Perspective taking in a novel signaling task: effects of world knowledge and contextual constraint

Justin Sulik

Royal Holloway, University of London

Gary Lupyan

University of Wisconsin-Madison

Word count: 12884

All data (as well as the R scripts for statistical analyses) are available at

`https://osf.io/frkeb/`

Author Note

Justin Sulik, Department of Psychology, Royal Holloway, University of London; Gary Lupyan, Department of Psychology, University of Wisconsin–Madison.

This research was partially supported by an award from The Templeton Foundation to the Metaknowledge Research Network. J.S. gratefully acknowledges a visiting-researcher stipend from the Max Planck Institute for Psycholinguistics during the preparation of this manuscript.

Some of these results were presented in a talk at the Evolang XI Conference (New Orleans, March 2016, abstract available at http://evolang.org/neworleans/papers/103.html).

Correspondence regarding this article should be addressed to Justin Sulik, Department of Psychology, Royal Holloway, University of London, Egham Hill, Egham TW20 0EX, UK. Email: justin.sulik@gmail.com

Abstract

Perspective taking – the ability to see things from someone else's point of view – can boost success in communication. A signaler might take perspective when designing an utterance that is informative from the receiver's point of view, or the receiver might take perspective when inferring the signaler's communicative intentions. Perspective taking is supposed to play a particularly vital role when people try to communicate in the absence of a conventional signaling system. However, the task demands in such cases are extremely different from those in typical experimental approaches to perspective taking. Thus, current evidence for perspective taking does not establish whether humans can take perspective in those cases where perspective taking is arguably most helpful. We describe experimental tests of perspective taking that are suitable for settling the matter. Our task focuses on the use of shared world knowledge rather than shared visual scenes, and it is suitable for both open-ended and contextually constrained responses. We show that people generally fail at perspective taking in a novel signaling task, but that perspective taking can be boosted by contextual constraint. In that case, however, it is context, rather than perspective taking or shared world knowledge, that explains communicative success.

Keywords: perspective taking, novel signaling task, knowledge, context, coordination

Perspective taking in a novel signaling task: effects of world knowledge and contextual constraint

Introduction

When we speak, we mostly communicate using conventional signals such as words. But we can also communicate using non-conventional signals. We are able to produce and understand novel gestures (Fay, Arbib, & Garrod, 2013; Schouwstra & de Swart, 2014), pictures (Garrod, Fay, Lee, Oberlander, & MacLeod, 2007), vocal sounds (Perlman & Cain, 2014), or deictics (Misyak, Noguchi, & Chater, 2016). This includes games involving spontaneous gesture (e.g., charades) or graphical signals (e.g., Pictionary), or gesturing with people who do not share our language. Call these all novel signaling tasks.

How do humans succeed at communicating in such tasks? A plausible answer is shared world knowledge (Clark, 1996; Levinson, 2006; Lewis, 1969; Schelling, 1960). Suppose a signaler intends to communicate the idea of a snake to a recipient. If they both know what snakes look like and how they behave, the signaler might use this shared world knowledge to produce a vocalization or gesture that resembles or imitates the snake in some way. However, signalers typically have several options (Mangold & Pobel, 1988). They could vocally imitate a snake's hiss, or gesturally imitate its fangs, its biting strike, winding movement, or swaying head movement. These various pieces of knowledge are unlikely to be equally successful in getting the recipient to guess that the intended message was 'snake'. For example, a gestural representation of a snake's winding motion might well make the recipient think of a fish.

How does a signaler choose the most effective cue? A common answer is that the speaker takes the perspective of the person receiving the signal, evaluating — from the receiver's point of view (POV) — which cue is likely to lead to the correct inference (e.g., Clark & Murphy, 1982; Levinson, 2006; Sperber & Wilson, 1995). Our main aims are to test whether (and if so, when) people can take the perspective of others in a novel signaling task and whether (and if so, when) shared knowledge leads to communicative success. We show that neither shared world knowledge nor perspective taking are *general* solutions to the question of how we communicate in the absence of convention. In particular, we show that (1) in the general case, people do not succeed well at perspective taking; (2) shared world knowledge can sometimes hinder communicative success; (3) but contextual constraints can promote perspective taking, thereby boosting communicative success.

Background

Coordination, world knowledge and salience

Lewis (1969) offers a game-theoretic account of how people coordinate in the absence of convention. If we can expect people to behave a certain way, we can coordinate with them without explicit agreement. For example, if one did not know which side of the road locals drive on in a foreign country, one could simply observe their behavior and drive on the same side as they do, based on the expectation that they will continue to drive on that side. One could thus behave conventionally without explicit agreement.

But although one can directly observe what side of the road someone drives on, one cannot directly observe the communicative intentions underlying a previously unseen signal. We will refer to tasks like deciding which side of the road to drive on 'cognitively transparent' and those like deciding what a new signal means 'cognitively opaque' (Gergely & Csibra, 2003). Cognitively opaque tasks need something more than observing others' behavior. To enable coordination in such tasks, Lewis appeals to shared world knowledge, and in particular, patterns of salience in that knowledge. Levinson makes a similar point in setting out preconditions for human interaction (including communication): '[Coordination] presupposes the notion of *mutual knowledge (or common ground)* ... But it also involves a notion of *mutual salience* — what leaps out of the common ground as a solution likely to independently catch our joint attention' (2006, p. 49, emphasis Levinson's).

Such accounts rely on focal or 'Schelling' points. Schelling (1960) describes a coordination task in which people are asked where and when they would meet someone in

New York City if they had not previously made any arrangements. There are many possibilities for places and times to meet, but some will be more salient than others: they stand out in some way, or are more likely to occur to people than the other possibilities. Most of Schelling's respondents said they would choose to meet at Grand Central Station. Schelling's interpretation is that this is a salient choice. Lewis extends this claim into a larger argument that patterns of salience in shared world knowledge allow people to coordinate in the absence of convention, and Scott-Phillips, Kirby, and Ritchie (2009) highlight the role of Schelling points in novel signaling tasks.

However, even though people can use salient aspects of world knowledge to coordinate, it is not clear how general this finding is. One reason for caution is that Schelling's participants shared a significant amount of context: they were all students at Yale in New Haven, Connecticut in the 1950s, so Grand Central Station would have been where most of them arrived in New York City (Verbeek, 2008). Perhaps this shared context (the participants all being from New Haven, rather than a random sample of Americans in general) is as much a driver of success as salience. We thus aim to study the effect of context by comparing open-ended tasks (where the target could be any English word) and contextually constrained tasks (where participants know the target must be one of a small set of words).

A second reason for caution is that, although studies such as Mehta, Starmer, and Sugden (1994) confirm the role of Schelling points in achieving coordination, their tasks do not involve communication, so it is unclear whether their results extend to novel signaling tasks. We thus aim to compare behavior in communicative and non-communicative tasks.

In sum, a proposed explanation for human performance in novel signaling tasks is that we share patterns of salience in world knowledge (Levinson, 2006; Lewis, 1969; Scott-Phillips et al., 2009). However, the evidence for this cited above typically reflect constrained contexts, or non-communicative tasks (or both). One aim of the present work is thus to compare communicative and non-communicative versions of a coordination task, and to manipulate the level of contextual constraint, in order to see how these factors affect people's responses. It remains possible that having shared knowledge would not help people coordinate communicatively in the absence of a tightly constrained shared context.

Why communication presents a particular challenge

Even if salience predicted participants' *responses* (i.e., signals or guesses) in communicative as well as non-communicative tasks, it might nonetheless not contribute equally to *success* across task type. The same salience-driven response may be a good coordination strategy in one type of task but not in another. One reason for this worry is that communication introduces two asymmetries that may hinder success.

The first source of asymmetry is differing patterns of salience in world knowledge. For example, money is a salient feature of banks. When people think of banks, they are likely to think of money. However, it does not follow that banks are a salient feature of money. When people think of money, they are vastly less likely to think of banks[1]. These claims are empirically supported by word association studies (e.g., Nelson, McEvoy, & Schreiber, 2004). In such studies, participants respond with the first word they think of when given a cue. Nelson et al. found that, of over 100 participants asked to think of a word given the cue 'bank', 80% responded with 'money'. However, given cue 'money', fewer than 2% responded 'bank'. Associative norms derived from such studies thus potentially serve as an empirical yardstick for salience (and Study 1 below demonstrates that they do).

In a non-communicative task, such salience asymmetries do not necessarily pose a problem. If two participants are given cue 'bank' and asked to try coordinate by generating the same one-word response, they would probably both respond with 'money' and thus succeed. However, the people in this example had the *same* starting point, whereas

[1] While there is a salience asymmetry between banks and money, other aspects of world knowledge are more symmetric: when people think about brides, many of them are likely to also think of grooms, and when they think of grooms, they are likely to also think of brides.

participants in a novel signaling task often work in *opposite* directions: the signaler has a target meaning they want to convey, and must generate a signal based on that meaning. The receiver, on the other hand, must infer the target meaning from the signal (Sperber & Wilson, 1995). Thus, communication introduces a second source of asymmetry. It is the combination of salience asymmetry and communicative asymmetry that may be a potential block to successful coordination, since this means that what is salient to the signaler is not necessarily salient to the receiver, and vice versa, *despite* sharing world knowledge.

Thus, even if there is a strong relationship between salience and success in a non-communicative coordination task, there might be a disjunction in a communicative task. Another aim of the present work is to see whether salience guides behavior and contributes to coordination success equally across different tasks.

Perspective taking

The above asymmetries imply a difference in perspective: what is salient from the signaler's point of may not be salient from the receiver's point of view. If asymmetry hinders communicative success, then the ability to take an interlocutor's perspective could be a potential counterbalance, a way to boost success (Hanna, Tanenhaus, & Trueswell, 2003; Surtees, Apperly, & Samson, 2013; Todd, Forstmann, Burgmer, Brooks, & Galinsky, 2015). This raises the question whether people are able to take perspective in a novel signaling task by working out what is allocentrically salient (salient from their interlocutor's point of view) or whether they are typically egocentric (responding based on what is salient from their own point of view).

According to the theory of pragmatics put forward by Sperber and Wilson (1995), people *should* be able to take perspective in such a task. For them, a key factor is what they term 'accessibility'. Something is accessible insofar as one is likely to think of it. For example, in thinking about snakes, perhaps their fangs are more accessible than their cold-bloodedness[2]. In this account, receivers generate hypotheses about meaning in order of accessibility and signalers should thus (to the best of their abilities) select or structure the information they share so that their target meaning is maximally accessible to the receiver. This is called audience or recipient design (Clark & Murphy, 1982) and it presumes perspective taking since signalers must have at least a rough idea what is accessible to the recipient.

Do people actually take perspective in communicative tasks? Experimental data on the subject is mixed (Brown-Schmidt & Hanna, 2011). There is evidence supporting the use of perspective taking or audience design (Brennan & Hanna, 2009; Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Clark & Murphy, 1982; Clark, Schreuder, & Buttrick, 1983; Hanna et al., 2003; Hilliard & Cook, 2016; Metzing & Brennan, 2003; Rubio-Fernández, 2008), as well as evidence highlighting its lack, both for speakers (Epley, Keysar, Van Boven, & Gilovich, 2004; Horton & Keysar, 1996; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Barr, & Horton, 1998) and listeners (Keysar, Barr, Balin, & Paek, 1998).

Various models attempt to account for these seemingly inconsistent patterns of behavior. (1) A constraint-based model (Hanna et al., 2003) argues that allocentric information is available early on in processing, but that it is just one source of information that probabilistically drives behavior, and that it can sometimes be overridden by egocentric information. (2) According to perspective-adjustment models (Epley, Keysar, et al., 2004; Keysar et al., 2000; Wu, Barr, Gann, & Keysar, 2013), early processing is purely egocentric, and that this perspective can be adjusted to include allocentric information, but that this adjustment occurs relatively late and only if it is necessary. (3) An

[2] Lewis (1969) and Levinson (2006) use the term 'salience', while Sperber and Wilson use 'accessibility'. However, the meaning of these terms is strikingly similar: if something is salient, it stands out from the other alternatives and should be more likely to occur to people, i.e., it should be more accessible. Mehta et al. (1994) call this 'primary salience'. We show that accessibility is a good predictor of behavior in a Schelling-like task (Study 1), and thus use 'salience' to mean 'salience or accessibility'. However, Study 4 and the general discussion offer a more nuanced position and discuss other kinds of salience.

anticipation-integration model (Barr, 2008) argues that listeners can anticipate allocentric information prior to hearing an utterance, but when interpreting a heard utterance, they fail to integrate that allocentric information.

Other factors affecting perspective taking include salience (Wardlow Lane & Ferreira, 2008), time pressure (Horton & Keysar, 1996), motivation (Epley, Keysar, et al., 2004), cognitive load (Cane, Ferguson, & Apperly, 2017), mood (Converse, Lin, Keysar, & Epley, 2008; Todd et al., 2015), anomalies in input (Bögels, Barr, Garrod, & Kessler, 2015), executive control (Brown-Schmidt, 2009; Nilsen & Graham, 2009), working memory (Cane et al., 2017; Wardlow, 2013), novelty and expertise (Gann & Barr, 2014), speaker identity (Metzing & Brennan, 2003), age (Dumontheil, Apperly, & Blakemore, 2010; Epley, Morewedge, & Keysar, 2004), or cultural differences (Wu & Keysar, 2007).

In these studies, differences in perspective are typically operationalized in terms of what participants can see: a common manipulation is to occlude one side of a cubbyhole, such that its contents are only visible to one participant[3]. But even if participants can sometimes take perspective on the basis of *seeing* what the other sees (a cognitively transparent task), this does not imply they can do so on the basis of *knowing* what the other knows (a cognitively opaque task), in the absence of visual common ground. Since shared knowledge is a proposed driver of behavior in a novel signaling task, another aim of the present work is to focus on the effect of shared world knowledge, as opposed to a shared visual scene.

