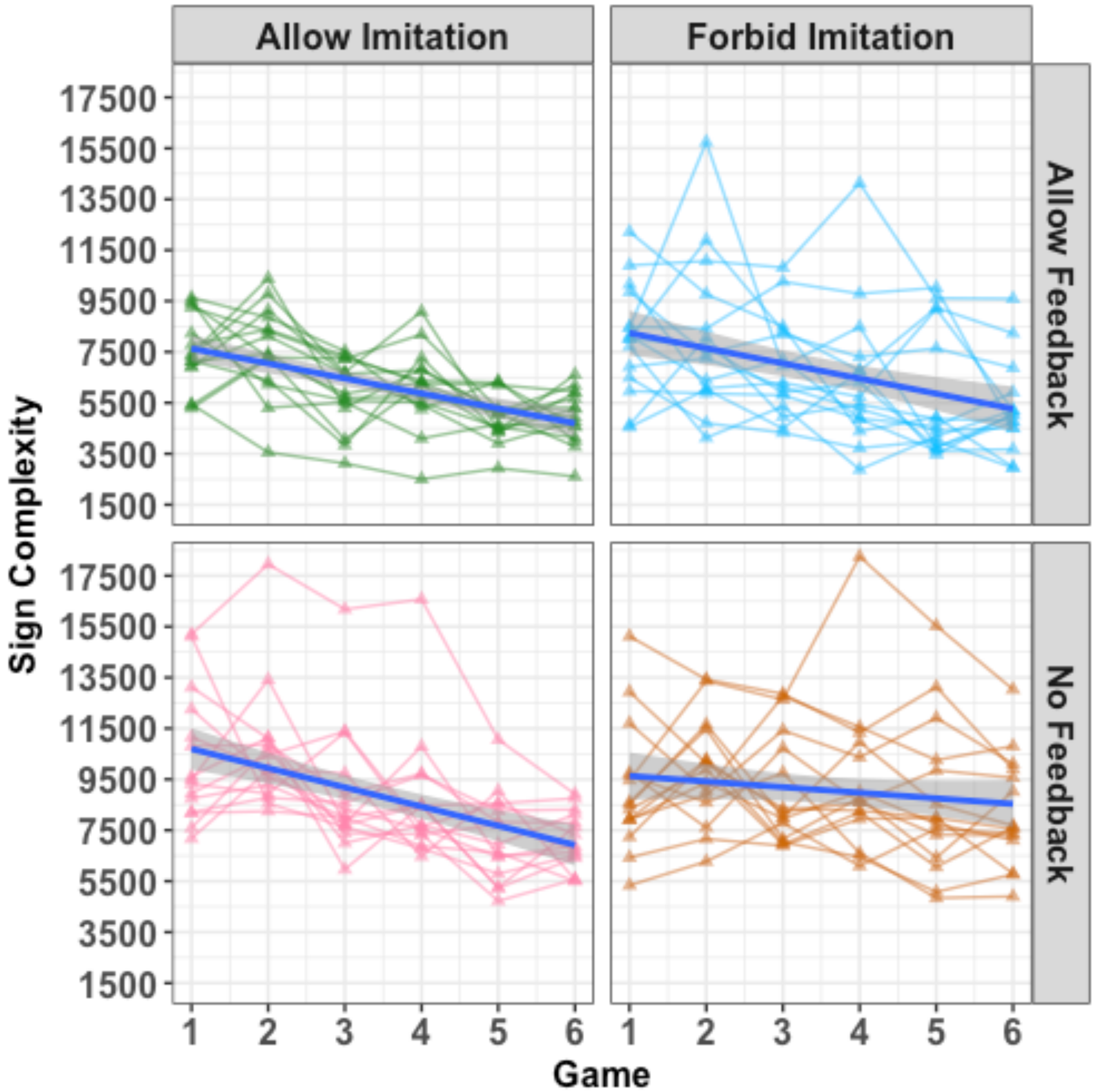

Fig. 10. Change in perimetric complexity of the signs (plotted for each dyad) for the different conditions across Games 1-6. The blue straight line is the linear model fit and the grey shaded area is the 95% confidence interval.

9. Experiment 2 discussion

Experiment 2 isolated two distinct aspects of social interaction – behaviour alignment and concurrent partner feedback – and examined the contribution of each, and their combination, to the creation of shared symbols. To examine the role of behaviour alignment, participants in half of the dyads tested were instructed not to imitate the signs produced by their partner. The manipulation worked; participants who were allowed to imitate their partner’s signs did so, and this led to increased sign alignment, via behaviour matching, over games. There was no evidence of sign alignment among participants for whom sign imitation was forbidden.

