# Conventions, coordination, and arbitrariness

Bob van Tiel

Bart Geurts

*Abstract*

A language is a system of conventions on many levels: phonological, morphological, syntactic, semantic, and so on. Like all social conventions, language conventions are associated with regularities in interactions between the members of a population. A theory of the evolution of language ought to explain, at a minimum, how such regularities can emerge and persist over time. There are two main theoretical approaches to these questions, which are quite different and often pitted against each other. This chapter argues, instead, for an inclusive approach, which makes it possible to capture continuities between animal and human communication, while at the same time allowing for marked discontinuities in relation to normativity, coordination and cooperation, and common ground.

## 1. Introduction

Our social lives are governed by myriads of conventions. We greet the people we meet, keep to the same side of the road, eat with knife and fork, exchange goods against money, flush the toilet after using, do not belch, break wind, or pick our noses in public, and so on. Our communicative exchanges, too, are governed by all sorts of language conventions, which are a major subclass of social conventions. We choose words to conform with established usage, pronounce them in accordance with phonological rules, observe the rules of grammar, and so on. In brief, conventions govern our social lives in general, and our communicative exchanges in particular.

On the other hand, conventions are also the products of everyday discourse and other forms of social activity. We live by the conventions that we make. Occasionally, conventions are established by explicit agreement or by

1

decree. But more often than not, they emerge from our interactions quasi-organically, as 'tacit agreements'. This is how language conventions generally establish themselves, and since the current chapter is primarily about language conventions, one of our main topics is how such 'tacit agreements' come about and are sustained.

There are many questions that can and have been raised about conventions, but the most important ones are these:

- What are conventions and how do they differ from mere behavioural regularities?
- What purpose(s) do conventions serve?
- How are conventions established, if not by explicit agreement or decree?
- Once established, how are conventions sustained?

All these questions are relevant from an evolutionary perspective, but the last two especially so, since they concern the dynamics of conventions, and are therefore the most obvious targets for evolutionary analyses.

About fifty years ago, subsidised thinking about conventions started in earnest with the work of David Lewis (1969; 1975), who offered the following as a general description of our quarry:

> Conventions are regularities in action [...] which are arbitrary but perpetuate themselves because they serve some sort of common interest. Past conformity breeds future conformity because it gives one a reason to go on conforming; but there is some alternative regularity which could have served instead, and would have perpetuated itself in the same way if only it had got started. (Lewis, 1975, pp. 4–5)

This passage contains several ideas that have figured prominently in most subsequent accounts of convention:

(i) *Regularities in action:* On Lewis's account, conventions are lawlike. For example, in most countries of the world, drivers are expected *always* to keep to the right side of the road. But despite the widespread belief to the contrary, it does not seem to be the case that conventions are inherently lawlike (Gilbert, 1989; Millikan, 1998; Moore, 2013). For example, wearing black at a funeral is a conventional pattern of behaviour that is still widely practiced, but as a rule it is not strongly expected anymore. Hence, strictly speaking, wearing black at funerals is not a Lewis convention, though it is still a 'regularity in action', and therefore fits Lewis's informal description above.

The distinction between lawlike conventions and others is especially relevant to our current purposes, because many language conventions belong to the others (Geurts, 2018; Millikan, 1998). For example, there are various conventional ways of saying goodbye in English, including 'Goodbye', 'Bye', and 'See you', and although their usage conditions are not quite the same, there is so much overlap that speakers are rarely required to employ any one in particular. More generally, the prevalence of synonyms and near-synonyms (e.g., 'couch/sofa', 'big/large', 'fortnight/two weeks', 'heart attack/myocardial infarction') indicates that lexical conventions are not lawlike. Lexical conventions *are* 'regularities in action', but they do not tell us how to act. Rather, they are templates on which we can model our actions, if we are so inclined.

(ii) *Persistence:* According to Lewis, conventions 'perpetuate themselves'. Although this way of speaking has caught on, it is misleading, because surely conventions do not sustain themselves, but are sustained by populations of agents. Still, the insight that conventions persist is fundamental, for this is what enables conventions to regulate and stabilise social interactions. Persistence was the main explanandum of Lewis's analysis of conventions, and it has remained a central topic ever since (cf. also Geurts, this volume, on the nexus between persistence and normativity).