This focus on shared world knowledge is a methodological departure from previous research in perspective taking. Thus, we do not explicitly evaluate the particular predictions of the models outlined above, but simply test whether people spontaneously

[3] Though this is a common task design in perspective taking research, other methodologies have been used. For instance, Epley, Keysar, et al. (2004) explore perspective taking in cases of sarcasm or irony, and Langdon, Davies, Coltheart, et al. (2002) used cartoon stories, though, neither case involves a novel signaling task.

take perspective in a novel signaling task.

There is evidence showing that people are generally egocentric in their use of knowledge. Camerer, Loewenstein, and Weber (1989) show that people do not discount their own privileged information when dealing with naive others in a decision-making task. This effect – 'the curse of knowledge' – has been shown to impact perspective taking (Birch, 2005; Birch & Bloom, 2007). However, these tasks are not communicative and thus do not speak directly to whether such knowledge plays a role in novel signaling tasks. Galantucci (2009) finds evidence for egocentric behavior in a novel signaling task. However, this does not speak directly to the question of shared world knowledge because egocentricity in this case seems to be principally a matter of whether people know that others are attending to them, or of not realizing that the signal's visual appearance is ambiguous.

The present experimental task

The above discussion has implications for our experiment design. First, we need a task that, with minor alterations, can be used for communicative or non-communicative coordination between people. Second, the task must be amenable to both contextually constrained responses and open-ended ones. Third, we need a task that offers few opportunities for coordination except those based on shared world knowledge. Finally, since we aim to test for a possible dissociation between response behavior and communicative success, we need an empirical measure of salience that is task-independent. Mehta et al. (1994) assume that salience guides coordinative behavior and conclude that whatever people respond with during a coordination task is thus salient. This introduces a degree of circularity that, while not fatal to their project, would be problematic here.

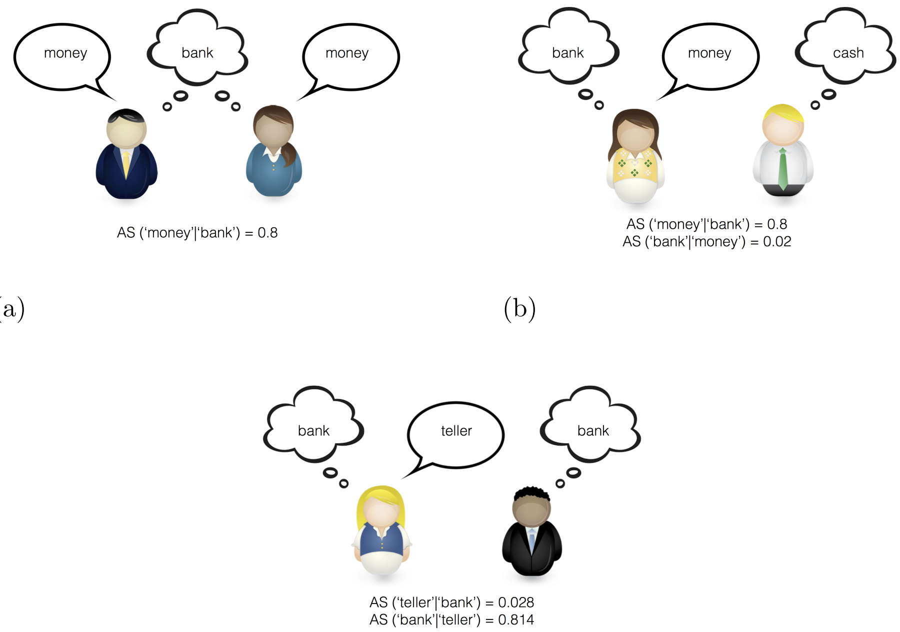
Our experimental task is illustrated in Fig. 1. In the non-communicative version (Fig. 1a, Study 1 below), both participants are given a one-word cue (in this example, 'bank') and they try coordinate by independently generating the same one-word response. For instance, if both participants generated the response 'money', they would succeed.

This is similar to the task used by Mehta et al. (1994), except for being more open-ended[4]. In the communicative version (Fig. 1b, c), the signaler is given the cue, and must generate a one-word signal to help their partner guess the cue. In a contextually unconstrained version (Study 2), both signaler and receiver can generate any English word. In other versions of the task, we constrain the context (Studies 3 and 4) by forcing the signaler or receiver to pick their response from a list of options, and we manipulate whether they are both given that list, or just one of them is.

There are several reasons for limiting cues and responses to single words. First, it affords an empirical, task-independent measure of salience. Associative strength (AS) values taken from published norms (e.g. Nelson et al., 2004) reflect the likelihood that someone given a particular cue will produce a particular response, all else being equal. If 80% of participants in such a study responded 'money' when given cue 'bank', then $AS(money|bank) = .8$ (for which read, the associative strength of response 'money' given 'bank'). On the common assumption that salience is a driver of coordination behavior (Clark et al., 1983; Levinson, 2006; Lewis, 1969; Mehta et al., 1994; Schelling, 1960), if AS predicts responses in our task, then AS is demonstrably a guide to salience. In that case, the fact that the vast majority of people responded with 'money' given 'bank' in Nelson et al. (2004) is evidence that money is a salient feature of banks. This provides a measure of salience that is independent of coordination success, so if participants produce 'money' as a response in both the communicative and non-communicative versions of the task, then we have evidence that salience drives their behavior, even if this response leads to success in Fig. 1a and failure in Fig. 1b.

Second, it affords a way to distinguish egocentric and allocentric salience, since AS is directional. In Nelson et al. (2004), some people were given 'bank' as a cue (and most responded 'money'). Others were given 'money' as a cue (and almost none responded

[4] In that task, participants coordinated by naming colors, makes of car, or types of flower, whereas here participants are not constrained by a particular semantic category.



(c)

Figure 1. (a) Non-communicative task. Both participants are given cue 'bank' and told to coordinate by generating the same word, i.e. respond with the same word that they think the other participant will. Here they both respond 'money' and are thus successful. (b) Communicative task where the signaler is given cue 'bank' and told to produce a one-word signal to help the receiver guess this target. Here, the signaler happens to signal 'money'. Given this signal, the receiver guesses 'cash'. Thus they are unsuccessful. (c) The same communicative task. The signaler is given cue 'bank' but this time happens to produce signal 'teller'. Given this signal, the receiver guesses 'bank' and they are thus successful. Associative Strength (AS) values (Nelson et al., 2004), represent how likely it is that the participants in their study produce a particular response when given a particular cue. $AS(y|x)$ represents the proportion of people in Nelson et al. who responded with y when given x.

'bank'). Thus, from the point of view of someone given 'bank', money is highly salient, but from the point of view of someone given 'money', banks are not salient. From the signaler's point of view in Fig. 1b, then, money is egocentrically salient, but not allocentrically informative. On the other hand, from the signaler's point of view in Fig. 1c, a teller is not egocentrically salient, but is allocentrically informative. AS values can thus be used to test whether people are taking perspective (i.e., are more likely to generate a signal such as 'teller' than one like 'money').

Third, a focus on single words allows us to isolate (as much as possible) effects of world knowledge. The one-word signal is the only information passing between signaler and receiver. Thus, the participants cannot rely on shared visual information (as in most perspective taking tasks), or on a shared history of interaction, which strongly affects signaling behavior (Brennan & Clark, 1996). Rather, they must rely on what they know about the referents of the various cues or signals.

Finally, it ensures a relatively clean measure of salience. In a game such as Pictionary, the graphical signal has several sub-parts. To signal 'Harrison Ford', for instance, someone might draw a man with a fedora, a whip and a gun. These may well be salient features of Harrison Ford construed as Indiana Jones, but it is difficult to isolate the contribution of each of these individual elements to guessing success, given the whole picture, and the same holds for spontaneous gesture. This issue is a potential confound when measuring communicative success (Sulik, 2018). Similarly, if the signal were an entire sentence (such as 'the place where money is deposited' for 'bank'), it would be prohibitively difficult to isolate the contribution of 'money', 'deposit', or even 'place' to the likelihood that someone would guess 'bank'.

Overall, this design allows us to test whether shared knowledge and perspective taking can explain responses or success in a novel signaling task. Signaler and receiver share a lot of world knowledge, so money is a salient feature of banks for both of them. However, only the signaler knows that banks are relevant to their current interaction and thus has 'bank' as a starting point. If the signaler acts egocentrically, they would probably produce 'money' as a signal (Fig. 1b). But 'money' does not make 'bank' salient to the receiver, in which case they would probably fail to guess 'bank'. In that case, they might fail to communicate *despite* shared world knowledge. Alternatively, the signaler might be able to take perspective, which involves suppressing the egocentric salience of 'money' and finding an allocentrically informative signal, such as 'teller' (Fig. 1c).

This task is similar to the television gameshow *Password* (examples can be found online at `www.youtube.com` with the search term 'password gameshow'), so it is certainly something humans are capable of. A similar task was used by Ryskin, Benjamin, Tullis, and Brown-Schmidt (2015) as a test of *prospective* perspective taking (perspective taking from the point of view of one's future self, with participants generating a cue such that would help them recall the target when given the cue a few days later). They found that AS predicted successful recall. We go further by exploiting the directional nature of AS to contrast egocentric and allocentric behavior, and by evaluating how likely people are to generate particular cues. Further, since the literature on novel signaling tasks shows that there is a difference in informativeness between generating a signal for one's self and doing so for another person (e.g., Garrod et al., 2007; Little, Eryılmaz, & de Boer, 2017), we test the effect of AS on response behavior when coordinating with another person.

It may be objected that the use of one-word cues and signals involves conventional language and is thus not a novel signaling task comparable with spontaneous gesture or graphical signaling. However, while 'money' is a conventional way to refer to money and 'teller' is a conventional way to refer to a teller, the key point is that they are not conventional ways to refer to banks, so the game, though using conventional *stimuli*, does not rely on conventional *signaling*. Since the task affords a neat, objective measure of salience, we consider this advantage to outweigh any potential negatives.

Summary and predictions

Two main factors are commonly argued to explain human success in novel signaling tasks: shared world knowledge (especially patterns of salience in that knowledge) and perspective taking. We have argued that, while salience has been shown to drive response behavior in a non-communicative task, it does not follow that it does so in a communicative task. Even if it does, it need not predict success. Similarly, while people are sometimes capable of perspective taking with shared visual scenes, it does not follow that they can do so in a novel signaling task that relies on shared world knowledge. We thus seek to explore the contributions of salience, perspective taking and context to success in a novel signaling task.

Study 1 explores coordination behavior based on world knowledge in a non-communicative task (Fig. 1a). We test whether associative strength predicts responses and success. If so, this measure serves as an empirical yard stick for salience in this Schelling-like task. Study 2 explores an otherwise-similar communicative task (Fig. 1b, c). Again, we test whether associative strength predicts responses and success. We also exploit the directionality of associative strength measures to test whether people are able to take perspective in a novel signaling task. We predict that salience will drive participant responses in both the communicative and non-communicative tasks, but that participants will behave egocentrically, and that success will be significantly lower in the communicative task.

Finally, in Studies 3 and 4 we explore contextual effects. Study 3 constrains the signal space by forcing the signaler to choose from a list of potential signals, while Study 4 constrains the meaning space by situating the target in a list of distractors. Additionally, it explores the role of common ground by manipulating whether the receiver shares this list. We predict that context will be a major driver of success. In that case, the explanatory burden must shift away from mutual salience and perspective taking and onto contextual factors. Finally, in study 5, we replicate the main results from studies 1-4 with a larger sample size.

Study 1: Coordination without communication

Overview

Participants took part in a non-communicative coordination task in which they were given a list of items. For each item, they were asked to coordinate by responding with the same word that other people would respond with. The main aims were (1) to provide a benchmark for coordination success. By drawing a comparison with the next study, this will allow us to assess whether communicative coordination is more of a challenge than non-communicative coordination. (2) To show that associative strength (AS) — a measure of accessibility derived from databases of word association norms (Fig. 1) — predicts coordination behavior. If it does, then associative strength can be used as a measure of salience, allowing us to test in the following study whether signalers are able to take perspective, seeing what is salient from the receivers' point of view.

Participants

We recruited 20 participants from Amazon's Mechanical Turk service. They received $1 in payment. Participation was limited to those registered as being in the USA, who had an approval rate of over 95%, and who had previously completed >1000 tasks. We ensured that no participant took part in more an one study, managing participation with Turkgate (Goldin & Darlow, 2013) and Turkprime (Litman, Robinson, & Abberbock, 2017). The study was approved by the University of Wisconsin-Madison Education and Social/Behavioral Science IRB.

Materials

We compiled a list of 20 one-word items to act as cues (see supplementary materials), based on word-association norms from Nelson et al. (2004). Our cue items were evenly divided into symmetric and asymmetric words, defined as follows. If the top associate of word X is word Y and the top associate of word Y is word X, then item X is symmetric. For instance, the top associate of 'day' is 'night' ($AS = .819$) and the top associate of 'night' is 'day' ($AS = .686$). On the other hand, an item is asymmetric if word X strongly cues Y but Y does not strongly cue X. For example, 'bank' strongly cues 'money' ($AS = .799$) but 'money' does not strongly cue 'bank' ($AS = .019$). This ensures that some items would provide a difference in perspective (cf. the discussion of asymmetry above, and

Fig. 1). Additionally, we filtered the shortlist of asymmetric items so that for each item, there exists a word Z such that Z is weakly associated with X, but X is strongly associated with Z. For instance, 'bank' weakly cues 'teller' ($AS = .028$) while 'teller' strongly cues 'bank' ($AS = .814$). This ensures that there exists a signal for the following study that would be informative for the receiver (cf. Fig. 1c). To allow participants some morphological and typographical leeway, 'goodbye', 'good-bye', 'good bye' and 'goodbyes' all counted as the same response. We collapsed the AS norm data across these distinctions.

