

The arena of language evolution: the emergence of symbolic referential signals in a common task framework

Seán G. Roberts^{1,*}, Kateryna Krykoniuk^{1,2}, Fiona M. Jordan³

¹School of English, Communication and Philosophy, Cardiff University, John Percival Building, Cardiff CF10 3EU, United Kingdom

²School of Languages, Arts & Societies, University of Sheffield, Jessop West, Sheffield S3 7RA, United Kingdom

³Department of Anthropology and Archaeology, University of Bristol, 43 Woodland Rd, Bristol BS8 1TH, United Kingdom

*Corresponding author. School of English, Communication and Philosophy, Cardiff University, John Percival Building, Cardiff CF10 3EU, United Kingdom. E-mail: robertss55@cardiff.ac.uk

Associate Editor: Sławomir Wacewicz

We present the common task framework approach to testing causal theories about the evolution of language. There are now many theories about how symbolic communication emerged, but less work trying to compare, synthesize and test these theories. We suggest that the first step is to formalize the theories as causal graphs using tools from the field of causal inference. This helps recognize the critical causal links that differentiate theories. The second step is to use methods from lab-based experimental semiotics to specify a ‘common task’ or ‘arena’: an experimental environment and a task for individuals to complete. The different theories suggest different designs for this arena, and the experimental results can be used as a measure of the relative success of each theory. In this paper, we provide an example from anthropological theories of the emergence of symbolic communication, suggesting that an effective arena contains an asymmetry of information, division of labour and contextually distal meanings. We run experiments in arenas based on collaborative construction and fire maintenance. The results indicate that the effectiveness of pointing can limit the emergence of symbolic signals, a problem that has previously not been worked into theories. In this way, we hope the common task framework can be used as a method to further develop theories of language evolution.

Keywords: symbolic communication; experimental semiotics; causal inference; anthropology.

1. Introduction

Human language is the most complex communication system on earth, with no other animal coming close to our linguistic abilities. How did this ability evolve, and why only in humans? While genetic and cognitive factors are obviously relevant (e.g. Fisher and Vernes 2015), perspectives from anthropology and archaeology are making it increasingly clear that social, economic, and ecological situations are also part of the story. Studies show that many species can acquire some advanced communicative behaviours under training that they do not exhibit in the wild (e.g. Premack 1971; Lyn and Christopher 2020; Leavens 2021; Pepperberg 2021; Kalan, Nakano, and Warshawski 2023). This suggests that their natural habitats do not provide the right selective pressures for these latent abilities to evolve further.

Only for humans did the right factors come together in a ‘language producing niche’ to motivate the evolution of complex language. However, the social, economic and ecological conditions of this niche are under-explored (see Roberts 2018). That is, at some point in the history of human evolution, our ancestors had the capacity to communicate using symbolic signals, but no such conventions had yet been invented. Some configuration of living conditions motivated them to start communicating in new ways. What were these conditions, and why did they seemingly only apply to recent hominid species? This question is similar to ‘the motivation problem’ (Hurford 2007: 131) or the ‘referential problem space’ (Leavens, Hopkins, and Bard 2005) and linked to the concept of ‘mitteilungsbedürfnis’ (Fitch 2010: 140). Following Hurford, we call these conditions the ‘arena of language evolution’ (see Hurford 1989: 189; 1990: 86).

Anthropology and archaeology have a wealth of theories about the kinds of arenas that plausibly created such a motivation for symbolic communication (for reviews and collections see e.g. Noble and Davidson 1996; Knight, Studdert-Kennedy, and Hurford 2000; Botha and Knight 2009; Allen et al. 2011; Dor, Knight, and Lewis 2014; Power, Finnegan, and Callan 2016). These include conditions related to predation (Hill and Dunbar 1998), hunting (Sterelny 2012; Knight and Lewis 2017), navigation (Bednarik 1997; Kazakov and Bartlett 2004), tool use (Davidson and Noble 1989), fire making (Twomey 2013; Wiessner 2014), and reproductive strategies (Knight, Power, and Watts 1995; Burkart, Hrdy, and Van Schaik 2009).

However, while theories abound, there are few attempts to synthesize them and systematically test them against each other. In this paper, we suggest a novel approach to addressing this gap. This endeavour presents two main challenges. The first is identifying relevant theories and formalizing their agreements or conflicts. Ideally, we would identify specific testable predictions to evaluate the relative plausibility of each theory. This process requires a systematic review of the literature and a way of formally representing the causal structure of theories. We suggest that causal inference (e.g. Pearl and Mackenzie 2018) can help researchers overcome this challenge, particularly using the causal hypotheses in evolutionary linguistics database (CHIELD, Roberts et al. 2020), and a database of hypotheses, which have been hand-coded as causal graphs. CHIELD allows users to discover formal links between hypotheses, assess conflicting evidence and spot weak links that currently have little support.

The second challenge lies in how to test these theories against each other empirically. Since we cannot observe language evolution directly, and naturalistic methods provide limited flexibility, support for theories must come from a robust combination of approaches (Irvine, Roberts, and Kirby 2013). We suggest a ‘common task framework’ can combine control from experimental semiotics and ecological validity from anthropology and archaeology. This framework is inspired by a synthetic approach to theory building—what Webb (2000) calls ‘simulation modelling’ and similar to the ‘emergent constructive approach’ (Hashimoto 2020). The principle is that if a theory is sound, then we should be able to construct a physical model that exhibits predicted emergent behaviours. In the context of language evolution, it should be possible to define an ‘arena’ including an environment and a task for agents to complete that reflects relevant and plausible analogues of early hominid life, but where communicating with modern language is prohibited. In the presence of the right conditions, the agents

are more likely to invent a new communication system, and in their absence they are less likely. Importantly, these simulation models do not necessarily constitute evidence that the theory is correct. Rather, the goal is to compel researchers to be explicit about their assumptions and to bring to light inconsistencies in the logic of the theory.

In this paper, we outline how causal inference and a common task framework can meet the two challenges above. We illustrate the approach by conducting experiments in two arenas to explore the emergence of *symbolic referential signals* (abbreviated hereafter as SRS): essentially ‘names for things’ or ‘semanticity’ (Hockett 1963: 8)—conventionally understood signals (whose form may be motivated, see Özyürek 2021) that can represent their referents with limited context (Greenfield et al. 2008: 34–35), as opposed to indexical (e.g. pointing, see Gontier 2013) or depictive communication (e.g. non-conventional pantomime, see Żywiec, Waciewicz, and Sibierska 2018). We demonstrate how the common task framework can help refine theories of the emergence of this property.

2. A common task framework for studying symbolic communication

A common task framework involves comparing different solutions to a common practical challenge. For example, in the *DARPA Grand Challenge*, autonomous robotic vehicles complete a task in the real world such as driving across a desert (Chung, Orekhov, and Maio 2023). Different designs can be evaluated against each other according to some evaluation criteria such as the time to reach the finish.

We suggest that a common task framework for language evolution involves designing an ‘arena’, which includes specifying an ‘environment’ and a ‘survival task’ to complete (e.g. a field with building materials and a task to build a shelter). These should reflect plausible analogues of early hominid life. An arena can be tested by placing agents into the arena and observing their behaviour. The agents should have a capacity for symbolic communication, but no existing pre-established conventions for communication within the context of the experiment. The arena is then evaluated based on whether the agents develop symbolic referential signals. As we show below, this will not necessarily occur unless the conditions are right. The experimenter must specify a set of minimally different conditions that separate emergence and non-emergence of the target behaviour. The results from different arena designs can be compared with evaluate the hypotheses against each other.

The aim is not to recreate the exact, true emergence of symbolic communication in early hominids—we cannot know what this situation was. Instead, the

procedures serve as grounded thought experiments, which explore and refine theories, similar to Webb's 'simulation modelling' (Webb 2000), Steels' embodied robotics (Steels 2003) or Braitenberg's synthetic psychology (Braitenberg 1986). There are also similarities with experimental archaeology (Stade 2017; Shilton 2019; Wadley 2023), which aims to 'test out hypothetical scenarios using potentially authentic materials and conditions' and allows researchers to 'be inventive and develop new ways to enlarge our understanding by proxy' (Outram 2008: 2).