As predicted, sign alignment improved communication success, establishing a causal link between behaviour alignment and comprehension. When sign imitation was forbidden concurrent partner feedback improved communication success, but not as strongly as behaviour alignment. Concurrent partner feedback proved to be important to sign symbolization. Allowing the matcher to interrupt the director, and bring the trial to an end via meaning selection, drove progressive sign simplification and abstraction over games. When unable to provide concurrent partner feedback (the functionality was removed from the interface), behaviour alignment reduced sign complexity, but not to the extent of concurrent partner feedback. Without the opportunity for behaviour alignment or the opportunity to provide concurrent partner feedback, there was no evidence of sign symbolization.

Taken together, the Experiment 2 results demonstrate that each process played a complementary role in the creation of shared symbols: behaviour alignment drove communication success and concurrent partner feedback drove sign symbolization.

10. General discussion

Experiment 1 examined the importance of social interaction to the creation of shared symbols. Interaction proved to be important to the evolution of communication systems that were effective, efficient and shared. Compared to the conditions where the opportunity for social interaction was removed, in the Interaction condition communication success was higher, the signs became simpler and more symbolic, and interlocutors increasingly used the same signs to communicate the same meanings (i.e., their behaviour aligned). These findings support the results of pragmatic (e.g., Garrod & Anderson, 1987; Krauss & Weinheimer, 1964; Schober & Clark, 1989) and semiotic experiments (e.g., Garrod et al., 2007, 2010) by demonstrating that inter-individual coordinative social learning is important to the creation of shared symbols.

The Experiment 1 results also indicate that observation and cognitive bias may play a role in the creation of shared symbols. When denied the opportunity to directly interact with their partner, participants in the Pseudo-Interaction: Precedent condition showed increased communication success, sign symbolization and sign alignment across the communication games. Although lower on each measure compared to participants who could directly interact with their partner, this finding suggests that individual-level observational learning positively contributed to the creation of shared symbols. These findings support theoretical accounts and empirical studies in which observation plus cognitive biases guide language evolution (e.g., Kirby et al., 2008; Kirby, Griffiths, & Smith, 2014; Thompson, Kirby, & Smith, 2016). Note that when a referential precedent was broken (Pseudo-Interaction: Broken Precedent) the influence of observation and cognitive bias was eliminated; there was no statistical evidence of a change in communication success, sign symbolization or sign alignment across the communication games. Like experimental pragmatic studies, breaking a referential precedent negatively impacted interpersonal communication (Kronmüller & Barr, 2015).

Having established that social interaction is important to the creation of shared symbols, Experiment 2 investigated the precise role played by two distinct aspects of social interaction: behaviour alignment and feedback. By experimentally manipulating the opportunity for behaviour alignment and concurrent partner feedback in a full factorial design, Experiment 2 demonstrated that each process made a distinct contribution to the creation of shared symbols: behaviour alignment primarily drove improvements in communication success and concurrent partner feedback primarily drove improvements in sign efficiency. Together, these complementary processes drove the interactive evolution of shared symbols.

The Experiment 1 and 2 findings suggest a possible solution to the symbol grounding problem (Harnad, 1990). Complex iconic signs ground shared meanings. Once grounded, social interaction drives sign alignment and refinement, the mechanisms through which effective and efficient shared symbols arise. This explanation offers a convincing candidate process through which iconic signs evolve into symbols, as originally proposed by Charles Sanders Peirce over 100 years ago (Peirce, 1931).

10.1. The interplay between cognitive bias & social interaction

Smith and Wonnacott (2010) examined the effect of intergenerational transmission on the elimination of unpredictable variation in a miniature artificial language. They found that as the miniature language was transmitted from person to person across a transmission chain it became increasingly regularized and language-like, suggesting that a bias for regularity was amplified across repeated transmission episodes, i.e., via unidirectional vertical transmission and without social interaction. Using an identical task, but one where participants were organized into interacting pairs as opposed to transmission chains, Smith, Fehér and Ritt (2014) showed that unpredictable variation was eliminated across participants' repeated social interactions with the same partner. This finding suggests that an individual-level bias for regularity may have been amplified by social interaction (see also Fehér, Wonnacott, & Smith, 2016). A key benefit of the interplay between cognitive bias and social interaction is the timescale on which it operates; when a cognitive bias is amplified via social interaction, rather than via intergenerational transmission, language change can be more rapid and responsive to environmental change.