Procedure

The participants were told they would play a word-guessing game in which they would be given a cue, such as 'puppy', and would have to think of one word in response to this cue, such as 'dog'. They were told that the aim of the game was to answer with the same word as another participant that they would be randomly paired with. This is similar to the verbal coordination task in Mehta et al. (1994), but more open-ended. Participants were then given the 20 cue items in a randomized order and were reminded each time to think of a response that would match someone else's. They were able to produce any English word as a response. There was no time limit on providing a response.

Analysis

There are two measures of interest. The first (coordination index) is a measure of how successful people were in coordinating over a given item. Mehta et al. (1994) define the coordination index as the probability that, over all possible pairings within the set of participants, the pairs responded with the same word. Let N be the number of participants, k the number of distinct responses to a given item and $m_1, ..., m_k$ the number of participants that gave each response. Thus, the coordination index is calculated as in formula 1.

$$c = \sum_{i=1}^{k} (m_i/N)[(m_i - 1)/(N - 1)] \tag{1}$$

This index serves as a baseline for comparison with the communicative task in the next study. The second measure (response count) is the number of participants producing each response ($m_1, ..., m_k$ above), indicating how likely it is that participants generate a particular response.

Our main predictor is AS. Since responses to such a task are assumed to be based on salience (cf. Lewis, 1969; Mehta et al., 1994), we will test whether AS significantly predicts response counts. If so, AS is a measure of salience. Additionally, we test whether the maximum AS of an item predicts the coordination index. Consider, for example, the top 5 associates of items 'cut' and 'bulb' (Fig. 2a). The top-ranking associate of 'bulb' has an AS of .788, and the next highest is just .027. In that case, 'light' should be strongly salient given 'bulb'. On the other hand, the top-ranking associate of 'cut' has an AS of just .168, so although 'blood' is relatively salient given 'cut', there is no word that is as salient for 'cut' as 'light' is for 'bulb'. It would therefore not be surprising if more people are able to coordinate for 'bulb' than for 'cut', and we therefore predicted that the coordination index for an item will be predicted by its maximum AS.

For this and all subsequent studies, reported AS values come from the University of South Florida (USF) Free Association Norms (Nelson et al., 2004). In addition, to ensure that these results reflect intersubjective rather than subjective salience, we checked for agreement with associative strengths drawn from other databases. Data collection for the Edinburgh Associative Thesaurus (EAT, Kiss, Armstrong, Milroy, & Piper, 1973) was similar to the USF norms, except that the participants spoke British rather than American English. The Small World of Words database (SWOW, De Deyne, Navarro, & Storms, 2012) differs from the USF and EAT norms in allowing multiple responses rather than just one, and in being a voluntary, mass online study rather than a supervised in-person study. We assessed the robustness of our results by examining whether they hold for all these measures of AS.

Results

Coordination success. Fig. 2b shows the coordination index for each item ($M = .46, SD = .24$). This is similar to the mean value for the verbal-coordination tasks in Mehta et al. (i.e., their questions 1 to 10; $M = .44, SD = .21$). In line with the 'light'|'bulb' and 'blood'|'cut' example in Fig. 2a, the variation in coordination indexes is significantly predicted by maximum AS of each item (linear regression $\beta = 1.08, SE = 0.18, t(18) = 5.86, p < .001$; Fig. 2c). The more salient the top-ranking associate, the easier it was for people to coordinate. The model accounts for much of the variance in the coordination index (adjusted $R^2 = .64$).

Response behavior. To model how likely it was that participants would generate each response, we used a binomial mixed-effects regression with the proportion of participants generating each response as the dependent variable, and AS as the predictor (for random effects structure, see supplementary material). The response proportions were positively predicted by AS ($\beta = 5.24$, $SE = 0.24$, $z = 21.83$, $p < .001$; Fig. 2d). For instance, in response to item 'bank', 15 of 20 participants responded 'money' ($AS = .799$) and just two of 20 responded 'account' ($AS = .035$). These conclusions held across word-association databases (see supplementary material).

Since 'money' is a more common word than 'account', it is possible that the effect of AS might reduce to an effect of word frequency. To rule this out, we model the effect of word frequency on coordination behavior using frequency data from $\text{SUBTLEX}_{\text{US}}$ (Brysbaert & New, 2009), though the frequency data is log-transformed into a Likert-like scale as recommended by Van Heuven, Mandera, Keuleers, and Brysbaert (2014). When it is the only explanatory variable in the model, word frequency predicts coordination behavior ($\beta = 0.36$, $SE = 0.08$, $z = 4.47$, $p < .001$): the more common a word, the more participants generated it. However, when both word frequency and AS are included in the

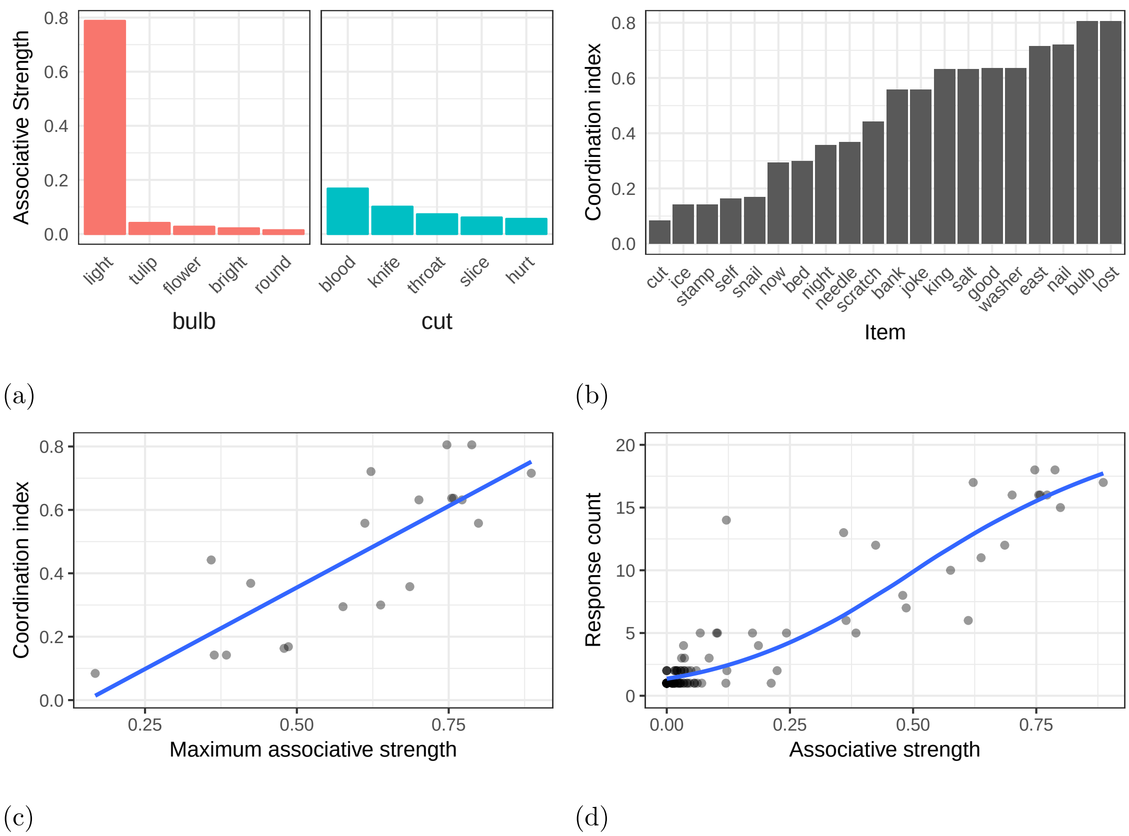


Figure 2. (a) An example of the associative strength values of the top-5 ranking associates of two cue items, 'bulb' and 'cut', from Nelson et al. (2004). (b) The coordination index by item. (c) Linear regression fit of maximum associative strength per item as a predictor of coordination index. Maximum associative strength represents the top-ranked associate of each item (e.g., 'light' for 'bulb' and 'blood' for 'cut'). (d) The number of people producing each response (dots) and binomial model fit (curve).

model, the effect of word frequency is no longer significant ($\beta = -0.1$, $SE = 0.08$, $z = -1.32$, $p = .188$), while AS remains a significant predictor ($\beta = 5.27$, $SE = 0.25$, $z = 21.37$, $p < .001$). The effect of AS is thus not a proxy for word frequency.

Discussion

The principal aims of Study 1 were (1) to quantify the level of coordination success to serve as a baseline for comparison with a communicative task in Study 2; and (2) to test whether associative strength (AS) between the cue and target words predicts response behavior. We found (1) that participants were able to coordinate 46% of the time, a result similar to verbal coordination tasks in Mehta et al. (1994); and (2) that AS strongly and significantly predicted coordination behavior. When there was a strong associate of an item, participants were likely to produce that as their response, and the likelihood of their doing so varied along with the strength of the association. This is not explicable as an effect of word frequency. Further, the more salient an item's top associate, the more likely people were to coordinate successfully.

Since various accounts (e.g., Lewis, 1969) argue that salience guides coordination behavior, a reasonable interpretation of result (2) is that AS is a guide to salience: the more salient Y is given X, the more likely it is that the word for Y occurs to people given the word for X in an association study, so the higher its AS. This makes AS a good empirical measure of salience (in as far as it predicts behavior here). Since the AS values were derived from large-scale studies that had nothing to do with coordination, using these as our predictor variables in what follows avoids the circularity of arguing that people coordinate by picking a salient response, and then claiming that their response is salient because they produced it when coordinating. Sperber and Wilson (1995) frame their account in terms of accessibility rather than in terms of salience, but accessibility is even more transparently related to AS than salience is: the more accessible a word Y is given word X, the sooner Y would occur to someone when given X.

To be clear, our finding that salience plays a role in non-communicative coordination does not mean that participants must represent whatever is salient to them as also being salient to others. The most parsimonious explanation is that participants simply respond with whatever is most salient from their own point of view, since that is what AS measures. Coordination is thus achievable without taking into account what others may be thinking. The main goal of Study 2 was to examine whether people would still take such an egocentric approach, even when taking the perspective of the receiver would improve the chance of success.

Study 2: Coordination for communication

Overview

Participants took part in a novel signaling task where signalers were given a list of target items, and had to come up with a one-word signal to help the receiver guess the target. Receivers were given these signals and had to make a guess what the target was. The main aims here are to test (1) whether coordination in an open-ended communicative context is significantly harder than in a similar non-communicative task and (2) to test whether people's responses are better predicted by egocentric (own-POV) or allocentric (other-POV) salience.

Participants

We recruited 10 participants from Amazon's Mechanical Turk service to serve as signalers. These signalers produced a total of 128 unique signals. To determine the communicative effectiveness of the signals, 80 receivers were each given a random sample of 16 signals (16 is a factor of 128), resulting in 10 guesses for every signal. Participants were paid $1.

The number of unique signals increases rapidly with the number of signalers. Thus, a small increase in the number of signalers means a large increase in the number of guessers. For practical reasons, we have thus kept N low in this study. An alternative strategy would be to collect guesses for a subset of signals (e.g., using all signals produced by more than one signaler, and then additionally sampling from the signals produced by just one signaler). We pursue this alternative with a larger N in the replication in Study 5.

Materials

This study used the same list of cue items from Study 1.

Procedure

The signalers were told that they would be playing a word-guessing game in which they would have to think of one-word signals that would help someone guess their items. They were talked through an example: if the item was ‘dog’, then a good signal would be ‘puppy’ since most people given ‘puppy’ would probably guess ‘dog’. They were given the 20 items in a randomized order. Under each item was a text entry box to input their signal. After all signal words were collected, a similar survey was presented to the receivers. They were told they would be playing a word-guessing game, and that someone else had chosen a one-word signal to help them guess the item. The instructions walked them through the ‘puppy’ example from the receiver’s point of view. In neither case was there a time limit on responding.

Analysis

In addition to allowing leeway in spelling and morphology as mentioned previously, we counted a guess as correct if it was a compound containing the item, but only when the item was the head of the compound (e.g., ‘lightbulb’ is a correct guess for ‘bulb’ since a lightbulb is a kind of bulb, whereas ‘doghouse’ is not a correct guess for ‘dog’ since it is a kind of house, not a kind of dog). This issue did not arise in Study 1.

Let k be the number of distinct signals produced for an item across all signalers. Let $s_1, ..., s_k$ be the number of signalers producing each of the signals $1, ..., k$ and let $g_1, ..., g_k$ be the number of guessers correctly guessing the item, given each of the signals $1, ..., k$. Thus, since N is the number of signalers or receivers, the correctness score for each item is:

$$c = \sum_{i=1}^{k} (s_i/N)(g_i/N) \tag{2}$$

Despite differences in formulas (1) and (2), it will be worth testing whether the ‘coordination index’ for the previous study and the ‘correctness score’ for the present one are related. They are similar in that they both represent the success criteria for each response. For Study 1, success is calculated as the number of other participants providing the same word; in the present study, success is the number of receivers guessing the target. In both cases, success is calculated per response, and then success values are aggregated per item.

One conceptual difference is that the interactions in the present study are determined by assigned communicative role, whereas the interactions in the previous study are calculated over all possible pairings. Despite this difference, since these tasks involve the same items, participants have access to the same world knowledge. Thus, by comparing success scores, we can investigate whether the same world knowledge can be leveraged to coordinate communicatively (where perspectives may differ) and non-communicatively (where perspectives align).