(iii) *Arbitrariness:* Lewis regards conventions as regularities in action for which there is 'some alternative regularity which could have served instead.' This is his way of saying that conventions are arbitrary. Arbitrariness is widely recognised as a defining feature of conventions, and many behavioural patterns do not count as conventional because they lack this feature. Examples are eating, breathing, and walking upright, all of which are non-arbitrary regularities in action, and therefore non-conventional.

Arbitrariness has become a topic of interest in connection with the evolution of language. In particular, it has been suggested that gesture provides a much better channel than speech for producing non-arbitrary signs, and that, therefore, gestural forms of communication must have played an important role in the earlier stages of language evolution (Arbib, Liebal, & Pika, 2008; Irvine, 2016; Planer & Sterelny, 2021). In recent years, this line of argument has received pushback from studies reporting correlations between the sound structure of signs and their meanings, both in development and across languages (Ćwiek et al., 2021; Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015; Sidhu, Westbury, Hollis, & Pexman, 2021). Such sound-meaning correlations are said to reflect the inherent 'iconicity' of language.

It is currently a fashionable notion that iconicity was a necessary requirement for language to emerge. However, as we will see, there is strong evidence that, at least in principle, iconicity is not needed for this. Moreover, the recent wave of iconicity studies that argue against an alleged 'dogma of arbitrariness' (Ćwiek et al., 2021, p. 2) use the word 'arbitrariness' in a different sense than Lewis and other authors on language conventions, thus setting up a straw man that may be easy to put down, but is irrelevant to our topic (Section 5).

Regularity in action, persistence, and arbitrariness are among the most widely cited features of conventions, but there are more, all of which were already discussed by Lewis:

(iv) *Normativity:* Driving on the right side of the road is obviously a regularity in action, but it is not just that: we consider it *correct* to do so, and expect each other to drive on the right *because* it is the correct thing to do. The same holds for greeting people we meet, flushing the toilet after using, uttering the word 'sofa' (rather than, say, 'potato') to refer to sofas, stressing the second syllable of 'tomato' (rather than the first or the last), putting articles before nouns (rather than the other way round), saying 'Okay' to express approval, and so on. In brief, we are normative about our conventions, language conventions included (Bartsch, 1987; Bicchieri, Muldoon, & Sontuoso, 2018).

More generally, normativity is a core feature of the human condition. We routinely treat each other's actions as correct or incorrect, appropriate or inappropriate, good or bad, and so on. It is a moot issue whether other species behave normatively at all, but it is agreed that we are unique for the extent and complexity of our normative interactions (Geurts, this volume).

The distinction discussed under (i) between conventions that are lawlike and those that are not can be sharpened by casting it in normative terms: the former are prescriptive, the latter are permissive. By convention, a traffic light has prescriptive force when red, and permissive force when green, and whereas keeping to the right side of the road is *the* correct strategy for drivers to adopt, using the word 'sofa' is *a* correct way of referring to sofas.

(v) *Coordination:* Lewis conventions are devices that enable us to coordinate our actions in situations where interdependent choices have to be made and all of us have more or less the same preferences regarding the combinations of our choices. More succinctly, Lewis conventions tell us how to coordinate

our actions when 'coincidence of interests predominates.' (Lewis, 1969, p. 69) The next section discusses in some detail what that means, but for now we illustrate the main idea with one of Lewis's own examples:

> In my hometown of Oberlin, Ohio, until recently all local telephone calls were cut off without warning after three minutes. Soon after the practice had begun, a convention grew up among Oberlin residents that when a call was cut off the original caller would call back while the called party waited. (Lewis, 1969, p. 43)

Imagine that, some day in the mid-1960s, Oberlin resident A calls her fellow resident B, and their conversation is duly cut off after three minutes. Supposing that both A and B want to resume their phone call, each is faced with the same choice: call back or wait. Since the caller pays for the call, each prefers the other to call back, but this preference is weaker than the wish to resume the call. So both A and B have a conflict of interests, but 'coincidence of interests predominates', and therefore it does not matter who calls and who waits, as long as they do not act alike. What are they to do? Convention resolves the quandary: A calls and B waits.