An example of a common task framework approach comes from Irvine and Roberts (2016). Pairs of human participants interacted in a 3D virtual world, using the video game *Minecraft* (see also Bun 2016). The environment was a field outdoors, and the survival task was to build a shelter together. Each participant had half of the plan for the shelter, providing something to communicate about. Participants were prevented from using speech during the experiment, but could knock on the table or use their avatar to point in the virtual world. This gave them the capacity to develop simple symbolic signals to refer to the four different building materials (e.g. one knock for red blocks, two knocks for yellow blocks, etc.). The crucial question was whether the arena would motivate them to do so. To the surprise of the researchers, it did not. Instead, participants used pointing and trial-and-error strategies to complete the task. Post-experiment interviews confirmed that participants were not motivated to develop referential signals because the pointing strategy was sufficiently effective. This might seem obvious in hindsight, but this conclusion was only reached by putting the theory into practice.

2.1 Relation to experimental semiotics

The empirical approach to language evolution is not new, and the approach here is inspired by 'experimental semiotics', which uses methods from psychology to explore the role of cognition, acquisition, and usage in shaping language (Galantucci 2009; Roberts 2017; Nölle and Galantucci 2022). Participants typically construct, use, and transmit artificial languages. For example, a 'director' is given a series of meanings to communicate to a 'matcher' in a communication game.

Experimental semiotics has been used to explore the emergence of compositional structure (Kirby et al. 2015). Initial results suggested that inter-generational transmission was a core causal component, but recent studies suggest structure can emerge without transmission in larger populations (Raviv, Meyer, and Lev-Ari 2019). Essentially, the experiments provided a way of comparing solutions to the 'common task' of creating conditions for the emergence of compositionality.

However, in order to obtain a high level of experimental control, these experiments tend to use idealized tasks based on formal communication games, which impose strict limits on how individuals can interact (e.g. the director is prevented from pointing at the target object when they want to request it; only one individual communicates at a time; individuals cannot initiate repair, Macuch Silva and Roberts 2016). As Nölle and Galantucci (2022: 10) put it, experimental semiotics has 'conceived elegant, but ultimately highly artificial, settings that hardly resemble real communicative interactions'.

In contrast, we assert that a valid hypothesis of the emergence of symbolic referential signals should assume embodied, active agents in a rich three-dimensional world that already have powerful non-linguistic communication. As Outram argues, a rich environment is one of the advantages of practical experimentation since 'unpredictable phenomena are often given more opportunity to act, thus enabling the refinement of hypotheses' (Outram 2008: 2).

3. Review of core causal components for an arena of language evolution

This section identifies core causal components of arenas that promote the emergence of symbolic communication. First, we establish our assumptions about the abilities of individuals. We assumed they have the physical capacity to produce variable signals and the cognitive capacity to process symbolic signals and learn from feedback. There is intense debate about what constitutes symbolic behaviour in hominid species or other animals (e.g. Slocombe and Zuberbühler 2005; Wheeler and Fischer 2012; Sievers and Gruber 2016; Dediu and Levinson 2018; Liebal and Oña 2018; Fitch 2020), but we took Hurford's (2007) stance that conceptual thinking is evolutionarily old, alongside more recent evidence that many species exhibit behaviour compatible with functional semantic reference (Townsend and Manser 2013; Lyn and Christopher 2020; Pepperberg 2021; Warren and Call 2022). We also assumed that individuals can achieve joint attention to tertiary objects via pointing or gesture, since this emerges early in human infancy (Liszkowski et al. 2012) and is at least comprehensible by a number of other species, such as primates, canines, birds (see e.g. Krause et al. 2018; McCreary, Jones, and Kuhlmeier 2023; Lyn et al. 2024), and even fish (Vail, Manica, and Bshary 2013). Finally, we assumed the individuals are embedded in a population of conspecifics who are motivated to cooperate (see e.g. Hurford 2007: 252, 307; Adornetti 2015). While there is a considerable body of work on the social and ecological conditions for cooperation and its relative rarity in non-human

animals (see e.g. Waciewicz and Żywicznyński 2018; Apicella and Silk 2019; Sterelny 2021), and these conditions are amenable to testing in a common task framework, we do not explore those conditions here.

Next, we identified the core causal components of the arena using a causal approach incorporating eleven theories from the CHIELD database. A full report of the methods and findings is included in the [Supplementary materials](#). The results suggested that there are three core causal components of an arena that motivates the emergence of symbolic referential signals. The first is asymmetry of information between individuals. Some individuals possess knowledge or skills that others do not, creating something to communicate about. However, the arena also needs to provide a reason to communicate about this thing—therefore, we suspect that a second causal component is a pressure for the division of labour: a reason that operating independently is not possible or not effective, creating a need to know about the thing. Critically, this division of labour creates dependencies between individuals, such as the need to coordinate and schedule, and likely involves trust and cooperation (Dunbar 2014).

However, there are many non-symbolic ways to transfer knowledge and skills, including pantomime (e.g. Zlatev, Żywicznyński, and Waciewicz 2020), social learning through observation (e.g. Tramacere and Moore 2018) and pointing. These require less effort to establish than symbolic referential signals. Indeed, given the communicative power of pointing and gesture (e.g. Macuch Silva et al. 2020), we expect symbolic referential signals to emerge only when the arena somehow discourages these strategies (see discussion of ‘groundedness’ in Számadó and Szathmáry 2006). Therefore, a third causal component is that the referents to be communicated about should be distant in time or space, making non-symbolic communication less effective. However, identifying ecologically valid distal meanings is more difficult than it first appears. If participants are free to move around the world, there are few realistic scenarios where individuals can be prevented from turning distal meanings into proximal meanings. Put another way, in a situation where you can point at something, you can also probably just walk over to it and pick it up without needing to communicate with anyone. Therefore, our arena should include contextually distal meanings: where at least one interlocutor needs to be distant from the referent at the point at which they are referring to it. That is, the ‘distal’ property of meanings is not necessarily inherent to the meaning or referent but emerges from an interaction between where the referents and interlocutors are in context.

Leavens et al. (2005) and Leavens (2021) note that contextually distal meanings may be rare in the wild

(e.g. needing to communicate across rivers). In contrast, captive environments with cages or screens change the ‘referential problem space’ to provide cases where desirable objects are visible but cannot be reached. Chimpanzees change their communicative behaviour in these environments (e.g. using manual pointing gestures, see Leavens et al. 2005; Leavens, Russell, and Hopkins 2010; but cf. Povinelli, Bering, and Giambrone 2003), demonstrating that a difference in the arena can motivate the emergence of a latent communicative ability. Leavens et al. (2005) suggest that extended neoteny creates ‘endogenous barriers’ (e.g., infants cannot reach things). However, it is not clear to us whether a need for *reciprocal* communication (Tomasello 1990) would be motivated in a parent-child relationship. Instead, we suggest that environments where coordination increases efficiency may be more likely candidates for motivating communication (e.g., honey bees’ harvesting of nectar and pollen). Distal meanings seem rarer in our more immediate relations but might include captive chimpanzee food calls (Leavens et al. 2005; Watson et al. 2015) and wild chimpanzee drumming to communicate information about travel activity (Gabrić 2022).

Meanings may also be contextually distal in time such as referring to events or coordinating plans. Some animals seem to conceptualize distal time (Hurford 2007: 71–83), but rarely communicate about them (c.f. orangutans, Lameira and Call 2018; and honeybees, von Frisch 1967).