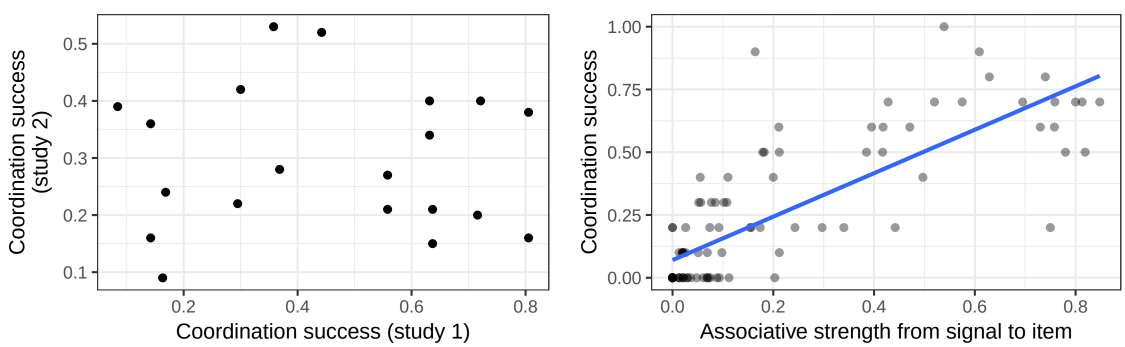
Since AS is directional, to streamline the presentation of results, we will call the AS from whatever the participant is given to whatever they produce 'forward' and the reverse 'backward'. Thus, in Fig. 1c, the signaler is *given* item 'bank' and *produces* signal 'teller', so from their point of view, forward $AS = .03$ and backward $AS = .8$. The receiver, on the other hand, is given signal 'teller' and produces guess 'bank' so from their point of view, forward $AS = .8$ and backward $AS = .03$. Thus, regardless of communicative role (signaler vs. receiver), egocentric salience (salience from one's own point of view) is represented by forward AS and allocentric salience (salience from the other's point of view) is represented by backward AS.

Results

Success. Coordination success ($M = .3, SD = .24$) was worse than in the previous task (difference in means = .16, bootstrapped 95% CIs [.045, .275]). Coordination in this communicative task was thus significantly more difficult than coordination in an otherwise similar non-communicative task.

In Study 1, the variation in coordination success across items was predicted by the maximum AS from an item to its top-ranking associate (cf. the ‘light’|‘bulb’ vs. ‘cut’|‘blood’ example above). In the present study, a linear regression finds no effect ($\beta = -0.18$, $SE = 0.13$, $t = -1.38$, $p = .186$), though success was higher for symmetric than for asymmetric items ($\beta = 0.13$, $SE = 0.06$, $t = 2.14$, $p = .047$). By item, there was no correlation between success at the previous task and success at the present one ($r = -.013$, $p = .96$, Fig. 3a).

Shifting focus from by-item success to by-signal success, the AS from signal to item was a significant positive predictor of how many receivers guessed each signal correctly ($\beta = 0.94$, $SE = 0.07$, $t = 12.85$, $p < .001$, Fig. 3b), whereas the AS from item to signal is now significant, but *negative* ($\beta = -0.16$, $SE = 0.08$, $t = -2.1$, $p = .039$). As illustrated by Fig. 1, the positive effect of signal-to-item AS means that success here is driven by receiver-POV rather than by signaler-POV salience. In fact, the negative effect of the item-to-signal AS suggests that signaler-POV salience can *hinder* communication.



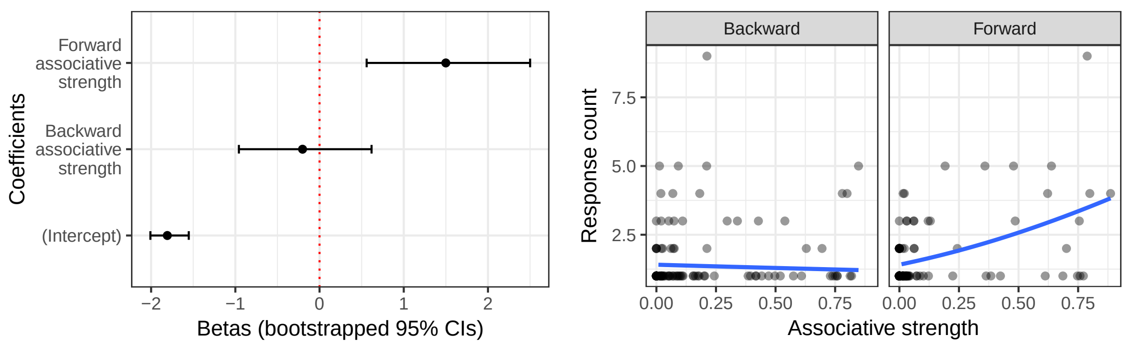
(a) (b)

Figure 3. (a) Each point represents an item. Coordination success for an item in Study 1 is uncorrelated with coordination success in Study 2. (b) Each point represents a signal, while the line represents a linear regression fit of the relationship between signal-to-item associative strength and coordination success.

Signaler behavior. As previously, we analyzed the relationship between AS and the proportion of signalers producing each signal with a binomial mixed-effects regression (Fig. 4). Forward AS significantly predicted how many signalers generated each signal

($\beta = 1.5$, $SE = 0.48$, $z = 3.1$, $p = .002$), but backward AS did not ($\beta = -0.2$, $SE = 0.43$, $z = -0.47$, $p = .639$). Thus, signalers behaved egocentrically: they were more likely to produce signals that were salient from their own point of view. See supplementary material for random effects structure, a demonstration that this conclusion holds across associative norms databases, and interaction terms.

When it is the only predictor in the model, word frequency falls just short of significance ($\beta = 0.13$, $SE = 0.08$, $z = 1.71$, $p = .088$). When forward AS is included in the model, the effect of word frequency remains nonsignificant ($\beta = 0.009$, $SE = 0.08$, $z = 0.1$, $p = .919$) and forward AS remains significant ($\beta = 1.6$, $SE = 0.3$, $z = 5.25$, $p < .001$). The contribution of AS to signaler behavior thus does not reduce to an effect of word frequency.

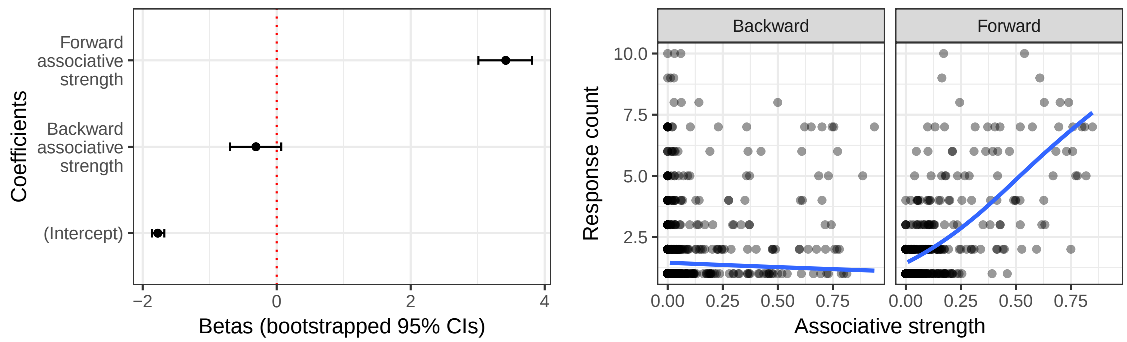


(a) (b)

Figure 4. (a) Binomial mixed-effects regression coefficients for the effect of associative strength on the proportion of signalers producing each response. (b) Model fit (curve) and data (points).

Receiver behavior. Like signalers, receivers behaved egocentrically since forward AS significantly predicted the proportion of receivers generating each guess ($\beta = 3.42$, $SE = 0.21$, $z = 16.38$, $p < .001$, Fig. 5) whereas the effect of backward AS was not significant ($\beta = -0.31$, $SE = 0.2$, $z = -1.57$, $p = .116$). See supplementary material for random effects structure, a demonstration that this conclusion holds across associative norms databases, and interaction terms.

Word frequency is a significant predictor of responses when it is the only predictor in the model ($\beta = 0.21$, $SE = 0.03$, $z = 6.4$, $p < .001$). When forward AS is included in the model, word frequency falls short of significance ($\beta = 0.07$, $SE = 0.04$, $z = 1.69$, $p = .092$), though forward AS remains significant ($\beta = 3.26$, $SE = 0.19$, $z = 17.27$, $p < .001$). As previously, AS is not a proxy for word frequency.



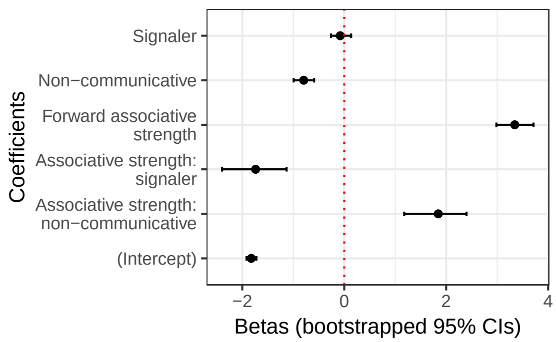
(a) (b)

Figure 5. (a) Binomial mixed-effects regression coefficients for the effect of associative strength on the proportion of receivers producing each response. (b) Model fit (curve) and data (points).

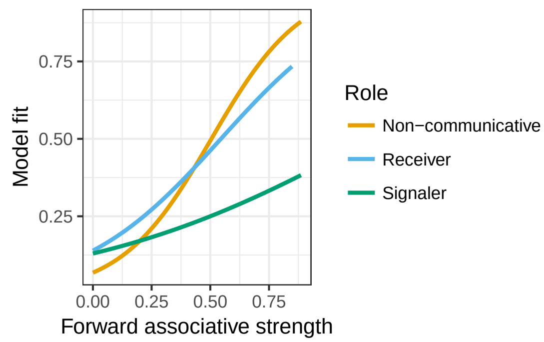
Comparing behavior across tasks. Potential differences in behavior across tasks include (a) the degree to which participants were egocentric, and (b) the degree to which salience predicted behavior.

To explore (a), we merged all data sets discussed so far, after introducing a variable to represent task (with values ‘non-communicative’, ‘signaler’ and ‘receiver’). We analyzed the relationship between forward AS and responses with a binomial mixed-effects regression (Fig. 6a) that included an interaction between task and forward AS. Participants in the non-communicative task were significantly more egocentric than receivers, who were in turn more egocentric than signalers (Fig. 6b). However, in Study 5 we find that the difference between communicative and non-communicative behavior replicates, but the difference between signalers and receivers does not.

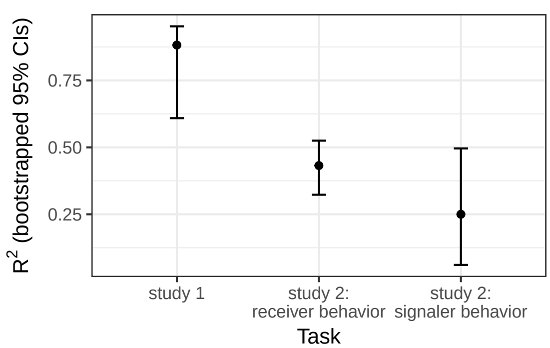
To explore (b), we noted that the spread of data points about the regression curves in Fig. 4b and Fig. 5b was wider than in Fig. 2d. We calculated bootstrapped confidence intervals for the pseudo-R^2 for each of these models (R_1^2, Kvålseth, 1985). Salience was a stronger predictor of behavior in the non-communicative task than it was in either of the communicative ones (Fig. 6c). Thus, there is a disjunction between participants' degree of egocentricity (the β representing the effect of forward AS in Fig. 2d, 4b, 5b) and the extent to which participants rely on salience at all (the pseudo-R^2 values of those models). While receivers and non-communicative participants were both more egocentric than signalers, participants' behavior in either communicative task is less predictable by AS values than in the non-communicative task.



(a)



(b)



(c)

Figure 6. (a) Binomial mixed-effect regression coefficients for the effect of forward associative strength on the proportion of receivers producing each response, including an interaction with task. The base level represents the receiver in the communicative task. (b) Model predictions for the effect of forward associative strength. (c) Adjusted pseudo-R^2 values, using R_1^2 from Kvålseth (1985).

Discussion

Participants typically behaved egocentrically rather than allocentrically. Since success here was driven by receiver-POV salience, and since signalers typically failed to respond allocentrically, coordination success was significantly lower than in the non-communicative task. Performance was worse when items involved a difference in perspective (asymmetric items). Thus, although diverse approaches (e.g., Levinson, 2006; Lewis, 1969; Sperber & Wilson, 1995) argue that shared world knowledge and perspective taking allow people to communicate in the absence of a conventional signaling system, we have found that neither signalers nor receivers consistently take an allocentric perspective when they have to rely on shared world knowledge to coordinate, and that this can hinder communicative success.

Mehta et al. (1994) show that Schelling-like focal points exist and that people are able to use them to coordinate in a non-communicative task. However, the results here show that such conclusions do not extend to coordination in open-ended communicative tasks, meaning that salience and shared world knowledge are thus not *general* solutions to the problem of coordination in the absence of convention, contrary to claims reviewed above. All the participants in both studies discussed so far presumably knew that money is a salient feature of banks. However, this only enabled them to coordinate when they have the same starting point (i.e., were both given cue 'bank', as in Study 1), but it was a hindrance to communication here (e.g., 'money' was the most popular signal for item 'bank', though it was uninformative from the receiver's POV).

Previous work has shown that people sometimes fail to take perspective (e.g., work by Keysar and colleagues), but these have tended to focus on visual salience — seeing what others can and cannot see in one's immediate environment. Our study focuses instead on salience in world knowledge. While AS was a strong predictor of behavior in the non-communicative task (meaning that it is a good empirical yardstick for salience in world knowledge), it was significantly less predictive in the communicative task. Thus, rather than claiming that world knowledge does or does not play an explanatory role in coordination behavior, researchers should focus on how it contributes differently according to the nature of the task.

A potential limitation of our results would be if signalers were egocentric because they simply did not have access to information that would allow for an allocentric response. For example, perhaps signalers simply did not know that ‘teller’ would work better than ‘money’ as a signal for ‘bank’. In Study 3 we therefore test whether participants know this and can use this information in a more constrained situation. If so, this would confirm that the failure to use it in Study 2 was not due to a lack of knowledge, but rather due to the inaccessibility of that knowledge and the difficulty inherent in taking someone else’s perspective.