The Experiment 1 findings support a role for cognitive biases in language evolution that is conditional on the communication context. In the absence of social interaction, a simplicity bias drives sign symbolization, and an alignment bias drives the evolution of a shared inventory of sign-to-meaning mappings among interlocutors. However, when a referential precedent was broken the influence of the cognitive biases was eliminated. By contrast, and similar to the aforementioned artificial language-learning studies, when participants could directly interact with their partner the cognitive biases were amplified, giving rise to a more powerful improvement in communication success, sign symbolization and sign alignment. This finding, that cognitive bias expression is conditional on the communication context, indicates that cognitive biases need not be deterministic, but can adapt to environmental change.

10.2. Scaling up to larger populations

Experiments 1 and 2 indicate that inter-individual social coordinative learning is important to the creation of shared symbols. How might our findings, based on dyadic interaction, scale-up to larger populations? Several experimental studies have examined the processes that operate when participants interact as part of a laboratory micro-society that includes between 8 and 24 members (Centola & Baronchelli, 2015; Fay et al., 2010; Garrod & Doherty, 1994). These studies, alongside agent-based computer simulations (Barr, 2004; Steels, 2003), indicate that the same social coordinative learning mechanisms identified in dyadic interaction experiments drive the evolution of referential conventions in larger populations. So, the processes identified in the present study are likely to be important to the creation of shared symbol systems in larger populations.

Using an identical task to that used in the present study, Fay et al. (2010) examined the evolution of shared symbol systems in 8-person micro-societies. Like the present study, participants interacted in pairs. After several games they switched partners, and continued in this way until they had interacted with each of the other members of their micro-society. Initially a diverse range of complex motivated signs was used to communicate each of the different meanings. Across interactions communication success improved and the initial sign variation was lost as participants aligned on a uniform inventory of single sign-to-meaning mappings. In addition, the signs used to communicate the different meanings became increasingly simplified and symbolic across repeated interactions in each micro-society. So, like the present study, social interaction improved communication success, behaviour alignment and sign symbolization, but at the population-level.

Increasing the population size also increased the diversity of signs that were used to communicate each meaning, and this increased competition between the different signs.

Tamariz et al. (2014) modelled the change in the frequency of the different communication variants used in each micro-society to communicate each meaning. They found that the data was best modelled by a combination of 'egocentric bias' and 'content bias'. When participants encountered a new sign-to-meaning mapping, they tended to reuse the sign they had used previously (egocentric bias) unless the newly encountered sign was perceived to be superior (content bias). In a large population, this preference to adopt the most informative sign (see Rogers & Fay, 2016 for empirical support) led to the selection of a set of sign-to-meaning mappings that were better designed, relative to those developed in interacting dyads, for comprehension and production by naïve learners (Fay & Ellison, 2013; Fay et al., 2008).

The findings of the present study scale up to larger populations, but larger populations add a selection dynamic that improves the ease of acquisition, and the transmission fidelity of the evolved signs, an outcome consistent with cumulative cultural evolution (Tennie, Call, & Tomasello, 2009; Tomasello, 1999).

11. Conclusion

This paper examined the social learning strategies important to the creation of shared symbols. Experiment 1 demonstrated that individual-level processes, via observation and cognitive biases, contributed to the evolution of effective, efficient and shared symbols. However, when a referential precedent was broken the benefits of these individual-level processes were eliminated. Importantly, the addition of inter-individual processes, via social interaction, produced the most effective, most efficient, and most shared symbols. These findings demonstrate that social coordinative learning plays an important role in the creation of shared symbols. Our findings also suggest that cognitive bias expression (simplicity and alignment bias) is conditional on the communication context, such that breaking a referential precedent eliminated the influence of the bias, and social interaction amplified the bias. Having established that social coordinative learning is important to the creation of shared symbols, Experiment 2 examined the precise contribution made by two distinct aspects of social interaction: behaviour alignment and concurrent partner feedback. Behaviour alignment primarily drove improvements in communication success and concurrent partner feedback primarily drove improvements in sign efficiency.

Social coordinative learning plays an important role in the evolution of shared symbols. The benefits of social coordinative learning arise through two complementary aspects of social interaction: behaviour alignment drives sign effectiveness, and concurrent partner feedback drives sign efficiency and symbolization.

Acknowledgements

We thank Alan Bailey who helped with the data collection for Experiment 2 and Casey Lister for her feedback on an earlier version of this paper. We are also grateful for the feedback provided by two anonymous assessors, Seth Frey (UC Davis) and by the editor, Todd Gureckis. N.F. and S.G. acknowledge support by an ARC Discovery grant (no. DP120104237).