To test these causal components, this study focuses on two arenas motivated by anthropological theories. First, we replicate the arena tested by Irvine and Roberts (2016) involving collaborative building (Arena A), where individuals need to co-ordinate to construct a complex entity such as shelters or boats, which may require asymmetry of specialize roles and division of labour (see e.g. Parker 1985; Coupé and Hombert 2002; Cuthbertson and McCrohon 2012). However, the meanings in this arena were not distal, and indeed the participants did not invent symbolic referential signals. Therefore, we use this arena as a baseline and compare it to a second arena involving fire maintenance (Arena B). Humans began using fire long before being able to create it, ‘borrowing’ fire from natural sources and keeping it supplied (Twomey 2013). Even today, fire *making* technology is not universal in human societies (McCauley, Collard, and Sandgathe 2020). This suggests that fire-making knowledge may be specialized, and therefore exhibit asymmetries between individuals. Furthermore, keeping a fire lit and fuelled creates pressures to organize and divide labour between individuals (Twomey 2014). This could provide the right pressure to start referring to distant

locations (sources of fuel) or points in time (agreements about tending the fire). We are not claiming that human language emerged only because of our interactions with fire, since non-human species regularly encounter wild-fire (Pruetz and Herzog 2017), and this affects their foraging and predation strategies (Herzog et al. 2014, 2016; Doherty et al. 2022). Some species also influence the frequency and spread of fires (Foster et al. 2020), with rare cases of intentional fire maintenance (e.g. by firehawk raptors, Bonta et al. 2017). Rather, we suggest that it was one factor that contributed to a rare combination of several critical environmental, social and economic conditions. For example, fire maintenance is also tied to domestication, cognitive control and cooperation (Twomey 2019). Further connections have not been commented on in the language evolution literature, as far as we know. For example, hunting animals provides bones that can be used as fuel for fires (though evidence of early use is debated, Roebroeks and Villa 2011; McCauley, Collard, and Sandgathe 2020). Wildfires can change food foraging opportunities (Parker et al. 2016) and fire can be used to reduce predation risk (Jaffe and Isbell 2009; Wiessner 2014; Geary et al. 2020), a motivating factor in several theories.

4. Arena A: collaborative building

4.1 Methods

Following Irvine and Roberts (2016), an arena was set up in *Minecraft*, a video game, which has various features that can simulate important aspects of the arenas; we intend to study such as the ability to collaboratively build shelters, fire dynamics, collectable fuel, and hearths that burn fuel. We created a modified version of *Minecraft* (link to code provided after review) that controlled relevant game properties (e.g. no enemies, constant daylight, and logging of participant activity). The setup was as follows:

4.1.1 Environment

A flat field with markers showing the outline of a building.

4.1.2 Task

Two participants needed to follow a plan to build an abstract ‘building’ from coloured blocks. Participants were not allowed to speak, but they could knock on the table or ‘gesture’ via their avatar’s movements in the game. Participants were given up to 20 min to complete the task. We had considered timing the participants until they completed the task, but we wanted to make this condition comparable to others. We reasoned that the amount of time the participants had to

potentially develop a communication system should be kept constant to make conditions comparable.

4.1.3 Asymmetry of information

Each participant had half of the plan of the building (Fig. 1). The plan was not predictable, meaning that participants had to communicate to each other the location and colours of blocks.

4.1.4 Division of labour

The plan included four colours of blocks, but each participant was only able to place two colours. This is analogous to individuals being specialized in the use of specific building materials. Participants were allowed to destroy blocks of any colour (analogous to generalized correction of construction mistakes).

4.1.5 Evaluation criteria

Establishing an agreed symbolic convention for referring to different block colours where the difference in the form of the signal aligns with a difference in referent.

Ethical approval for all experiments was granted by the research ethics committee for the School of English, Communication and Philosophy at Cardiff University (SREC reference 221123ENCAPKrykoniukRoberts). Sixty-six participants were recruited (eleven pairs × three conditions, see [Supplementary materials](#) for more information). Participants were trained in the basic controls for *Minecraft* and built a small test structure from a plan to check their understanding (see [Supplementary materials](#)). Each participant’s screen was digitally recorded. After the main task, participants completed a questionnaire about their communication strategy and were then informally interviewed. The videos, interviews, and participant questionnaires were analysed for various categories of communication strategies. A pair was considered to have established an SRS if both participants’ questionnaires reported the same signal for referring to a block colour, or if the interview revealed they had developed such a signal. The signals could be in any modality (knocking, jumping, spinning etc.). Other strategies were identified from the data rather than being assumed *a priori* and are described below.

Three conditions were tested. The main ‘four colours’ condition was as described above. However, an alternative explanation for Irvine and Roberts (2016) results might be that, rather than the critical barrier being pointing *per se*, it was the low cost of the trial-and-error strategy. Alternatively, the system of meanings might have been too simple to require dedicated signals (see Nowak, Plotkin, and Jansen 2000; Swarup and Gasser 2010). To test these alternative

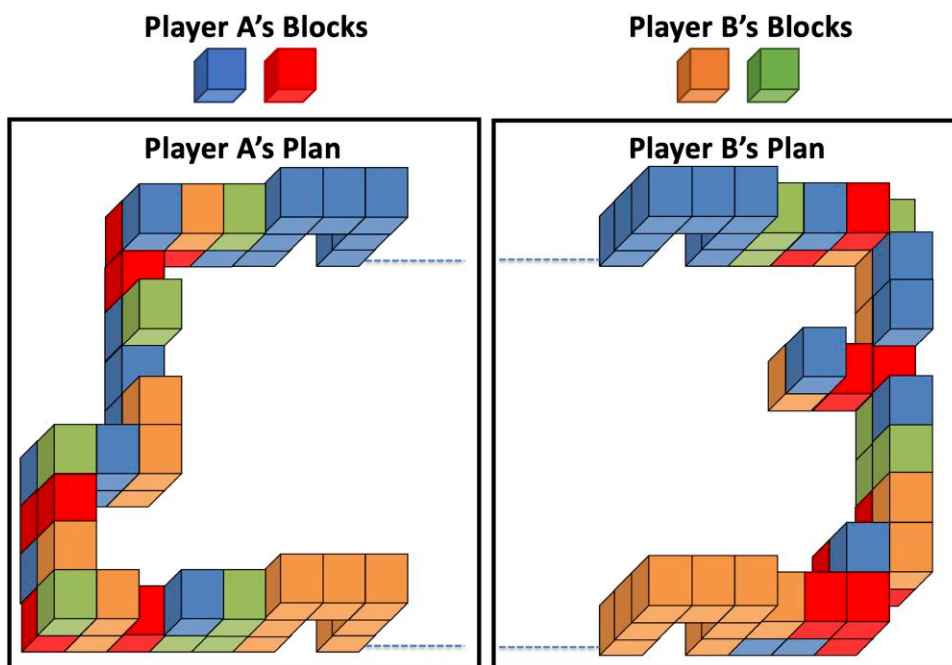


Figure 1. Top-down isometric view of the plan for the building the players had to build. The two 'gates' (blue, above and orange, below) were placed into the world for the participants in the correct positions.



Figure 2. A screenshot from the experiment showing a completed building, with a landmark hill in the background.

explanations, two more conditions were tested: one where destroying a block took three times longer to reduce the effectiveness of trial-and-error ('hard blocks' condition), and one with a total of eight differently coloured block types shared between the players to increase the complexity of the meaning space ('eight colours' condition, see [Supplementary materials](#)).

4.2 Results

Every pair of participants succeeded in establishing a strategy to solve the task (e.g. [Fig. 2](#)). The typical procedure was that each participant would start by building some portion of their own side of the building. Then, they would realize that they needed the help of their partner and seek their attention. Pairs typically

built one side at a time, so one participant would take the role of ‘director’, indicating locations and colours, and the other taking the role of the ‘builder’ who placed the blocks (Fig. 3). Participants used pointing to identify locations and trial and error to identify block colour: a director would indicate a place for the builder to place a block, but destroy it if it was the wrong type.