Study 3: A constrained signal space

Overview

The previous study explored perspective taking in an open-ended task: the only restrictions were the instruction to try coordinate, and the requirement that the response be an English word. The present study tests whether people can take perspective in a constrained context. For example, if the target is ‘bank’, instead of open-endedly generating any word they wish, now signalers must choose one signal from the list: ‘money’, ‘teller’, ‘vault’, ‘loan’, ‘safe’. In particular, since the previous study found success to be driven by receiver-POV salience, we test whether signalers are able to select the most allocentrically informative signal given a constrained signal space.

Participants

We recruited 20 signalers from Amazon’s Mechanical Turk service, 10 to choose a signal that they thought would help someone else to guess the item, and 10 to choose a signal that they thought would help themselves guess the item (this instructional manipulation has no effect on any of the results below, for which see supplementary material, so we do not discuss it here). Each participant received $1. From the receiver's point of view, this task is precisely the same as in Study 2, so where available, we simply reused data from Study 2 to provide guesses used in calculating correctness scores (formula 2). For signals not generated previously, we recruited further receivers (again, 10 per signal), but this meant that receivers saw a variable number of signals. They were paid an amount proportional to $1 for 20 signals.

Materials

For every item, we generated a list of five potential signals as follows. The list contained the top-ranking associate of the item in the USF database (e.g., for item 'bank', the list contained 'money'). Where different, it also contained the word in the USF database for which the item was the top-ranking associate, since this was an informative signal (e.g., 'bank' was the top-ranking associate of 'teller'). If the most popular signal in Study 2 was not one of these, we additionally included it. The rest of the list was sampled from signals generated in Study 2.

Procedure

The signaling task was explained to participants as previously. They were then given all 20 items in random order and along with each item, a list of five potential signals (also in random order). Participants were asked to choose from the five signals. Half the signalers were asked to pick which item they thought would help someone else to guess the item, and the other half were asked to pick which signal they thought would help themselves guess the item. There was no time limit on each response.

Results

Overall coordination success ($M = .5, SD = .19$) was similar to performance in Study 1, but a significant improvement over Study 2 (difference in means= .20, bootstrapped 95%

CI [.118, .282]). Constraining the signaling space thus significantly boosts coordination success.

We entered forward and backward AS as fixed effects into a binomial mixed-effects regression. The outcome variable was the proportion of signalers choosing each signal. While forward AS predicted signaling choices in both Study 1 and Study 2, it falls just short of significance here ($\beta = 0.63$, $SE = 0.33$, $z = 1.9$, $p = .057$, Fig. 7a, b), and backward AS is now a significant predictor ($\beta = 2.27$, $SE = 0.42$, $z = 5.41$, $p < .001$). Constraining the signaling space thus promotes perspective taking. The model pseudo-$R^2 = .32$, so salience explained behavior less than in Study 1, and about the same as in Study 2. See supplementary materials for random effects structure, and a demonstration that this effect does not reduce to one of word frequency.

Across word association norms databases, there is consistently a significant effect of backward AS, but for the other databases (EAT and SWOW), there is additionally a significant (though smaller) main effect of forward AS. The conclusion, then, is that when the signal space is constrained, participants are more allocentric than egocentric, though they are nonetheless somewhat egocentric.

The inclusion of a two-way interaction between forward and backward AS significantly improves model fit ($\chi^2(1) = 14.65$, $p < .001$). In addition to the effect of backward AS ($\beta = 3.29$, $SE = 0.475$, $z = 6.93$, $p < .001$) there is a significant (though smaller) effect of forward AS ($\beta = 1.0$, $SE = 0.46$, $z = 4.1$, $p < .001$). The interaction term ($\beta = -4.04$, $SE = 1.08$, $z = -3.75$, $p < .001$) means that backward AS has less of an effect when forward AS is high (Fig. 7c, right panel), and participants still behave somewhat egocentrically at times, though only for low values of backward AS (Fig. 7c, left panel).

Discussion

A constrained signal space boosted coordinative success as high as it was in the non-communicative task (Study 1). Unlike in previous tasks, participants behaved

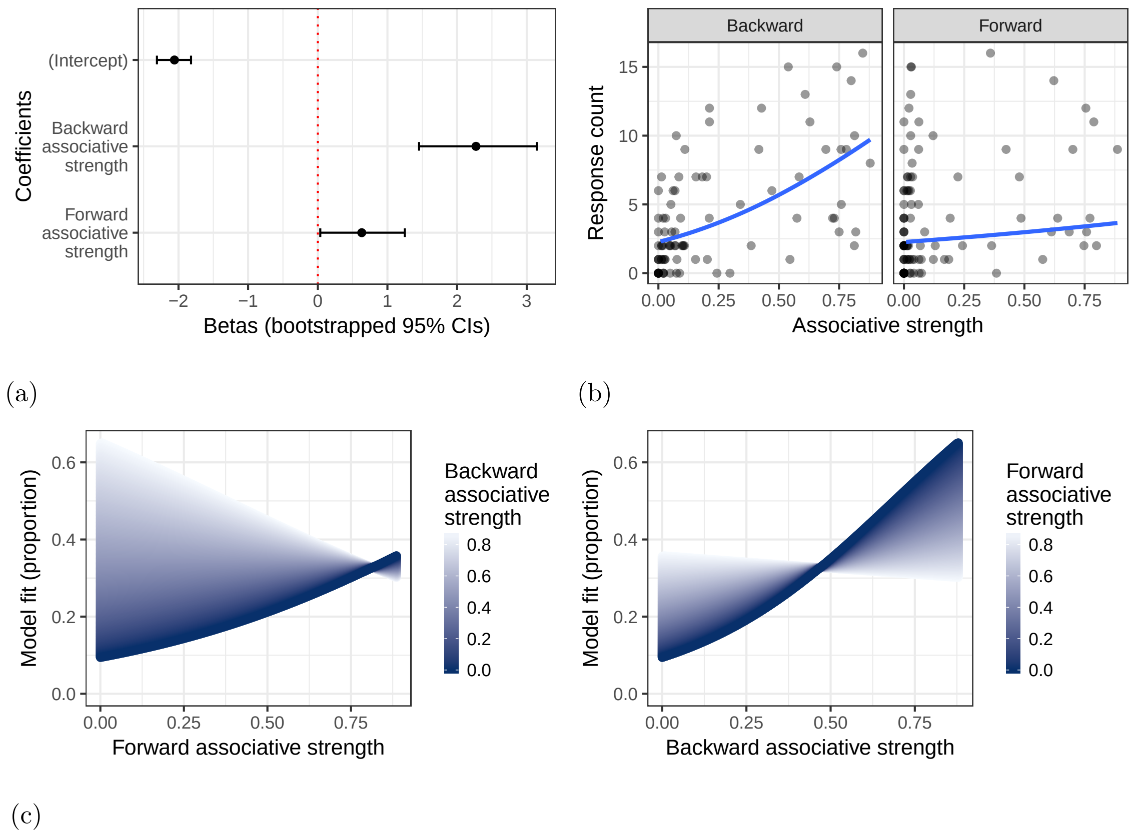


Figure 7. (a) Binomial mixed-effects regression coefficients (without an interaction term) for the effects of forward and backward associative strength on the proportion of signalers producing each response. (b) Model predicted response counts (curve) and data (dots). (c) Model fit for the proportion of signalers generating each response, including an interaction term between backward and forward associative strength. For each panel, the darkest line represents the main effect of the variable on the x-axis (i.e., when the value of the other variable, labelled on the legend, is 0). As the value of the other variable increases, the color becomes lighter. Thus, a comparison of the left and right panels shows that the main effect of backward associative strength is larger than that of forward associative strength. The right panel shows that the effect of backward associative strength is positive for all but the highest values of forward associative strength, while the left panel shows that forward associative strength only has a positive effect for lower values of backward associative strength.

allocentrically. Salience was as poor a guide to behavior as it was in the open-ended communicative task (Study 1).

There are two (potentially compatible) ways of framing this result. One is that people are egocentric in the general, open-ended case but can behave allocentrically in constrained communicative contexts, such as when the signal space is limited in this way. The second is that people are egocentric when generating hypotheses about communicative choices (since in Study 2, they had to generate their own signals) but that they can be allocentric when evaluating hypotheses (since in Study 3, they just had to evaluate which is the best of the given signals). Untangling these two possibilities may have implications for how we should explain the evolution of language in our species, since the first focuses more on the communicative context (open vs. constrained signal space), and the second focuses more on cognitive abilities (hypothesis generation vs. evaluation).

Either way, the present result is useful since it demonstrates that the signalers in the open-ended task possess the relevant information: it's not that they simply didn't *know* 'teller' would make a better signal than 'money' for 'bank'; it's that 'teller' simply didn't *occur* to them as a signal, given that their behavior was driven by egocentric salience, so they don't even have the chance to evaluate the informativeness of 'teller'. People do indeed share the relevant world knowledge, but the trick lies in bringing that world knowledge to bear on a particular problem. Signalers managed to do so here, but not in the previous study. Thus, it is overly simplistic to claim that salience drives coordination behavior. A more realistic claim is that salience, task (e.g., communicative vs. non-communicative) and context (e.g., constrained vs open-ended signal space) interact to do so.

Study 4: A constrained meaning space and common ground

Overview

Here we test whether constraining the meaning space has an effect on communicative behavior, since constraining the signal space did so in study 3. There are doubtless several ways of doing this, but here we explore the effects of placing a target item in the context of distractor items which share patterns of salience. For example, the most salient signal given target 'bank' is 'money', so we place 'bank' in the context of four distractor items for which ‘money’ is highly salient.

Additionally, we manipulate whether or not the signaler knows that the receiver has access to the same constrained meaning space. This allows us to test the effect of common ground on perspective taking. In general, a signaler and receiver can assume that much of their world knowledge broadly overlaps. However, sharing world knowledge in this unconstrained sense did not help them take perspective in Study 2. Here we test whether people take perspective when attention is focused on a constrained meaning space, either for the *signaler only* (the ‘no-common ground task’), or for *both* the signaler and receiver (the ‘common ground task’).

Participants

Payment and requirements for participants are the same as described previously. As previously, 10 signalers saw each item, and 10 receivers saw each signal. Since the number of unique signals varied across items, the number of signals seen by each receiver varied.

From the receiver’s point of view, the no-common-ground task is precisely the same as in Study 2 and 3, so where available, we simply reused data from those studies to provide correctness scores. For signals not generated in previous studies, we recruited further receivers as described for Study 3.

Materials

For each target item, we identified the egocentrically most salient associate, and then constructed a set of distractors by selecting four other words that strongly cue the same associate. For example, the associate with the highest forward AS from item ‘bank’ is ‘money’, so the set of distractors was ‘cash’, ‘fund’, ‘wallet’ and ‘profit’. The target and the distractors together constituted the constrained or focal meaning space.

Procedure

Participants were given similar instructions as previously, including the same 'puppy' example from previous studies. Signalers were told they would be shown a list of five potential target items (e.g., 'bank', 'cash', 'fund', 'wallet', 'profit' in a randomized order) and told to read them over. After clicking a button to indicate they had read through the list, one of the five items was highlighted, and they were told that this was the item they should get someone to guess. Though it appeared to them that it was a random selection, each time the target item was simply the item that has been used in all the studies above (in this case, 'bank').

Additionally, they were told during the initial instructions that the receiver either would or would not have the same list of 5 items in front of them while guessing. Thus, in the common-ground task, their goal was just to get the receiver to pick the target from the list. In either task, they were reminded about whether the guesser would have the list before every trial, and could generate any English word as in Study 2.

Receivers in the common-ground task were given the same list of five potential items (in a randomized order) and told to guess which one the signaler intended to signal. Receivers in the no-common-ground task are not given a list to choose from, and could guess any English word, as in Study 2. In neither task was there a time limit.

Results

Accuracy. Coordination success for the common-ground task ($M = .72, SD = .15$) was higher than the no-common-ground task ($M = .40, SD = .17$, bootstrapped 95% CI for the difference in means [.227, .424]). Constraining the meaning space for the receiver thus improved performance. Fig. 8 illustrates that this was the best performance across all studies: making the receiver pick from a list of items (4b) improved accuracy even more than making the signaler pick from a list of signals in Study 3 (bootstrapped 95% CI for the difference in means [.138, .31]).

In the no-common ground task (4a), performance was better than the open-ended task 2 (bootstrapped 95% CI for the difference in means [.003, .191]) and poorer than task 3 (bootstrapped 95% CI for the difference in means [-.197, -.003]). Thus, constraining the meaning space for the signaler improved performance relative to the open-ended task, but not as much as constraining the signaling space.

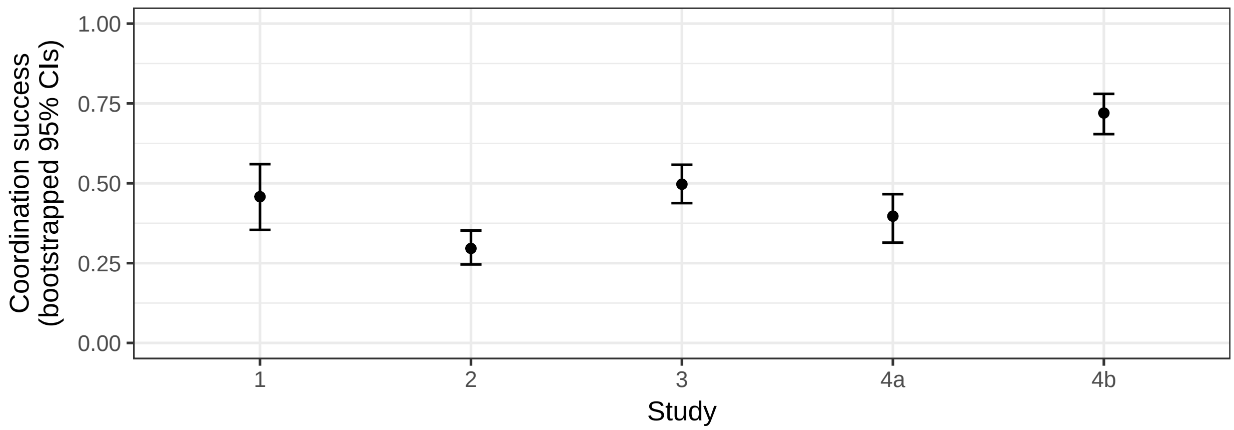


Figure 8. Mean values for coordination success (formula 1 for Study 1, formula 2 for others) across studies. 4a represents the no-common-ground task; 4b the common-ground task.