References

Bandura, A. (1977). *Social learning theory*. General Learning Press.

Barr, D. J. (2004). Establishing conventional communication systems: Is common knowledge necessary? *Cognitive Science*, *28*(6), 937–962.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear mixed-effects models using Eigen and S4. *R Package Version*, *1*(4).

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology*, *79*(6), 941–952.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*(2), B13–B25.

Branigan, H. P., Pickering, M. J., McLean, J. F., & Cleland, A. A. (2007). Syntactic alignment and participant role in dialogue. *Cognition*, *104*(2), 163–197.

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, *121*(1), 41–57. https://doi.org/10.1016/j.cognition.2011.05.011

Brennan, S. E., & Clark, H. H. (1996). Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493.

Bronowski, J. (1967). *Human and animal languages*. The Hague: Mouton.

Caldwell, C. A., & Smith, K. (2012). Cultural Evolution and Perpetuation of Arbitrary Communicative Conventions in Experimental Microsocieties. *PLoS ONE*, *7*(8), e43807. https://doi.org/10.1371/journal.pone.0043807

Centola, D., & Baronchelli, A. (2015). The spontaneous emergence of conventions: An experimental study of cultural evolution. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1418838112

Chater, N., & Vitányi, P. (2003). Simplicity: a unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*(1), 19–22.

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

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*(1), 62.

Clark, H. H., & Schaefer, E. F. (1987). Collaborating on contributions to conversation. *Language and Cognitive Processes*, *2*, 19–41.

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

Deacon, T. (1997). *The Symbolic Species: The co-evolution of langauge and the brain*. New York, NY: Norton.

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., … others. (2015). Universal principles in the repair of communication problems. *PloS One*, *10*(9), e0136100.

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

Fay, N., Ellison, M., & Garrod, S. (2014). Iconicity: From sign to system in human communication and language. *Pragmatics & Cognition*, *22*(2), 244–263.

Fay, N., & Ellison, T. M. (2013). The Cultural Evolution of Human Communication Systems in Different Sized Populations: Usability Trumps Learnability. *PLoS ONE*, *8*(8), e71781. https://doi.org/10.1371/journal.pone.0071781

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), 3553–3561.

Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science*, *34*(3), 351–386.

Fay, N., Lister, C. J., Ellison, T. M., & Goldin-Meadow, S. (2014). Creating a Communication System from Scratch: Gesture Beats Vocalization Hands Down. *Frontiers in Psychology*, *5:354*.

Fehér, O., Wonnacott, E., & Smith, K. (2016). Structural priming in artificial languages and the regularisation of unpredictable variation. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2016.06.002

Frishberg, N. (1975). Arbitrariness and iconicity: Historical change in American Sign Language. *Language*, *51*, 696–719.

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to Terms: Quantifying the Benefits of Linguistic Coordination. *Psychological Science*, *23*(8), 931–939.

Fusaroli, R., & Tylén, K. (2016). Investigating Conversational Dynamics: Interactive Alignment, Interpersonal Synergy, and Collective Task Performance. *Cognitive Science*, *40*(1), 145–171. https://doi.org/10.1111/cogs.12251

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science*, *29*(5), 737–767.

Galantucci, B. (2017). Experimental Semiotics. *Oxford Research Encyclopedia of Linguistics*, 1–18. https://doi.org/10.1093/acrefore/9780199384655.013.210

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181–218.

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.

Garrod, S., Fay, N., Rogers, S., Walker, B., & Swoboda, N. (2010). Can iterated learning explain the emergence of graphical symbols? *Interaction Studies*, *11*(1), 33–50.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1–3), 335–346.

Healy, P. G. T., Swoboda, N., & King, J. (2002). A tool for performing and analysing experiments on graphical communication. In X. Faulkner, J. Finlay, & F. Detienne (Eds.), *People and computers XVI: Proceedings of HCI2002: The 16th British HCI Group Annual Conference* (pp. 55–68). London: Springer-Verlag.

Healy, P. G. T., Swoboda, N., Umata, I., & King, J. (2007). Graphical language games: Interactional constraints on representational form. *Cognitive Science*, *31*, 285–309.

Hupet, M., & Chantraine, Y. (1992). Changes in repeated references: Collaboration or repetition effects? *Journal of Psycholinguistic Research*, *21*, 485–496.

Keysar, B., & Barr, D. J. (2002). Anchoring Comprehension in Linguistic Precedents. *Journal of Memory and Language*, *46*(2), 391–418. https://doi.org/10.1006/jmla.2001.2815

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686.