To assist this strategy, various communication sub-strategies emerged (Table 1). Most pairs established conventions for signalling ‘correct’ or ‘incorrect’, usually motivated by the real-world convention of nodding and shaking the head. While this is a symbolic convention, it does not refer to objects in the game world, therefore we did not categorize them as symbolic referential signals. Two sub-strategies used a feature of *Minecraft* where players can see which colour blocks their partner is currently holding. In the main condition, the most frequent sub-strategy was for the director to switch the blocks they were holding as a cue for the builder to switch their own blocks. While this relies on an analogy, the signal’s meaning is ‘change your block’ or just ‘incorrect’. It was highly contextual and could not be used to refer to a specific colour in a different context.

In the ‘hard blocks’ condition where blocks took longer to destroy, the dominant sub-strategy shifted to a similar system, but this time the builder would switch to holding a different block type and wait for confirmation from the director that it was correct. This is essentially the same as pointing at a candidate object, and is a logical strategy to adopt when the cost of placing the wrong block colour is higher.

In the ‘eight colour’ condition, the dominant strategy was an indexical one. To indicate the colour of the next block, participants pointed to existing blocks to a set of ‘reference blocks’ that they had placed to one side of the main building. This strategy relies solely on pointing and is more efficient than trial and error, which requires more guesses on average with a larger number of possible referents.

In summary, the condition affected the dominant strategy that emerged (Fisher’s exact test of director/builder/indexical strategy frequency, $P = 0.038$), indicating that the task demands were sufficiently different to motivate different communication strategies. However, only one pair in each condition established SRSs. In one case in the ‘four colours’ condition, the pair established a system immediately before doing anything else. One participant knocked once holding a blue block, knocked twice holding a red block, then knocked once holding a blue block. This redundant repetition signalled an ostensive action. Their partner understood the idea and did the same with their blocks. This process took only 17 s. This illustrates that, while establishing SRSs is clearly possible, the arena does not provide sufficient

motivation for this to emerge frequently. Indeed, in many cases, participants reported that they had considered establishing SRSs, but decided that it was not worth the effort. In one condition, the pairs even managed to complete much of the task without communicating directly: one participant placed blocks randomly and the other destroyed incorrectly placed blocks.

After the main part of the experiment, participants rated the usefulness of pointing and knocking using a five-point Likert scale (three observations excluded due to participants not answering the question). These participant-level data were analysed using Bayesian mixed effects modelling controlling for participant pair. The β statistics reported below represent the relevant coefficient estimate and confidence intervals (see the [Supplementary materials](#) for full details). There was no difference in ratings between conditions for pointing (β estimates straddle zero) or knocking (β estimates straddle zero). On average, pointing is rated as more useful than knocking [average rating for pointing = 4.09, for knocking = 2.75, $\beta = 1.55$ (0.55, 2.59)]. Furthermore, there was a negative correlation between ratings for knocking and pointing [Kendall’s rank correlation = -0.24 , $\beta = -0.34$ (-0.66 -0.02)], suggesting that the usefulness of pointing limits the usefulness of symbolic communication. Interestingly, there was evidence for an inverse relationship between the usefulness of symbolic signalling and player experience. Less experienced players rated knocking as more useful than more experienced players on average (though this was not significant) and experiments with players who had never played *Minecraft* before were more likely to attempt to establish SRSs (6 per cent of experiments where both players had played before, 11 per cent of experiments where one player had played before, and 60 per cent of experiments where both players had played before; 37 per cent of players who had never played attempted an SRS compared with 8.9 per cent of experienced players, Fisher’s exact $P = 0.029$).

In summary, in the building arena, the ability to point at objects makes an SRS redundant. This suggests that an arena that motivates the emergence of an SRS needs to involve meanings that cannot be pointed to.

5. Arena B: fire maintenance

The theories about fire maintenance above inspired an alternative arena where participants had to collect raw materials and ‘smelt’ them into refined materials.

5.1 Methods

Forty-four new participants were recruited (eleven pairs $\times$ two conditions, see [Supplementary materials](#)).



Figure 3. A screenshot from the experiment, from the perspective of Player B, looking at Player A (labelled ‘ExperPlayer3’) who is in the process of destroying a block.

Table 1. Strategies adopted in each condition of the building arena (dominant strategy in bold). Strategies are not mutually exclusive.

Condition	Convention for ‘correct’/‘incorrect’ (%)	Director switches blocks (%)	Builder switches blocks (%)	Indexical system (%)	Established SRS (%)
Four colours	91	55	27	18	9
Hard blocks	73	9	45	9	9
Eight colours	82	18	27	64	9

There was no difference in experience with Minecraft between arenas (30 per cent of participants in the building arena had never played before, compared with 41 per cent in the fire arena, Fisher’s exact test $P=0.43$). The arena was constructed as follows:

5.1.1 Environment

A narrow strip of land between a lake and a sheer mountain. A furnace was placed at one end near a source of fuel, and a ‘mine’ was placed at the other end with a source of gold ore and green ore (Fig. 4).

5.1.2 Task

Smelt as many ores into ‘ingots’ as possible by adding ore and fuel to the furnace.

5.1.3 Asymmetry of information

The ‘smelter’ had a set of cards that indicated the order in which ingots should be produced. This was not observable to the ‘miner’.

5.1.4 Division of labour

The smelter was taught to use the furnace, and the miner was taught to obtain ore from the mine (though the smelter was not prohibited from going to the mine and the miner was not prohibited from using the furnace).

5.1.5 Evaluation criteria

Establishing an agreed symbolic convention for referring to different block types where the difference in the form of the signal aligns with a difference in referent.

The rest of the methods were identical to Arena A. The expected optimal strategy was for the smelter to communicate the type of ore required to the miner, who would then retrieve the ore while the smelter managed fuel, monitored the furnace and produced the ingots before the miner’s return. The arena was designed so that the time it took to get ore was roughly equal to the time to ‘maintain’ the furnace and produce an ingot. This meant that, if the participants were acting



Figure 4. Layout of the arena based on theories of fire maintenance.

Table 2. The frequency of established and attempted SRSs in each arena. A trial with an established SRS implies a trial with an attempted SRS.

Arena	Number of trials	Established SRS	Attempted SRS
Building (all conditions)	33	3 (9%)	4 (12%)
Fire maintenance	11	4 (36%)	9 (82%)
Fire maintenance (short path)	11	1 (9%)	1 (9%)

efficiently, the smelter would remain at the furnace and both participants would not be in the same place with both types of ore, avoiding an opportunity for pointing at a required object. The analogue in the real world might be needed to tell someone to collect a specific type of fuel while they kept a fire going. A control condition moved the mine to be next to the furnace (Fig. 4). We also ran four pilot experiments, which helped refine the design of the experiment and which we document and discuss elsewhere (Roberts et al. in press).

5.2 Results

Four trials were excluded from the analysis (in two cases, participants used spoken English to announce a symbolic strategy to their partner before the experimenter finished explaining the rules, see below; in one trial, one of the participants did not engage with the task; and in one trial a participant revealed that they had taken part in the building condition). These trials were repeated with new participants to yield eleven trials in each condition.

Table 2 shows the number of trials with an established SRS for the building arena and the two fire maintenance conditions. In the main fire maintenance condition, 36 per cent of pairs established SRSs. This is four times more frequent than in the building arena though categorically not significant (Fisher's exact $P=0.053$). More tellingly, at least one participant in each pair attempted to establish an SRS in 86 per cent of trials. This is significantly higher compared with 12 per cent of trials in the building arena (Fisher's exact $P=0.0001$). Only two trials in the main fire condition did not attempt to establish an SRS, and in one of them a smelter reported that they would have done but thought that they were not allowed to go to the mine. In fact, we had to exclude two trials from the data in this condition because participants violated the experiment rules and shouted out a symbolic referential strategy to their partner before the trial began. This is clear evidence that the arena created a pressure to establish an SRS.