Signaler behavior. For both the common-ground and no-common-ground tasks, we modeled the effect of AS on the proportion of signalers generating each signal with a binomial mixed-effects regression. See the supplementary material for random effects structures and a demonstration that the effects below do not reduce to an effect of word frequency.

For the common-ground task, signalers' behavior was predicted by backward AS ($\beta = 0.56$, $SE = 0.39$, $z = 2.69$, $p = .007$, Fig. 9a, b) but not by forward AS ($\beta = 0.73$, $SE = 0.4$, $z = 1.4$, $p = .162$, model pseudo-$R^2 = .3$). There was some inconsistency across association norms databases (see supplementary material), though the Akaike Information Criterion for the USF database reported here was the lowest. Thus, instead of concluding that signalers were straightforwardly allocentric in this task, we draw the weaker conclusion that they were at least somewhat allocentric. There was no significant interaction.

For the no-common-ground task, both forward and backward AS were significant predictors, though the former has a larger effect (forward AS $\beta = 1.52$, $SE = 0.5$, $z = 3.06$, $p = .002$; backward AS $\beta = 0.85$, $SE = 0.38$, $z = 2.24$, $p = .025$; model pseudo-$R^2 = .25$, Fig. 9c, d). Across AS norms databases, the effect of forward AS was consistently significant and consistently larger than that of backward AS.,The effect of backward AS was not significant for the EAT database. Thus, signalers were more egocentric than allocentric, though they were still allocentric relative to the open-ended task (Study 2). There was no significant interaction term.

We combined data for the two versions of this task in order to explicitly model the effect of the instructional manipulation (telling signalers that the receiver had access to the same list of targets or not). We included a pair of two-way interactions: between task and forward AS, and between task and backward AS. There were significant main effects for both forward AS ($\beta = 1.04$, $SE = 0.35$, $z = 2.99$, $p = .003$) and backward AS ($\beta = 0.82$, $SE = 0.34$, $z = 2.41$, $p = .016$) but not for task ($\beta = -0.28$, $SE = 0.17$, $z = -1.62$, $p = .105$). There was no significant interaction between task and forward AS ($\beta = -0.91$, $SE = 0.52$, $z = -1.74$, $p = .082$) or backward AS ($\beta = 0.21$, $SE = 0.50$, $z = 0.42$, $p = .67$). On the whole, then, when the meaning space was constrained in this way, signalers exhibited both egocentric and allocentric behavior, though the former effect was larger.

Receiver behavior. From the receiver's point of view, the no-common-ground task is precisely the same as in Study 2. Thus we focus on the common-ground task here (though see supplementary material for the no-common-ground task, which replicates the results from Study 2). Again, we use a binomial mixed-effects regression to model the effect of forward and backward AS on the proportion of receivers that selected each guess. There is a significant positive effect of forward AS ($\beta = 9.16$, $SE = 2.29$, $z = 4.0$, $p < .001$) and a smaller, negative effect of backward AS ($\beta = -0.97$, $SE = 0.39$, $z = -2.48$, $p = .013$, model pseudo-$R^2 = .18$, Fig. 10a, b). See supplementary material for random effects structure, word frequency and results across norms databases.

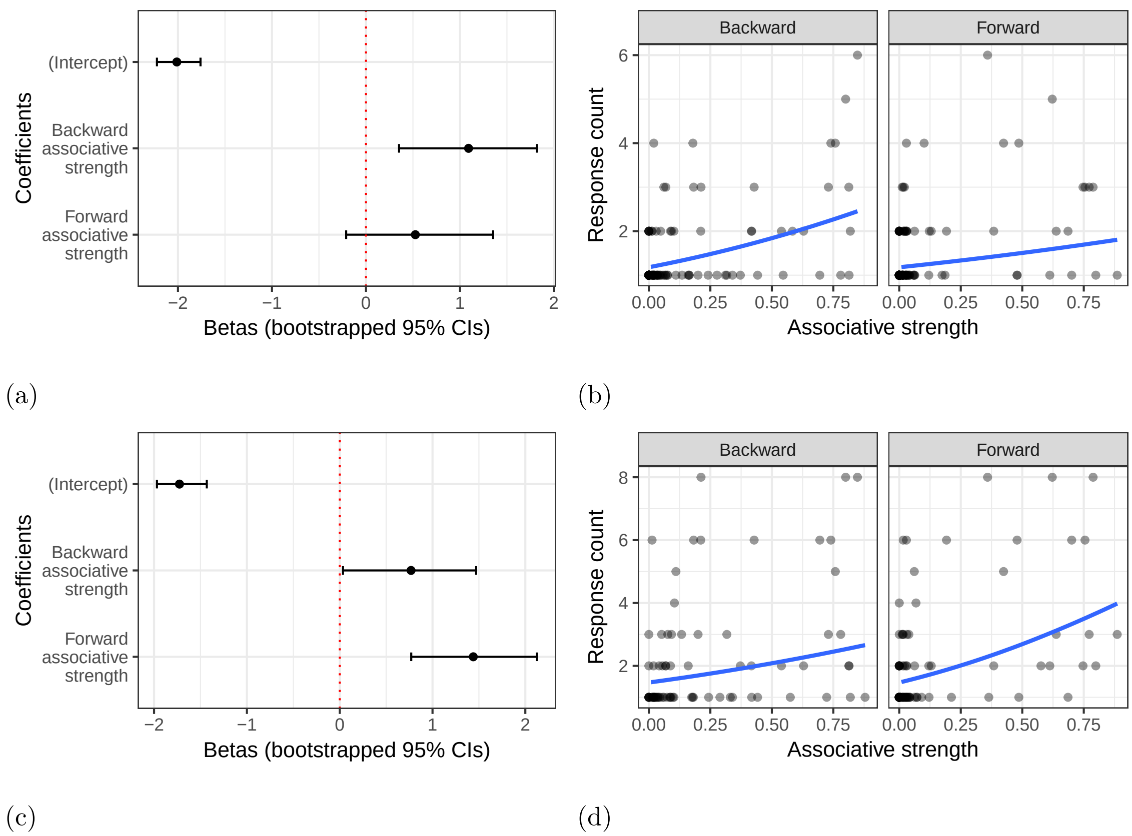


Figure 9. (a) Binomial mixed-effect regression coefficients for the effect of associative strength on the proportion of signalers producing each signal in the common-ground task. (b) Model fitted response counts (curve) and data (points). (c) The same model of behavior in the no-common-ground task. (d) Model fitted response counts (curve) and data (points) in the no-common-ground task.

However, it is obvious from Fig. 10b and from the low pseudo-R^2 values that something other than AS predicts almost all of the variation in guessing behavior. Like most other studies, the responses in Fig. 10b are spread widely, but unlike any other study so far, they cluster strongly at AS=0, do so across the full range of response counts, and do so for both forward and backward AS. Most of the variation in receiver behavior, then, is not meaningfully captured by salience as measured by AS. However, high response counts are spread across the range of values of forward AS, and low response counts across the range of values of backward AS, and this seems to be driving the model estimates.

Nonetheless, *something* must be guiding receiver behavior, since receivers were more likely to converge on correct guesses (Fig. 10c). However, it would be explanatorily vacuous to say that receivers are driven by correctness here since that raises the question of how the receiver knows what the correct guess is. The above model is thus misspecified in that it is missing an important predictor. What that might be, if not AS, is discussed below.

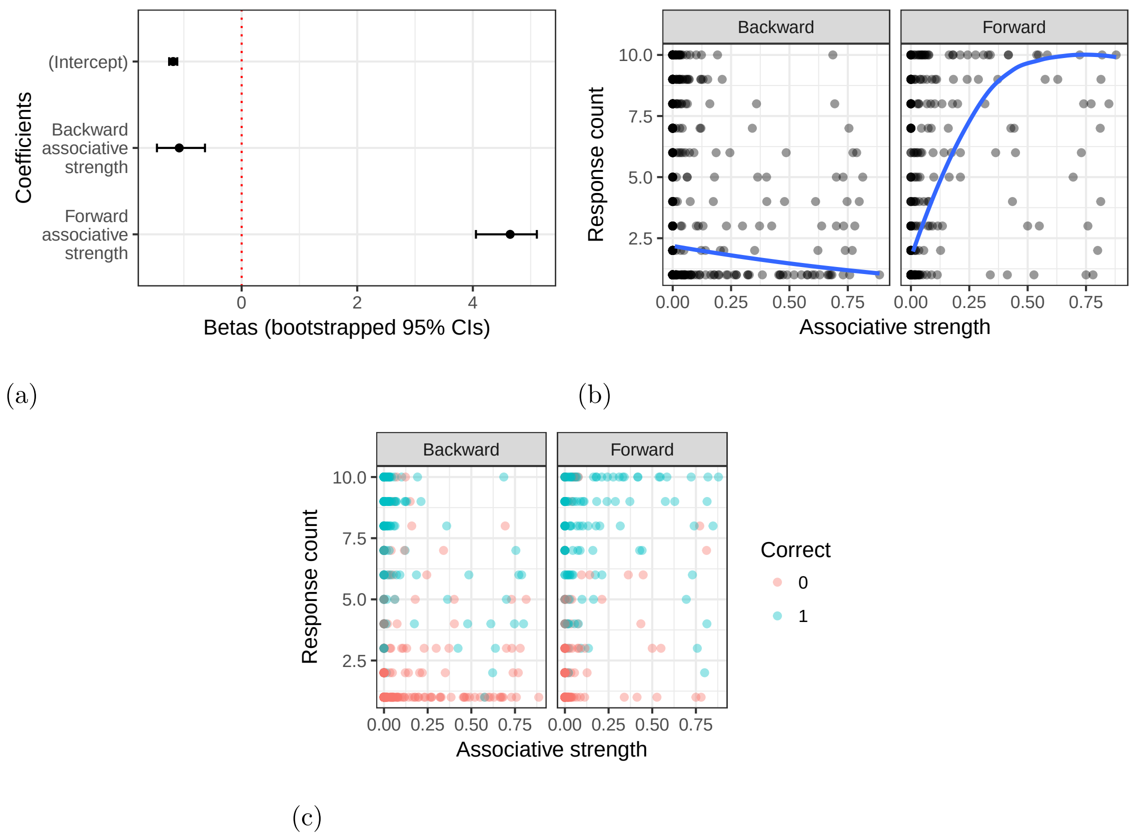


Figure 10. (a) Binomial mixed-effect regression coefficients for the main effects of associative strength on the proportion of receivers selecting each guess in the common-ground task (bootstrapped 95% CIs). (b) Model predictions (curve) and data (points). (c) Data colored by whether the guess was correct or not to indicate that receivers were able to converge on correct guesses.

Discussion

Constraining the meaning space promoted coordination success. Constraining it for the receiver yielded the highest success rate across all studies. Constraining it only for the signaler also promoted success, but did so to a smaller extent than constraining the signaling space (cf. Study 3).

Additionally, constraining the meaning space promoted a degree of perspective taking in the signaler. Across task versions, signalers exhibited a mixture of egocentric and allocentric behavior. If common ground drives coordination in the absence of convention, it does so in the sense that participants' attention is drawn to a small subset of shared knowledge, rather than having to rely on the vast background body of shared world knowledge.

Constraining the meaning space for the receiver caused AS to be a poor predictor of guessing behavior. To illustrate, consider a couple of signals unique to this task. To cue item 'bank', one participant signaled 'pig' and 9 of 10 receivers given this signal correctly guessed 'bank'. Coordination success was thus high, though pigs are not ordinarily salient features of banks, thus having low AS. Another participant signaled 'building'. Again, 9 of 10 receivers guessed correctly. Although banks are buildings, this feature is not usually salient (likely because it is non-specific). One is a metaphor or metonym ('pig', though possibly the signaler intended 'piggy' as a collocation) and the other is a semantic relationship not captured by AS ('building'). It is the prominence of such signals that distinguishes this task from all others, in terms of the cluster of responses at AS = 0 (Fig. 10b).

Perhaps there are several kinds of salience, then: one that is measured by AS and (at least) one other that is not. While being a building is not ordinarily a salient feature of banks, it may be salient *in this context* since none of the other items are buildings. It is thus a 'fully discriminative attribute' (Mangold & Pobel, 1988, p. 182). Mehta et al. (1994) call this 'Schelling salience', whereas they would describe AS as a measure of 'primary salience'. Schelling salience involves identifying a rule of selection that distinguishes one particular strategy (an item or signal) from all others (only banks are buildings, in the context of this task).

In that case, the present study supports the distinction between kinds of salience proposed by Mehta et al. However, they draw the distinction based on whether or not participants were told to coordinate, whereas we find that primary salience predicts most of the behavior in a non-communicative coordination task, some of the behavior in an open-ended communication task or one where the signal space is constrained for the signaler, and very little when the meaning space is constrained for the receiver. A potential explanation is that Mehta et al. find a role for Schelling salience because their task is much more constrained than our Study 2, and in that regard it is more like receiver behavior here. For instance, in their verbal coordination task, participants had to name makes of car or types of flower. In that case, their results do not represent a general solution to the problem of coordination, but rather reflect behavior when the semantic space is narrowly constrained, as it is in the present study.

Because this form of salience involves distinguishing one item from the others, it must depend on what the others are. In that case, the identification of a fully discriminative attribute could involve some kind of context-sensitive (i.e., flexible) reasoning, whereas primary salience simply requires that people respond with whatever occurs to them first, without further reflection. Our results are thus compatible with the claim that adult humans have two cognitive systems for inferring others' beliefs: one that is cognitively efficient but inflexible, and another that is more flexible, but cognitively demanding (Apperly & Butterfill, 2009). In that case, our results imply that contextual constraint in the meaning space is one factor driving differential recruitment of these systems.