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28*(0), 108–114. https://doi.org/10.1016/j.conb.2014.07.014

Klima, E., & Bellugi, U. (1979). *The Signs of Language*. Cambridge MA: Harvard University Press.

Krauss, R. M., & Weinheimer, S. (1964). Changes in the length of reference phrases as a function of social interaction: A preliminary study. *Psychonomic Science*, *1*, 113–114.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, *4*, 343–346.

Kronmüller, E., & Barr, D. J. (2015). Referential precedents in spoken language comprehension: A review and meta-analysis. *Journal of Memory and Language*, *83*, 1–19.

Mein, C., Fay, N., & Page, A. C. (2016). Deficits in joint action explain why socially anxious individuals are less well liked. *Journal of Behavior Therapy and Experimental Psychiatry*, *50*, 147–151. https://doi.org/10.1016/j.jbtep.2015.07.001

Morgan, T. J. H., Laland, K. N., & Harris, P. L. (2015). The development of adaptive conformity in young children: effects of uncertainty and consensus. *Developmental Science*, *18*(4), 511–524. https://doi.org/10.1111/desc.12231

Peirce, C. S. (1931). *Collected Papers of Charles Sanders Peirce* (Vol. 1–8). Cambridge, MA: Harvard University Press.

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, *46*(28), 4646–4674.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526–3529. https://doi.org/10.1073/pnas.1012551108

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–226.

R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Reitter, D., & Moore, J. D. (2014). Alignment and task success in spoken dialogue. *Journal of Memory and Language*, *76*, 29–46. https://doi.org/10.1016/j.jml.2014.05.008

Roberts, G., Lewandowski, J., & Galantucci, B. (2015). How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition*, *141*, 52–66.

Rogers, S. L., & Fay, N. (2016). Stick or Switch: A Selection Heuristic Predicts when People Take the Perspective of Others or Communicate Egocentrically. *PloS One*, *11*(7), e0159570.

Schegloff, E. A. (2000). When'others' initiate repair. *Applied Linguistics*, *21*(2), 205–243.

Schober, M. F., & Clark, H. H. (1989). Understanding by Addressees and Overhearers. *Cognitive Psychology*, *21*(2), 211–232.

Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, *131*(3), 431–436.

Smith, K., Feher, O., & Ritt, N. (2014). Eliminating Unpredictable Linguistic Variation through Interaction. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 1461–1466). Cognitive Science Society.

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*(3), 444–449.

Sperber, D., & Wilson, D. (1987). Precis of relevance: Communication and cognition. *Behavioral and Brain Sciences*, *10*(4), 697–710.

Steels, L. (2003). Evolving grounded communication for robots. *Trends in Cognitive Sciences*, *7*(7), 308–312.

Stolk, A., Verhagen, L., & Toni, I. (2016). Conceptual Alignment: How Brains Achieve Mutual Understanding. *Trends in Cognitive Sciences*, *20*(3), 180–191.

Tamariz, M. (2017). Experimental Studies on the Cultural Evolution of Language. *Annual Review of Linguistics*, *3*(1), 389–407. https://doi.org/10.1146/annurev-linguistics-011516-033807

Tamariz, M., Ellison, T. M., Barr, D. J., & Fay, N. (2014). Cultural selection drives the evolution of human communication systems. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1788), 1–6. https://doi.org/10.1098/rspb.2014.0488

Tamariz, M., & Kirby, S. (2014). Culture: copying, compression, and conventionality. *Cognitive Science*, *39*(1), 171–183.

Tan, R., & Fay, N. (2011). Cultural transmission in the laboratory: Agent interaction improves the intergenerational transfer of information. *Evolution & Human Behavior*, *32*(6), 399–406.

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: on the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2405–2415. https://doi.org/10.1098/rstb.2009.0052

Theisen, C. A., Oberlander, J., & Kirby, S. (2010). Systematicity and arbitrariness in novel communication systems. *Interaction Studies*, *11*(1), 14–32.

Thompson, B., Kirby, S., & Smith, K. (2016). Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences*, *113*(16), 4530–4535. https://doi.org/10.1073/pnas.1523631113

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences*, *16*(3), 495–511.

Vaccari, O., & Vaccari, E. E. (1961). *Pictorial Chinese-Japanese characters* (4th ed.). Vermont, USA: Tokyo, Japan: Charles E. Tuttle.

Wescott, R. W. (1971). Linguistic iconism. *Language*, *47*, 416–428.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. New York: Addison-Wesley.