Although there are many differences between the building arena and the fire arena, contextually distal meanings were a key explanatory factor. This can be seen by comparing the two fire arena conditions, which were identical except for the distance of the mine. In the short path condition, only 9 per cent of trials established or attempted an SRS, significantly lower than in the main condition (Fisher's exact $P=0.002$). The interviews at the end of the trial corroborated that the distance to the mine was the reason for the difference in communicative strategy. When asked about whether they needed SRSs, participants in the short path condition frequently stated that they did not need one because they could point at the targets. Furthermore, when we explained to the short path participants that the mine was much further away in the main condition,

several pairs spontaneously suggested that they would have needed to establish SRSs if they had been in that condition.

An SRS was usually established by both participants going to the mine, pointing to a type of ore and producing a signal with a knock. That is, pointing was important to help ground the signals, but the necessity of signals was created by the conditions of the arena.

Pointing was rated as less effective in the long path fire arena compared with the building arena (mean in building arena = 4.09, mean in fire arena = 3.05, $\beta = -1.48$ [$-2.54, -0.45$]), but there were no differences in the ratings for knocking between the two arenas (β estimates straddle zero). Unlike the building arena, there was no significant difference in ratings for pointing and knocking in the fire arena (β estimates straddle zero). Taken together, this suggests that the key difference is that the fire arena makes pointing less effective, rather than making knocking/signalling more effective.

6. Discussion

The building arena consistently failed to motivate symbolic referential signals. Instead, participants found creative solutions involving pointing. Manipulating the opportunity cost and the number of meanings changed the secondary communication strategy but did not affect the likelihood of referential symbols emerging. Therefore, it seems that it is the ability to point at referents which inhibits the emergence of symbolic communication (in line with Irvine 2016). This suggests that archaeological evidence of complex collaborative constructions may not be good evidence of symbolic language (see also, Cuthbertson and McCrohon 2012). In contrast, SRSs frequently emerged in the fire arena, and participants were more likely to report feeling a need for SRSs. Changing the fire arena so that the referents were nearby reduced the likelihood of an SRS back to the same level as the building arena.

We suggest the explanatory causal model in Fig. 5. Conditions of the arena create pressures for specific types of social interaction, and these change the effectiveness of various communication strategies. The core of the model is that asymmetry of information and division of labour create a need for communication, which in turn increases the effectiveness of a pointing or SRS strategy (and hence a likelihood of one emerging). Asymmetry of information was provided by dividing the building plan in the building arena and by specialized training in the fire arena, with exclusive knowledge of the ore sequence restricted to the smelter. Division of labour was motivated by having specialized blocks in the building arena but was more complex in the fire arena, involving specialized training, a need to monitor the

fire, and a distant mine. The likelihood of establishing an SRS is affected by the need for communication, how much time the SRS costs to set up, and the presence of an effective pointing strategy is. If pointing can solve the task, this directly prevents the need for SRSs and indirectly increases the opportunity cost of setting one up. Increasing the number of blocks was predicted to reduce the effectiveness of trial-and-error, but it also increased the cost of setting up SRSs. Increasing the hardness of blocks was predicted to increase the cost of mistakes, so reduce the relative cost of setting up SRSs since both pointing and symbols would provide more confidence. However, apparently it did not reduce the relative effectiveness of the pointing system enough for a symbolic system to emerge. In contrast, the fire arena created distal meanings: the most efficient task solution involved the smelter needing to request an item when it was not immediately present. This reduced the effectiveness of pointing, reducing the relative cost of setting up a symbolic system (compared with both participants travelling to the mine), and motivated participants to establish SRSs. When the distance to the mine was reduced, the meanings were no longer distant, and participants were no longer motivated to attempt an SRS.

However, it is clear that the fire arena does not guarantee the establishment of an SRS. Furthermore, the window of time where participants were motivated to establish a symbolic system was quite narrow. Participants were unlikely to attempt to establish one after the first few rounds of smelting because at this point the smelter had access to refined referents that they could point to in order to request the next raw ore (perhaps analogous to pointing to stored or partially burned fuel). This is still compatible with the model: when the meanings were rendered non-distal, the likelihood of an SRS reduced. Similarly, in one condition the participants switched roles: the smelter went to get the ores (since they knew which ores to get), and the miner did the smelting (having observed the smelter using the forge, analogous to observational learning). In this case, there was still an asymmetry of information, but there was no division of labour since there was no longer any need to share the information. Accordingly, an SRS did not emerge, which is also compatible with the causal model. Still, these creative solutions by participants highlight that there may be other core causal components (e.g. disruptions to the task, the need for role-switching), or some aspect that prevents observation learning (e.g. complex or invisible manufacturing steps).

7. Conclusion

We presented a common task framework approach to investigating the evolution of symbolic communication.



Figure 5. A causal model of pressures that influence the emergence of symbolic referential signals. Green: positive effect, red: negative effect (lightness is only to distinguish different arrows).

Causal inference tools were used to formally relate theories and identify critical differences. The resulting causal graphs were used to design a common task framework to evaluate the theories against each other. This involved observing participants under simulated conditions and listening to their feedback in order to identify hidden assumptions and issues with the theories. The participants were active cocreators of research knowledge in the senses defined by Bonney et al. (2009), an approach rarely seen in experimental linguistic science (for further discussion, see Roberts et al. in press).

We suggest that theories related to fire maintenance include core causal components for the emergence of symbolic communication. We presented evidence from two arenas that both included asymmetry of information and division of labour. However, the fire arena also included contextually distal meanings and was more likely to motivate the emergence of SRSs. These findings suggest that a sound theory about the arena of language evolution needs to have at least these three elements.

Of course, insights from this approach need to be integrated with aspects such as cooperation, which we have assumed is present in these arenas, and other theories of the emergence of displacement (Zhang, Shi, and Benítez-Burraco 2024). Importantly, the fire arena in this study led to an established SRS about a third of the time. It is likely that there are other arenas that motivate the evolution of symbols to a greater extent, perhaps including negotiating the division of labour, teaching, or the need to refer to meanings that distant in time. For example, Sterelny (2012; see also Planer and Sterelny 2021) suggests that the potential rewards

of hunting big game would provide pressures to cooperate and communicate. It involves an asymmetry of information (skilled and unskilled hunters, knowing the location of prey), division of labour (flushing and ambushing), and distal meanings (distal prey or scheduling in time). Conditions could vary the number of participants or the speed of the animal. Fast animals might encourage strategies involving stealth or splitting the group into different roles, requiring communication about locations, time and coordinated action. We note that recent studies of cooperative hunting in wild chimpanzees have found that vocal signals are important for coordinating action (Mine et al. 2022). The common task framework can also test these arenas and compare the results to evaluate the theoretical soundness of each proposal. We hope other researchers take up the common task challenge to refine theories of how symbolic signals emerged. Indeed, because many questions about the evolution of language involve complex causal connections from multiple sources of evidence, and where direct experimentation is impossible, we hope that the common task framework will be a useful more generally for the field of evolutionary linguistics.

Acknowledgements

We thank the participants for their creative insights.

Supplementary data

Supplementary data is available at *Journal of Language Evolution* online.

Funding

This project was supported by an Arts and Humanities Research Council grant AH/T006927/1.

Data availability

Videos of the experiments are available online: For reviewers this is spread over three OSF repositories due to file size limitations: <https://doi.org/10.17605/OSF.IO/SMRB9>; <https://doi.org/10.17605/OSF.IO/85AZS>; <https://doi.org/10.17605/OSF.IO/A8Q3N>.