However, just how to characterize the relevant systems is currently an open question. For instance, Postema (2008) argues that reasoning about salience is a creative process, while Apperly and Butterfill (2009) make no such claim. Similarly, Samson, Apperly, Kathirgamanathan, and Humphreys (2005) argue that different cognitive processes are involved in inhibiting egocentric perspective and identifying allocentric perspective, and it is not clear how this lines up with the distinction in Apperly and Butterfill (2009).

Note that is not enough to say that Schelling salience explains behavior here, since that does not explain why people made *these* particular choices and not others. Having a door and being a building are both fully discriminative attributes of banks in this context, but the latter seems more natural. The idea that 'naturalness' has something to do with salience is found already in Lewis (1969), but Cubitt and Sugden interpret this to mean a 'natural association of ideas' (2003, 201), which sounds very much like the sort of associative relationship captured by AS. Thus, if naturalness plays a role in Schelling salience, then the distinction between Schelling salience and primary salience becomes blurred. Regardless, while we could confidently predict what responses people would give in a task like Study 1 (and could do so, allowing for more error, in a task like Study 2), it is far from clear that anyone could do so for the current task in a non-post-hoc way. An explanation based on Schelling salience is thus unscientific, as things currently stand.

Study 5: Replication of main results

To test the robustness of the main findings in studies 1–4, we conducted a replication with a larger sample size. Below, 'task 1' refers to the task from Study 1; 'task 2' to that from Study 2, etc.

Participants

Participants were recruited from the same population as studies 1–4 using the same inclusion criteria. We recruited 200 signalers, 40 for each task. We recruited 400 receivers to guess the most common signals produced in the communicative task. We did not recruit receivers for the common-ground version of the constrained-meaning-space task since, as discussed above, it is behaviorally dissimilar to the other tasks.

Materials

The stimuli were identical to Studies 1–4.

Procedure

The procedure was identical to Studies 1–4, with one exception. Each receiver was given 20 signals, and only one signal per item. This change is motivated by a potential confound. In Study 2, receivers saw a random sample of 18 signals. This means that some receivers may have seen more than one signal for a given item. Even though they were unaware of this (since they did not know what the items were, so could not have known which item prompted each signal), they may have been motivated not to produce the same guess more than once, for different signals[5]. Here we avoid this potential confound by recruiting receivers for the 20 most common signals per item, and by showing each receiver only one signal per item.

Results

Success. The success scores for each task are displayed in Fig. 11. The pattern of results resembles that observed in Fig. 8. Modeling the effect of task on performance with a binomial mixed-effects regression (including a maximal random-effects structure), we found that performance in a constrained signal space (task 3) was better than in the open-ended task 2 ($\beta = 0.41$, $SE = 0.18$, $z = 2.32$, $p = .026$). When the random effects structure includes only intercepts for item and participant, performance in the non-communicative task 1 is significantly better than task 2 ($\beta = 0.38$, $SE = 0.09$, $z = 3.94$, $p < .001$). However, with the inclusion of a by-task random slope for item, this difference becomes non-significant ($\beta = 0.34$, $SE = 0.3$, $z = 1.11$, $p = .27$). We thus conclude that the previously observed difference between tasks 1 and 2 is not robust, though constraining the signal space (task 3) consistently boosts performance. As previously, there is no relationship between by-item success in task 1 and task 2 ($r = -.08$, $p = .72$). In task 2, success is positively predicted by receiver-POV salience ($\beta = 14.95$, $SE = 0.48$, $t = 31.01$, $p < .001$), and *negatively* predicted by signaler-POV salience ($\beta = -1.87$, $SE = 0.45$,

[5] We thank Dale Barr for this observation.

$t = -4.15$, $p < .001$). Thus, in this open-ended communicative task, shared knowledge can sometimes be a hindrance to success.

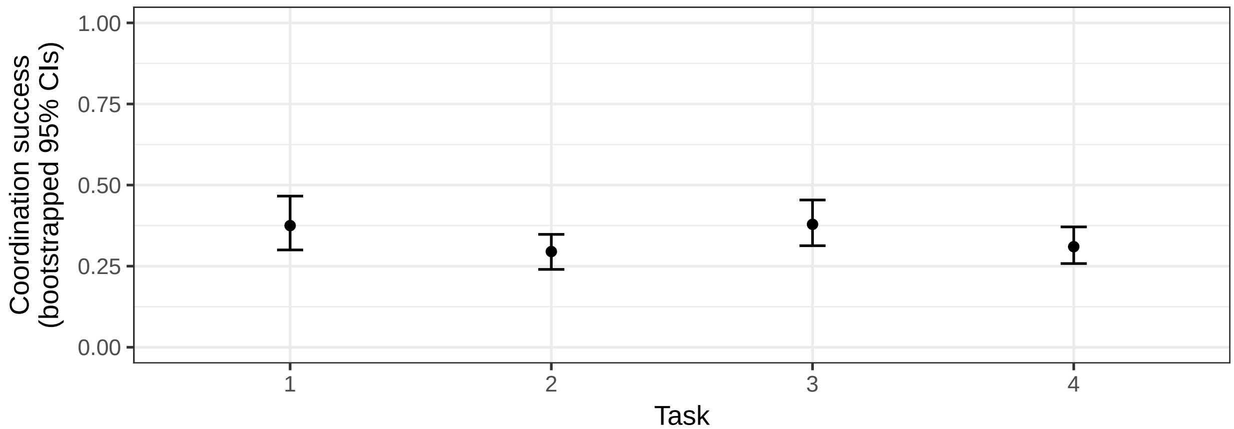


Figure 11. Mean values for coordination success (formula 1 for Study 1, formula 2 for others) across studies. Task 4 represents the no-common-ground task – see discussion above.

Response likelihood. As previously, forward associative strength (AS) was a strong predictor of response likelihood in non-communicative task 1 ($\beta = 5.78$, $SE = 0.29$, $z = 20.8$, $p < .001$). In communicative task 2, we included both forward and backward AS as predictors to test whether signalers were able to take perspective. We find a significant effect of forward AS ($\beta = 4.57$, $SE = 0.39$, $z = 11.55$, $p < .001$) but not backward AS ($\beta = 0.36$, $SE = 0.34$, $z = 1.06$, $p = .29$), indicating that signalers failed to take perspective in this open-ended signaling task.

These results were robust across different databases of word association norms (see supplementary material), except that Small World of Words (SWOW) norms produced a significant effect of backward AS ($\beta = 0.86$, $SE = 0.28$, $z = 3.09$, $p = .002$), though this was smaller than the effect of forward AS with the same norms ($\beta = 5.26$, $SE = 0.37$, $z = 14.3$, $p < .001$). Thus, there is evidence for a large effect of egocentricity across all norms databases, and evidence from one norms database for a small effect of allocentricity. We conclude that signaling in the open-ended task is predominantly egocentric.

The receivers were also egocentric (forward AS: $\beta = 4.4$, $SE = 0.18$, $z = 24.03$,

$p < .001$; backward AS:$\beta = 0.14$, $SE = 0.19$, $z = 0.73$, $p = 0.47$). Unlike Study 2, we found no evidence that receivers were more egocentric than signalers, since the interaction between communicative role and forward AS is not significant (forward AS $\beta = 4.53$, $SE = 0.17$, $z = 27.34$, $p < .001$; role $\beta = 0.42$, $SE = 0.06$, $z = 7.29$, $p < .001$; role:forward AS $\beta = 0.02$, $SE = 0.17$, $z = 0.11$, $p = .91$).

When we constrained the signal space (task 3), we observed a significant effect of backward AS ($\beta = 1.77$, $SE = 0.69$, $z = 2.57$, $p = .01$) and no effect of forward AS ($\beta = -0.76$, $SE = 0.86$, $z = -0.88$, $p = .38$). This confirms that constraining the signal space produces allocentric behavior. The result is consistent across word-association norms databases (see supplementary material).

In task 4, when common ground was emphasized by informing signalers that receivers would be choosing from a short list of targets visible to both receivers and signalers, forward AS was a significant predictor of signaling ($\beta = 2.41$, $SE = 0.35$, $z = 6.78$, $p < .001$) but backward AS was not ($\beta = 0.35$, $SE = 0.39$, $z = 0.88$, $p = .38$). However, using SWOW norms again showed an effect of backward AS ($\beta = 0.95$, $SE = 0.19$, $z = 5.05$, $p < .001$) which, though significant, was smaller than the forward AS effect with the same norms ($\beta = 3.02$, $SE = 0.32$, $z = 9.54$, $p < .001$).

In the no-common-ground version of task 4, signalers were given a short list of possible targets, but were told that receivers would not have access to the list when guessing. Here, we found an effect of both forward AS ($\beta = 3.71$, $SE = 0.41$, $z = 9.15$, $p < .001$) and backward AS ($\beta = 1.22$, $SE = 0.36$, $z = 3.41$, $p < .001$).

In this replication, the instructional manipulation (i.e., telling the signaler whether the receiver would see the same list of targets) did not have the same effect as in Study 4. In the original study, signalers were more allocentric in the common-ground task, and more egocentric in the no-common-ground task. This inconsistency suggests that the instructional manipulation is unreliable.

In order to model the effect of this instructional manipulation explicitly, we combined data for both the common-ground and no-common-ground tasks, and included task version as a main effect, along with a pair of two-way interactions, one between task and forward AS; the other between task and backward AS. There was a significant main effect of both forward AS ($\beta = 3.05$, $SE = 0.4$, $z = 7.53$, $p < .001$) and backward AS ($\beta = 0.68$, $SE = 0.32$, $z = 2.1$, $p = .035$). There were also significant effects for the interaction between task and forward AS ($\beta = 1.46$, $SE = 0.26$, $z = 5.52$, $p < .001$) and between task and backward AS ($\beta = 0.56$, $SE = 0.26$, $z = 2.19$, $p = .029$). However, there was no main effect of task ($\beta = -0.08$, $SE = 0.08$, $z = -1.03$, $p = .3$).

Thus, when the meaning space was constrained, participants produced both ego- and allocentric responses, though the effect of egocentric salience was stronger. This coheres with the analysis in Study 4 with the same predictors. The results also show that emphasizing common ground causes AS (whether forward or backward) to be a weaker predictor of signaling behavior. Thus, the instructional manipulation focusing on common ground does not necessarily boost allocentric behavior (as may have been suggested by the results from Study 4). Rather, it promotes signaling behavior that is poorly predicted by AS, unlike all the other tasks, and unlike the no-common-ground version of this task.

Fig. 12 compares the regression coefficients presented here with those from studies 1–4.

Discussion

We replicate the findings that (1) salience is a driver of both non-communicative and communicative behavior, but success in a non-communicative task does not predict success in a communicative task; (2) signalers are egocentric in an open-ended communicative task; (3) they are allocentric when the signaling space is constrained; (4) constraining the meaning space can also boost allocentricity to a degree, but responses were still more egocentric on the whole.

Task 4 appears to be qualitatively different from all the others, and is in need of

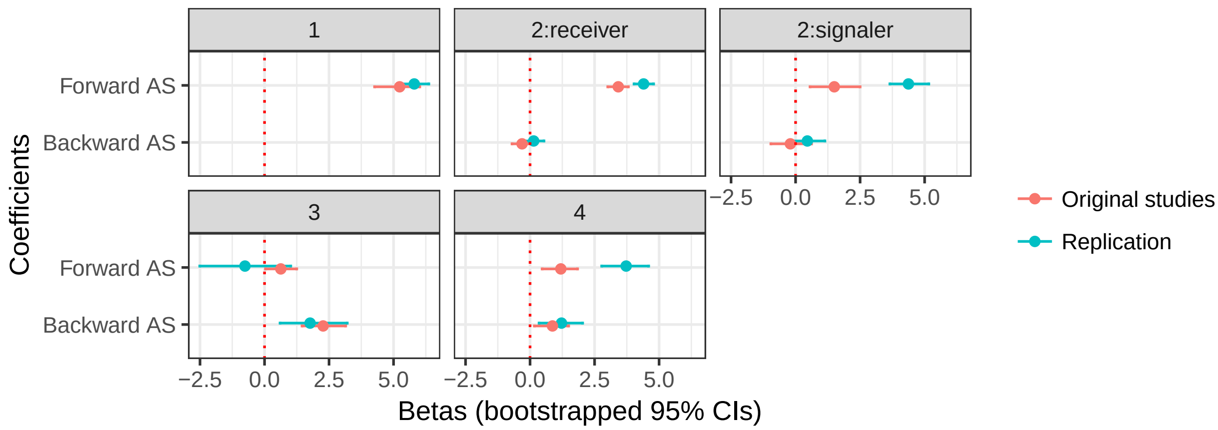


Figure 12. A comparison of the model parameters for forward and backward AS across tasks 1–4 reported here and those reported in studies 1–4 above. The parameters representing task 4 are those that model both datasets combined with task version as a fixed effect. Overall, behavior in tasks 1 and 2 was egocentric, task 3 was allocentric. The responses in task 4 were a mixture of ego- and allocentric, but predominantly the former.

further study, especially since so many lab-based studies of coordination during communication assume a constrained meaning space. The precise contents of the constrained meaning space likely play an important role. It is currently unclear whether the small effect of backward AS in task 4 means that the meaning space caused a small percentage of people to be more allocentric; or that it caused all people to be more allocentric to a small degree. We are currently investigating individual differences in signaling, which will hopefully address issues such as this.

Finally, we reiterate that while salience might drive response behavior (since forward AS predicts response rates in both tasks 1 and 2), this does not mean that salience predicts success across tasks, since success in task 1 was unrelated to success in task 2.

General discussion

Our principal aim was to test whether perspective taking and shared world knowledge (especially patterns of salience in that knowledge) explain human success at novel signaling tasks. We found that associative strength (AS) predicts behavior and success in a non-communicative Schelling task (Study 1), and thus serves as an empirical measure of salience. The same measure shows that people do not take perspective in a communicative task (Study 2). Constraining the signaling space by allowing a choice among just five signals (Study 3) boosted both coordination success and perspective taking. Constraining the meaning space such that the target could only be one of five cues (Study 4) boosted success. It also promoted signaler perspective-taking to a degree.