References

- Adornetti, I. (2015) 'Competition and Cooperation in Language Evolution: A Comparison Between Communication of Apes and Humans', in D'Errico, F., (eds.) *Conflict and Multimodal Communication: Social Research and Machine Intelligence*, pp. 91–101. Cham: Springer.
- Allen, N. J., et al. (2011) *Early Human Kinship: From Sex to Social Reproduction*. Malden, MA: John Wiley & Sons.
- Apicella, C. L., and Silk, J. B. (2019) 'The Evolution of Human Cooperation', *Current Biology*, 29: R447–50. <https://doi.org/10.1016/j.cub.2019.03.036>
- Bednarik, R. G. (1997) 'The Origins of Navigation and Language', *Artefact*, 20: 16. <https://doi.org/10.3316/informit.338874718097625>
- Bonney, R., et al. (2009) *Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education*. Washington, DC: Center for Advancement of Informal Science Education (CAISE).
- Bonta, M., et al. (2017) 'Intentional Fire-Spreading by "Firehawk" Raptors in Northern Australia', *Journal of Ethnobiology*, 37: 700–18. <https://doi.org/10.2993/0278-0771-37.4.700>
- Botha, R., and Knight, C., eds. (2009) *The Cradle of Language*. Oxford: OUP.
- Braitenberg, V. (1986) *Vehicles: Experiments in Synthetic Psychology*. Cambridge, MA: MIT Press.
- Bun, J. (2016). 'Novel Communication Systems in Minecraft', Master thesis, Tilburg University. <http://arno.uvt.nl/show.cgi?fid=141621>
- Burkart, J. M., Hrdy, S. B., and Van Schaik, C. P. (2009) 'Cooperative Breeding and Human Cognitive Evolution', *Evolutionary Anthropology*, 18: 175–86. <https://doi.org/10.1002/evan.20222>
- Chung, T. H., Orehhov, V., and Maio, A. (2023) 'Into the Robotic Depths: Analysis and Insights From the DARPA Subterranean Challenge', *Annual Review of Control, Robotics, and Autonomous Systems*, 6: 477–502. <https://doi.org/10.1146/annurev-control-062722-100728>
- Coupé, C., and Hombert, J. (2002) 'Language at 70,000 BP: Evidence From Sea-crossings', in 4th International Conference on the Evolution of Language (EVOLANG4), Cambridge, MA: World Scientific.
- Cuthbertson, C., and McCrohon, L. (2012) 'Sea Crossings are an Unreliable Indicator of Language Ability in Hominids', in Scott-Phillips, T. C., et al. (eds.) *The Evolution of Language*, pp. 88–95. London: World Scientific.
- Davidson, I., and Noble, W. (1989) 'The Archaeology of Perception: Traces of Depiction and Language', *Current Anthropology*, 30: 125–55. <https://doi.org/10.1086/203723>
- Dediu, D., and Levinson, S. C. (2018) 'Neanderthal Language Revisited: Not Only Us', *Current Opinion in Behavioral Sciences*, 21: 49–55. <https://doi.org/10.1016/j.cobeha.2018.01.001>
- Doherty, T. S., et al. (2022) 'Fire as a Driver and Mediator of Predator–Prey Interactions', *Biological Reviews of the Cambridge Philosophical Society*, 97: 1539–58. <https://doi.org/10.1111/brv.12853>
- Dor, D., Knight, C., and Lewis, J. (2014) *The Social Origins of Language*. Oxford: OUP.
- Dunbar, R. I., et al. (2014) 'The Road to Modern Humans: Time Budgets, Fission-Fusion Sociality, Kinship and the Division of Labour in Hominin Evolution', in Dunbar, R. I. M. (ed.) *Lucy to Language: The Benchmark Papers*, pp. 333–55.
- Fisher, S. E., and Vernes, S. C. (2015) 'Genetics and the Language Sciences', *Annual Review of Linguistics*, 1: 289–310. <https://doi.org/10.1146/annurev-linguist-030514-125024>
- Fitch, W. T. (2010) *The Evolution of Language*. Cambridge, PA: CUP.
- Fitch, W. T. (2020) 'Animal Cognition and the Evolution of Human Language: Why we Cannot Focus Solely on Communication', *Philosophical Transactions of the Royal Society B*, 375: 20190046. <https://doi.org/10.1098/rstb.2019.0046>
- Foster, C. N., et al. (2020) 'Animals as Agents in Fire Regimes', *Trends in Ecology & Evolution*, 35: 346–56. <https://doi.org/10.1016/j.tree.2020.01.002>
- Gabrić, P. (2022) 'Overlooked Evidence for Semantic Compositionality and Signal Reduction in Wild Chimpanzees (Pan Troglodytes)', *Animal Cognition*, 25: 631–43. <https://doi.org/10.1007/s10071-021-01584-3>
- Galantucci, B. (2009) 'Experimental Semiotics: A New Approach for Studying Communication as a Form of Joint Action', *Topics in Cognitive Science*, 1: 393–410. <https://doi.org/10.1111/j.1756-8765.2009.01027.x>
- Geary, W. L., et al. (2020) 'Predator Responses to Fire: A Global Systematic Review and Meta-Analysis', *The Journal of Animal Ecology*, 89: 955–71. <https://doi.org/10.1111/1365-2656.13153>
- Gontier, N. (2013) 'Pointing and the Evolution of Language: An Applied Evolutionary Epistemological Approach', *Humana Mente Journal of Philosophical Studies*, 6: 1–26. <https://www.humanamente.eu/index.php/HM/article/view/148>
- Greenfield, P., Lyn, H., and Savage-Rumbaugh, S. E. (2008) 'Protolanguage in Ontogeny and Phylogeny: Combining Deixis and Representation', *Interaction Studies*, 9: 34–50. <https://doi.org/10.1075/is.9.1.04gre>
- Hashimoto, T. (2020) 'The Emergent Constructive Approach to Evolving Linguistics: Considering Hierarchy and Intention Sharing in Linguistic Communication', *Journal of Systems Science and Systems Engineering*, 29: 675–96. <https://doi.org/10.1007/s11518-020-5469-x>