The above findings were replicated in Study 5 with a larger N, except that we found the instructional manipulation in Study 4 to be unreliable. However, the results combining common-ground and no-common-ground versions of task 4 were consistent in the original study and in the replication: constraining the meaning space boosted allocentric behavior, though responses were still egocentric overall.

Comparing results across studies, one finding is that salience is a general driver of behavior (though less so when the meaning space is constrained, especially in the common-ground task), but not a general driver of success. People share a great deal of world knowledge – surely everyone knows that money is a salient feature of banks – and this common knowledge drove participants to frequently generate 'money' in response to 'bank' in both communicative and non-communicative tasks. However, by-item success in the non-communicative task does not predict success in the communicative tasks, so the relationship between salience and success is task-dependent. Further, participants generated egocentric responses when signaling open-endedly, but generated allocentric ones when the signal space was constrained. Thus, it is not shared world knowledge that explains perspective taking, but contextual constraint. Successful perspective taking in a novel signaling task represents a special case, rather than a general explanation of human success across tasks.

Our results also problematize appeals to 'mutual salience'. We all share a great deal of world knowledge, and the patterns of salience in that knowledge do not differ wildly, at least, in the broad strokes relevant here (as mentioned, everyone knows that money is a feature of banks). However, the existence of *mutual* salience depends on whether people approach a problem from the same or from different directions. In a novel signaling task, signaler and receiver approach the problem from different directions, and this affects performance negatively, unless their attention is focused on a restricted meaning space. Thus, if common ground plays a role in perspective taking in novel signaling tasks, it is in this focal sense, as opposed to the vast, unconstrained body of background knowledge that people typically share.

Constraining the meaning space for the receiver boosted success tremendously (Study 4), but this caused AS to become a poor predictor of behavior. We discussed one possibility for what else might predict such behavior: Schelling salience, as opposed to primary salience. AS only measures the latter, and we identified some gaps that must be filled before the former can serve as a scientific explanation of behavior.

A common theme in the literature on perspective taking is the time-course for when (if at all) allocentric information becomes available or is integrated into utterance design (Barr, 2008; Hanna et al., 2003; Keysar et al., 2000). However, rather than focusing on this time-course, we have focused on contextual effects. Details about the time-course of ego- or allocentric information probably vary between our different tasks (perhaps signalers in study 3 fixate on ‘money’ before selecting ‘teller’), but this matter must be left for future research. The main motivating factor for our focus on contextual factors is that human performance at novel signaling tasks is relatively unconstrained (e.g., in Pictionary, where the target meaning could be one of thousands of possibilities), so if we explored only constrained contexts, we would not be able to draw conclusions about open-ended novel signaling tasks.

Overall, the results suggest that a signaling system based on salience is unlikely to afford perspective taking in the absence of a highly constraining context. A similar point could be made about any account where accessibility is foundational, such as Sperber and

Wilson (1995). While an ideal, rational agent (the sort described in game-theoretic accounts such as Lewis, 1969) might be able to use its world knowledge to make a choice that is both salient and allocentric, the evidence here shows that humans, though capable of evaluating the relevant world knowledge appropriately, are not always able to bring that relevant knowledge to bear on a particular problem. In the open-ended case, they are trapped by the salience of whatever is most likely to occur to them first, and do not spontaneously escape this egocentric bias.

These results paint a pessimistic view of perspective taking, raising the question of how our ancestors could ever evolve signaling conventions. One potential solution to this question is interaction. Apart from the signal itself, our participants did not interact, so we are currently exploring the effect of feedback and practice in follow-up studies. Garrod and Pickering (2004) argue that interaction boosts alignment, which may diminish the need to explicitly represent how one's interlocutor's representations differ from one's own, potentially reducing the explanatory burden placed on perspective taking. Indeed, Garrod et al. (2007) show that interaction ultimately leads to conventionalization in a graphical novel signaling task. Even though we think interaction would ultimately shoulder much of the explanatory burden, our aim was to evaluate the common claim that salience and perspective taking are key drivers of success. To evaluate the ego- or allocentricity of people's responses based on world knowledge, it was necessary to exclude any potential effect of communicative interaction, hence the use here of a one-shot task.

Since cognitive opacity was one of the issues identified above, a second solution would be to shift some of the inferential burden from cognitively opaque information (such as a novel signal) to cognitively transparent information, such as inferring someone's goals based on non-communicative behavior (cf. Tomasello, 1999). For instance, if one person observes another picking up their spear, they might infer that they are going hunting. This would constrain the context *prior* to a novel signaling interaction, and we have shown that a constrained context boosts success and (depending on the task) perspective taking.

Our results also suggest that producing a signal is cognitively different from producing an interpretation. Although some contexts promoted perspective taking in signalers, none did so for receivers. Constraining the receiver's choices (Study 4) had a markedly different effect from constraining the signaler's choices (Study 3), not only in terms of promoting success, but also in terms of the extent to which primary salience explains behavior.

We do not wish to claim that people cannot take the perspective of others in a broad sense. It is possible to put yourself in the emotional shoes of another person, or to work out that someone else can see something you can't, but neither of these involve overriding egocentric salience in world knowledge to find something allocentrically salient. Nor are we claiming that no one behaves allocentrically. One of our participants had a knack for doing this (see supplementary material), but a failure of perspective taking represents the more general case. We are currently undertaking an individual-differences study to identify why some people are better at this task than others. Nor do our results speak to a theory of mind. Person A might know that Person B knows that A knows that banks have tellers, but the results show that this needn't imply success at a novel signaling task about banks. In any case, A knowing that B knows that A knows that banks have tellers does not suffer from the asymmetries identified above.

Conclusions

We have shown that patterns of salience in world knowledge and perspective taking are not general drivers of success in a novel signaling task. Success in a non-communicative task did not generalize to an otherwise-similar communicative task. The same patterns of salience drive responses in both cases. However, the asymmetries inherent in communication mean that the same responses can be successful in one task type and not in the other. Though signalers and receivers share a great deal of world knowledge, signalers were typically unable to leverage this knowledge to override whatever was egocentrically salient and find something allocentrically salient.

It was contextual constraint rather than mutual salience that helped signalers behave allocentrically in certain specific situations. Receivers, on the the other hand, were always egocentric, perhaps because they were biased to assume informative signalers, or perhaps because interpreting a signal involved more uncertainty than generating a signal. We leave open the possibility that interaction is a major driver of human success in novel signaling tasks, since our aim here was to test particular claims about shared world knowledge and perspective taking.

Supplementary Information

All data (as well as the R scripts for statistical analyses) are available at `https://osf.io/frkeb/`.

References

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological review*, *116* (4), 953–970.

Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, *109* (1), 18–40.

Birch, S. A. (2005). When knowledge is a curse: Children's and adults' reasoning about mental states. *Current Directions in Psychological Science*, *14* (1), 25–29.

Birch, S. A., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, *18* (5), 382–386.

Bögels, S., Barr, D. J., Garrod, S., & Kessler, K. (2015). Conversational interaction in the scanner: mentalizing during language processing as revealed by MEG. *Cerebral cortex*, *25* (9), 3219–3234.

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.

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, *1* (2), 274–291.

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic bulletin & review*, *16* (5), 893–900.

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107* (3), 1122–1134.

Brown-Schmidt, S., & Hanna, J. E. (2011). Talking in another person's shoes: Incremental perspective-taking in language processing. *Dialogue & Discourse*, *2* (1), 11–33.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior research methods*, *41* (4), 977–990.

Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of political Economy*, *97*(5), 1232–1254.

Cane, J. E., Ferguson, H. J., & Apperly, I. A. (2017). Using perspective to resolve reference: The impact of cognitive load and motivation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 591–610.

Clark, H. H. (1996). *Using language.* Cambridge university press.

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. *Advances in psychology*, *9*, 287–299.

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground at the understanding of demonstrative reference. *Journal of verbal learning and verbal behavior*, *22*(2), 245–258.

Converse, B. A., Lin, S., Keysar, B., & Epley, N. (2008). In the mood to get over yourself: Mood affects theory-of-mind use. *Emotion*, *8*(5), 725.

Cubitt, R. P., & Sugden, R. (2003). Common knowledge, salience and convention: A reconstruction of David Lewis' game theory. *Economics and Philosophy*, *19*, 175-210.

De Deyne, S., Navarro, D., & Storms, G. (2012). Better explanations of lexical and semantic condition using networks derived from continued rather than single word associations. *Behavior Research Methods*, *45*, 450–498.

Dumontheil, I., Apperly, I. A., & Blakemore, S.-J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental science*, *13*(2), 331–338.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of personality and social psychology*, *87*(3), 327.

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, *40*(6), 760–768.

Fay, N., Arbib, M., & Garrod, S. (2013). How to bootstrap a human communication system. *Cognitive science*, *37*(7), 1356–1367.

Galantucci, B. (2009). Experimental semiotics: A new approach for studying communication as a form of joint action. *Topics in Cognitive Science*, *1*(2), 393–410.

Gann, T. M., & Barr, D. J. (2014). Speaking from experience: Audience design as expert performance. *Language, Cognition and Neuroscience*, *29*(6), 744–760.

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., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8-11.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: the naïve theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287-292.

Goldin, G., & Darlow, A. (2013). *Turkgate (version 0.4.0)[software].* Available from https://github.com/gideongoldin/TurkGate.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*(1), 43–61.

Hilliard, C., & Cook, S. W. (2016). Bridging gaps in common ground: Speakers design their gestures for their listeners. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(1), 91.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*(1), 91–117.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38.

Keysar, B., Barr, D. J., Balin, J. A., & Paek, T. S. (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory* *and Language*, *39*(1), 1–20.

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current directions in psychological science*, *7*(2), 46-50.

Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. J. Aitken, R. W. Bailey, & N. Hamilton-Smith (Eds.), *The computer and literary studies.* Edinburgh: Edinburgh University Press. Available online at http://eat.rl.ac.uk.

Kvålseth, T. O. (1985). Cautionary note about R^2. *The American Statistician*, *39*(4), 279–285.

Langdon, R., Davies, M., Coltheart, M., et al. (2002). Understanding minds and understanding communicated meanings in schizophrenia. *Mind and Language*, *17*(1 & 2), 68–104.

Levinson, S. C. (2006). On the human "interaction engine". In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 39–69). Oxford: Berg.

Lewis, D. (1969). *Convention: A philosophical study.* Cambridge, MA: Harvard University Press.

Litman, L., Robinson, J., & Abberbock, T. (2017). Turkprime. com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior research methods*, *49*(2), 433–442.

Little, H., Eryılmaz, K., & de Boer, B. (2017). Conventionalisation and discrimination as competing pressures on continuous speech-like signals. *Interaction studies*, *18*(3), 352–375.

Mangold, R., & Pobel, R. (1988). Informativeness and instrumentality in referential communication. *Journal of Language and Social Psychology*, *7*(3-4), 181–191.

Mehta, J., Starmer, C., & Sugden, R. (1994). The nature of salience: An experimental investigation of pure coordination games. *The American Economic Review*, *84*(3), 658-673.

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, *49*(2), 201–213.

Misyak, J., Noguchi, T., & Chater, N. (2016). Instantaneous conventions: The emergence of flexible communicative signals. *Psychological science*, *27*(12), 1550–1561.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida word association, rhyme, and word fragment norms. *Behavior Research Methods*, *36*(3), 402–407. Available from http://w3.usf.edu/FreeAssociation/.

Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive psychology*, *58*(2), 220–249.

Perlman, M., & Cain, A. A. (2014). Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture*, *14*(3), 320–350.

Postema, G. J. (2008). Salience reasoning. *Topoi*, *27*, 41-55.

Rubio-Fernández, P. (2008). On the automaticity of egocentricity: A review of the egocentric anchoring and adjustment model of perspective taking. *UCL Working Papers in Linguistics*, *20*, 247–274.

Ryskin, R. A., Benjamin, A. S., Tullis, J., & Brown-Schmidt, S. (2015). Perspective-taking in comprehension, production, and memory: An individual differences approach. *Journal of Experimental Psychology: General*, *144*(5), 898.

Samson, D., Apperly, I. A., Kathirgamanathan, U., & Humphreys, G. W. (2005). Seeing it my way: A case of a selective deficit in inhibiting self-perspective. *Brain*, *128*(5), 1102–1111.

Schelling, T. (1960). *The strategy of conflict.* Cambridge, MA: Harvard University Press.

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

Scott-Phillips, T. C., Kirby, S., & Ritchie, G. R. (2009). Signalling signalhood and the emergence of communication. *Cognition*, *113*, 226-233.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Malden, MA: Blackwell Publishing.

Sulik, J. (2018). Cognitive mechanisms for inferring the meaning of novel signals during symbolisation. *PloS One*, *13*(1), e0189540.

Surtees, A., Apperly, I., & Samson, D. (2013). The use of embodied self-rotation for visual and spatial perspective-taking. *Frontiers in human neuroscience*, *7*.

Todd, A. R., Forstmann, M., Burgmer, P., Brooks, A. W., & Galinsky, A. D. (2015). Anxious and egocentric: How specific emotions influence perspective taking. *Journal of Experimental Psychology: General*, *144*(2), 374.

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

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190.

Verbeek, B. (2008). Conventions and moral norms: The legacy of Lewis. *Topoi*, *27*(1-2), 73–86.

Wardlow, L. (2013). Individual differences in speakers' perspective taking: The roles of executive control and working memory. *Psychonomic bulletin & review*, *20*(4), 766–772.

Wardlow Lane, L., & Ferreira, V. S. (2008). Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1466–1481.

Wu, S., Barr, D. J., Gann, T. M., & Keysar, B. (2013). How culture influences perspective taking: differences in correction, not integration. *Frontiers in human neuroscience*, *7*, 822.

Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological science*, *18*(7), 600–606.