- Herzog, N. M., et al. (2014) 'Fire and Home Range Expansion: A Behavioral Response to Burning Among Savanna Dwelling Vervet Monkeys (*Chlorocebus aethiops*)', *American Journal of Physical Anthropology*, 154: 554–60. <https://doi.org/10.1002/ajpa.22550>
- Herzog, N. M., et al. (2016) 'What's Burning Got to Do With It? Primate Foraging Opportunities in Fire-Modified Landscapes', *American Journal of Physical Anthropology*, 159: 432–41. <https://doi.org/10.1002/ajpa.22885>
- Hill, R. A., and Dunbar, R. I. (1998) 'An Evaluation of the Roles of Predation Rate and Predation Risk as Selective Pressures on Primate Grouping Behaviour', *Behaviour*, 135: 411–30. <https://doi.org/10.1163/156853998793066195>
- Hockett, C. F. (1963) 'The Problem of Universals in Language', in Greenberg, J. (ed.) *Universals of Language*, pp. 1–29. Cambridge: MIT Press.
- Hurford, J. R. (1989) 'Biological Evolution of the Saussurean Sign as a Component of the Language Acquisition Device', *Lingua*, 77: 187–222. [https://doi.org/10.1016/0024-3841\(89\)90015-6](https://doi.org/10.1016/0024-3841(89)90015-6)
- Hurford, J. R. (1990) 'Nativist and Functional Explanations in Language Acquisition', *Logical Issues in Language Acquisition*, 85: 136. <https://doi.org/10.1515/9783110870374-007>
- Hurford, J. R. (2007) *The Origins of Meaning: Language in the Light of Evolution, Volume 1*. Oxford: OUP.
- Irvine, E. (2016) 'Method and Evidence: Gesture and Iconicity in the Evolution of Language', *Mind & Language*, 31: 221–47. <https://doi.org/10.1111/mila.12102>
- Irvine, E., and Roberts, S. G. (2016), 'Deictic Tools Can Limit the Emergence of Referential Symbol Systems' in Roberts, S. G., et al. (eds.) *The Evolution of Language: Proceedings of the 11th International Conference (EVOlang11)*, pp. 99–100. New Orleans, LA: EvoLang Scientific Committee. <http://evolang.org/neworleans/papers/99.html>
- Irvine, L., Roberts, S., and Kirby, S. (2013) 'A Robustness Approach to Theory Building', *Proceedings of the Annual Meeting of the Cognitive Science Society*, 35: 2614–9. <https://escholarship.org/uc/item/1rw611dp>
- Jaffe, K., and Isbell, L. A. (2009) 'After the Fire: Benefits of Reduced Ground Cover for Vervet Monkeys (*Cercopithecus aethiops*)', *American Journal of Primatology*, 71: 252–60. <https://doi.org/10.1002/ajp.20644>
- Kalan, A. K., Nakano, R., and Warshawski, L. (2023) 'What We Know and Don't Know About Great Ape Cultural Communication in the Wild', *American Journal of Primatology*, 87: e23560. <https://doi.org/10.1002/ajp.23560>
- Kazakov, D., and Bartlett, M. (2004) 'Cooperative Navigation and the Faculty of Language', *Applied Artificial Intelligence*, 18: 885–901. <https://doi.org/10.1080/08839510490509072>
- Kirby, S., et al. (2015) 'Compression and Communication in the Cultural Evolution of Linguistic Structure', *Cognition*, 141: 87–102. <https://doi.org/10.1016/j.cognition.2015.03.016>
- Knight, C., and Lewis, J. (2017) 'Wild Voices: Mimicry, Reversal, Metaphor, and the Emergence of Language', *Current Anthropology*, 58: 435–53. <https://doi.org/10.1086/692905>
- Knight, C., Power, C., and Watts, I. (1995) 'The Human Symbolic Revolution: A Darwinian Account', *Cambridge Archaeological Journal*, 5: 75–114. <https://doi.org/10.1017/S0959774300001190>
- Knight, C., Studdert-Kennedy, M., and Hurford, J., eds. (2000) *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*. Cambridge, UK: CUP.
- Krause, M. A., et al. (2018) 'Animal Pointing: Changing Trends and Findings From 30 Years of Research', *Journal of Comparative Psychology*, 132: 326–45. <https://doi.org/10.1037/com0000125>
- Lameira, A. R., and Call, J. (2018) 'Time-space-displaced Responses in the Orangutan Vocal System', *Science Advances*, 4: eaau3401. <https://doi.org/10.1126/sciadv.aau3401>
- Leavens, D. A. (2021) 'The Referential Problem Space Revisited: An Ecological Hypothesis of the Evolutionary and Developmental Origins of Pointing', *Wiley Interdisciplinary Reviews: Cognitive Science*, 12: e1554. <https://doi.org/10.1002/wcs.1554>
- Leavens, D. A., Hopkins, W. D., and Bard, K. A. (2005) 'Understanding the Point of Chimpanzee Pointing: Epigenesis and Ecological Validity', *Current Directions in Psychological Science*, 14: 185–9. <https://doi.org/10.1111/j.0963-7214.2005.00361.x>
- Leavens, D. A., Russell, J. L., and Hopkins, W. D. (2010) 'Multimodal Communication by Captive Chimpanzees (*Pan troglodytes*)', *Animal Cognition*, 13: 33–40. <https://doi.org/10.1007/s10071-009-0242-z>
- Liebal, K., and Oña, L. (2018) 'Different Approaches to Meaning in Primate Gestural and Vocal Communication', *Frontiers in Psychology*, 9: 478. <https://doi.org/10.3389/fpsyg.2018.00478>
- Liszkowski, U., et al. (2012) 'A Prelinguistic Gestural Universal of Human Communication', *Cognitive Science*, 36: 698–713. <https://doi.org/10.1111/j.1551-6709.2011.01228.x>
- Lyn, H., and Christopher, J. L. (2020) 'How Environment Can Reveal Semantic Capacities in Nonhuman Animals', *Animal Behavior and Cognition*, 7: 159–67. <https://doi.org/10.26451/abc.07.02.10.2020>
- Lyn, H., et al. (2024) 'Do Dogs Really Get the Point?', in Nölle, J., et al. (eds.) *The Evolution of Language: Proceedings of the 15th International Conference (EVOlang XV)*, pp. 363–4. Nijmegen, The Netherlands: The Evolution of Language Conferences. <https://doi.org/10.17617/2.3587960>
- Macuch Silva, V., et al. (2020) 'Multimodality and the Origin of a Novel Communication System in Face-to-Face Interaction', *Royal Society Open Science*, 7: 182056. <https://doi.org/10.1098/rsos.182056>
- Macuch Silva, V., and Roberts, S. (2016) 'Exploring the Role of Interaction in the Emergence of Linguistic Structure', in Roberts, S., and Mills, G. (eds.) *Proceedings of EvoLang XI, Language Adapts to Interaction Workshop*, pp. 9–10. New Orleans, LA: EvoLang Scientific Committee. http://evolang.org/neworleans/workshops/papers/LATI_9.html
- McCauley, B., Collard, M., and Sandgathe, D. (2020) 'A Cross-Cultural Survey of On-Site Fire Use by Recent

- Hunter-Gatherers', *Journal of Paleolithic Archaeology*, 3: 566–84. <https://doi.org/10.1007/s41982-020-00052-7>
- McCreary, M. K., Jones, S. V., and Kuhlmeier, V. A. (2023) 'Following the Human Point: Research With Nonhuman Animals Since Povinelli, Nelson, and Boysen (1990)', *Learning & Behavior*, 51: 34–47. <https://doi.org/10.3758/s13420-022-00546-0>
- Mine, J. G., et al. (2022) 'Vocal Signals Facilitate Cooperative Hunting in Wild Chimpanzees', *Science Advances*, 8: eabo5553. <https://doi.org/10.1126/sciadv.abo5553>
- Noble, W., and Davidson, I. (1996) *Human Evolution, Language and Mind: A Psychological and Archaeological Inquiry*. Cambridge, UK: CUP Archive.
- Nölle, J., and Galantucci, B. (2022) 'Experimental Semiotics: Past, Present and Future', in García, A. M., Ibanez, A. (eds.) *The Routledge Handbook of Semiosis and the Brain*, pp. 4. Philadelphia, PA: Routledge.
- Nowak, M. A., Plotkin, J. B., and Jansen, V. A. A. (2000) 'The Evolution of Syntactic Communication', *Nature*, 404: 495–8. <https://doi.org/10.1038/35006635>
- Outram, A. K. (2008) 'Introduction to Experimental Archaeology', *World Archaeology*, 40: 1–6. <https://doi.org/10.1080/00438240801889456>
- Özyürek, A. (2021) 'Considering the Nature of Multimodal Language From a Crosslinguistic Perspective', *Journal of Cognition*, 4: 42. <https://doi.org/10.5334/joc.165>
- Parker, S. T. (1985) 'A Social-Technological Model for the Evolution of Language', *Current Anthropology*, 26: 617–26. <https://doi.org/10.1086/203350>
- Parker, C. H., et al. (2016) 'The Pyrophilic Primate Hypothesis', *Evolutionary Anthropology: Issues, News, and Reviews*, 25: 54–63. <https://doi.org/10.1002/evan.21475>
- Pearl, J., and Mackenzie, D. (2018) *The Book of Why: The New Science of Cause and Effect*. New York, NY: Basic Books.
- Pepperberg, I. M. (2021) 'Nonhuman and Nonhuman-Human Communication: Some Issues and Questions', *Frontiers in Psychology*, 12: 647841. <https://doi.org/10.3389/fpsyg.2021.647841>
- Planer, R., and Sterelny, K. (2021) *From Signal to Symbol: The Evolution of Language*. Cambridge, MA: MIT Press.
- Povinelli, D. J., Bering, J. M., and Giambrone, S. (2003) 'Chimpanzees "Pointing": Another Error of the Argument by Analogy?', in Kita, S. (ed.) *Pointing*, pp. 43–76. New York, NY: Psychology Press.
- Power, C., Finnegan, M., and Callan, H. (2016) *Human Origins: Contributions from Social Anthropology*. Oxford, UK: Berghahn Books.
- Premack, D. (1971) 'Language in Chimpanzee', *Science*, 172: 808–22. <https://doi.org/10.1126/science.172.3985.808>
- Pruetz, J. D., and Herzog, N. M. (2017) 'Savanna Chimpanzees at Fongoli, Senegal, Navigate a Fire Landscape', *Current Anthropology*, 58: S337–50. <https://doi.org/10.1086/692112>
- Raviv, L., Meyer, A., and Lev-Ari, S. (2019) 'Larger Communities Create More Systematic Languages', *Proceedings of the Royal Society of London: Series B, Biological Sciences*, 286: 20191262. <https://doi.org/10.1098/rspb.2019.1262>
- Roberts, G. (2017) 'The Linguist's Drosophila: Experiments in Language Change', *Linguistics Vanguard*, 3: 20160086. <https://doi.org/10.1515/lingvan-2016-0086>
- Roberts, S. (2018) 'What Are the Social, Economic and Ecological Conditions for the Evolution of Complex Communication Systems?', *Physics of Life Reviews*, 26: 152–4. <https://doi.org/10.1016/j.plrev.2018.06.016>
- Roberts, S., et al. (2020) 'CHIELD: The Causal Hypotheses in Evolutionary Linguistics Database', *Journal of Language Evolution*, 5: 101–20. <https://doi.org/10.1093/jole/lzaa001>
- Roberts, S. G., et al. (in press) 'Get Data Early, Get Data Often, Iterate Constantly: An Explorative, Participatory Approach to Studying Language Evolution', *Journal of Language Evolution*.
- Roebroeks, W., and Villa, P. (2011) 'On the Earliest Evidence for Habitual Use of Fire in Europe', *Proceedings of the National Academy of Sciences*, 108: 5209–14. <https://doi.org/10.1073/pnas.1018116108>
- Shilton, D. (2019) 'Is Language Necessary for the Social Transmission of Lithic Technology?', *Journal of Language Evolution*, 4: 124–33. <https://doi.org/10.1093/jole/lzz004>
- Sievers, C., and Gruber, T. (2016) 'Reference in Human and Non-human Primate Communication: What Does it Take to Refer?', *Animal Cognition*, 19: 759–68. <https://doi.org/10.1007/s10071-016-0974-5>
- Slocombe, K. E., and Zuberbühler, K. (2005) 'Functionally Referential Communication in a Chimpanzee', *Current Biology*, 15: 1779–84. <https://doi.org/10.1016/j.cub.2005.08.068>
- Stade, C. M. (2017) 'Lithic Morphological Variability as a Proxy for Palaeolithic Linguistic Ability: A Knapping Training Study Exploring Cultural Transmission, Theory of Mind and Language', Doctoral thesis, University of Southampton, Southampton, UK, pp. 226.
- Steels, L. (2003) 'Evolving Grounded Communication for Robots', *Trends in Cognitive Sciences*, 7: 308–12. [https://doi.org/10.1016/S1364-6613\(03\)00129-3](https://doi.org/10.1016/S1364-6613(03)00129-3)
- Sterelny, K. (2012) 'Language, Gesture, Skill: The Co-Evolutionary Foundations of Language', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367: 2141–51. <https://doi.org/10.1098/rstb.2012.0116>
- Sterelny, K. (2021) *The Pleistocene Social Contract: Culture and Cooperation in Human Evolution*. Oxford, UK: OUP.
- Swarup, S., and Gasser, L. (2010) 'The Classification Game: Combining Supervised Learning and Language Evolution', *Connection Science*, 22: 1–24. <https://doi.org/10.1080/09540090802638766>
- Számadó, S., and Szathmáry, E. (2006) 'Competing Selective Scenarios for the Emergence of Natural Language', *Trends in Ecology & Evolution*, 21: 555–61. <https://doi.org/10.1016/j.tree.2006.06.021>
- Tomasello, M. (1990) 'Cultural Transmission in Chimpanzee Tool Use and Signaling', in Parker, S. T., and Gibson, K. R. (eds.) *Language and Intelligence in Monkeys and Apes: Comparative Developmental Perspectives*, pp. 274–311. Cambridge, UK: CUP.
- Townsend, S. W., and Manser, M. B. (2013) 'Functionally Referential Communication in Mammals: The Past,

- Present and the Future', *Ethology*, 119: 1–11. <https://doi.org/10.1111/eth.12015>
- Tramacere, A., and Moore, R. (2018) 'Reconsidering the Role of Manual Imitation in Language Evolution', *Topoi*, 37: 319–28. <https://doi.org/10.1007/s11245-016-9440-x>
- Twomey, T. (2013) 'The Cognitive Implications of Controlled Fire Use by Early Humans', *Archaeological Journal*, 23: 113–28. <https://doi.org/10.1017/S0959774313000085>
- Twomey, T. (2014) 'How Domesticating Fire Facilitated the Evolution of Human Cooperation', *Biology & Philosophy*, 29: 89–99. <https://doi.org/10.1007/s10539-013-9402-2>
- Twomey, T. (2019) 'Domestic Fire, Domestic Selves: How Keeping Fire Facilitated the Evolution of Emotions and Emotion Regulation', *Handbook of Cognitive Archaeology*, pp. 415–30. New York, NY: Routledge.
- Vail, A. L., Manica, A., and Bshary, R. (2013) 'Referential Gestures in Fish Collaborative Hunting', *Nature Communications*, 4: 1765. <https://doi.org/10.1038/ncomms2781>
- von Frisch, K. (1967) *The Dance Language and Orientation of Bees*. Harvard, MA: HUP.
- Waciewicz, S., and Żywiczyński, P. (2018) 'Language Origins: Fitness Consequences, Platform of Trust, Cooperation, and Turn-Taking', *Interaction Studies*, 19: 167–82. <https://doi.org/10.1075/is.17031.wac>
- Wadley, L. (2023) 'Experimental Archaeology Enables Inferences About Human Cognition', in Wynn, T., Overmann, K. A., and Coolidge, F. L. (eds.) *The Oxford Handbook of Cognitive Archaeology*, online edn, pp. 391–410. Oxford, UK: Oxford Academic.
- Warren, E., and Call, J. (2022) 'Inferential Communication: Bridging the Gap Between Intentional and Ostensive Communication in Non-Human Primates', *Frontiers in Psychology*, 12: 718251. <https://doi.org/10.3389/fpsyg.2021.718251>
- Watson, S. K. et al. (2015) 'Vocal Learning in the Functionally Referential Food Grunts of Chimpanzees', *Current Biology*, 25: 495–9. <https://doi.org/10.1016/j.cub.2014.12.032>
- Webb, B. (2000) 'What Does Robotics Offer Animal Behaviour?', *Animal Behaviour*, 60: 545–58. <https://doi.org/10.1006/anbe.2000.1514>
- Wheeler, B. C., and Fischer, J. (2012) 'Functionally Referential Signals: A Promising Paradigm Whose Time Has Passed', *Evolutionary Anthropology: Issues, News, and Reviews*, 21: 195–205. <https://doi.org/10.1002/evan.21319>
- Wiessner, P. W. (2014) 'Embers of Society: Firelight Talk Among the Ju/'hoansi Bushmen', *Proceedings of the National Academy of Sciences*, 111: 14027–35. <https://doi.org/10.1073/pnas.1404212111>
- Zhang, E. Q., Shi, E. R., and Benítez-Burraco, A. (2024) How Displacement Might Have Evolved. PsyArXiv Preprints. <https://doi.org/10.31234/osf.io/9j8gm>.
- Zlatev, J., Żywiczyński, P., and Waciewicz, S. (2020) 'Pantomime as the Original Human-Specific Communicative System', *Journal of Language Evolution*, 5: 156–74. <https://doi.org/10.1093/jole/lzaa006>
- Żywiczyński, P., Waciewicz, S., and Sibierska, M. (2018) 'Defining Pantomime for Language Evolution Research', *Topoi*, 37: 307–18. <https://doi.org/10.1007/s11245-016-9425-9>