While Lewis held that conventions are coordination devices by definition, it has since been argued that this view is too narrow (Davis, 2003; Geurts, 2018; Millikan, 1998; Sugden, 2005). For example, fashion in clothing may be seen as a form of action coordination, but not necessarily in Lewis's sense of the word, and although we heavily rely on language to coordinate our activities, it does not seem right to say that we are addressing Lewisian coordination problems whenever we use the word 'potato', a falling intonation contour, or interrogative syntax. Hence, although it may well be true that, generally speaking, conventions are coordination devices, our pre-theoretic concept of coordination may be broader than what is covered by Lewis's definition. Having said this, we will follow the bulk of the literature on convention by concentrating our attention on coordination problems. More about coordination in Section 4.

(vi) *Common ground:* Many of the conventions we depend upon in our everyday interactions are common ground between us: we know, and trust each other to know, that we are supposed to greet the people we meet, drive on the right, and eat with knife and fork, and we know, and trust each other to know, the established uses of the words and grammatical constructions of the language we share. All this and much more is common ground, and

accordingly Lewis writes common ground into his analysis of convention: a regularity in action is a Lewis convention only if it is common ground that it is arbitrary and serves to solve a coordination problem.[1]

It has often been claimed that the cognitive requirements for having common ground are so demanding that Lewis conventions are bound to be the prerogative of adult humans, and unattainable for small children and non-human species (e.g., Millikan, 1998; Moore, 2013; Skyrms, 1996, 2010). However, this objection is off the mark, because it is premised on the assumption that common ground is a configuration of psychological states, notably knowledge or beliefs, which has a special recursive structure. On this view, $p$ is common ground in a population if and only if all its members believe that $p$, all believe that all believe that $p$, and so on.

However, that wasn't Lewis's view. Rather than defining common ground in psychological terms, Lewis defined it in terms of 'reasons to believe'. According to Lewis, $p$ is common ground in a population if and only if all its members have reason to believe that $p$, all have reason to believe that all have reason to believe that $p$, and so on. Clearly, reasons to believe are not beliefs, any more than reasons to complain are complaints. A man can have reason to believe that his house is on fire without even considering the possibility that his house is on fire. Hence, on Lewis's view, common ground is not a configuration of psychological states, and he made it very clear that the objections that were to be raised by his critics did not affect his theory (Lewis, 1969, pp. 52–57). Apparently, his critics skipped this part of the book.

Still, even if one of the main motivations for developing alternatives to Lewis's account turns out to be mistaken, the alternatives are of considerable interest in their own right, as they offer austere and elegant explanations not only of what conventions are, but also of the ways they evolve. It is to this family of theories that we turn now.

## 2. Coordination games and the evolution of signalling conventions

Lewis conventions are behavioural regularities that enable coordination in recurring situations where coincidence of interests predominates. Lewis develops this idea using the mathematical theory of 'games', which in this con-

---

[1]In his monograph of 1969, Lewis uses the term 'common knowledge', which was a patent misnomer, as he conceded six years later (Lewis, 1975). We use 'common ground' instead of Lewis's 'common knowledge'.

text is a playful label not only for many games in the everyday sense of the word, but for interactions between agents ('players') in general. Coordination games are an important subclass of games, exemplified by the following scenarios:

(a) Two drivers approach each other on the same road. Both drivers have to decide whether to keep to the left or to the right; the optimal outcome for both is that they make the same choice, regardless which one.

(b) A telephone call is cut off, and the two parties have to decide whether to call back or wait; the best outcome for both is that one calls back and the other waits. However, each also has a personal preference for the other party to call back, because the caller pays for the call.

(c) Two people are invited to a dinner party. Both have to decide whether or not to bring a gift for their host. The best outcome for both is that they make the same choice, preferably to not bring a gift, which saves money and the effort of buying one.

As shown in Fig. 1, these game situations can be represented by means payoff matrices, in which cell specifies the payoffs for each combination of strategies. (The first value is the payoff for the row player; the second for the column player.) For example, the first matrix shows that both drivers have two possible strategies: driving on the right and driving on the left. These payoffs represent how satisfied each driver is with the outcome.

These coordination games have two main features in common. First, in each case, coincidence of interests predominates, which means that players' payoffs are strongly correlated. But payoffs need not be *perfectly* coordinated, as Figs. 1b and 1c illustrate. Second, each scenario has more than one 'coordination equilibrium': a combination of choices that, once made, rules out that any player might improve his payoff by unilaterally deviating from his choice. Players may prefer different equilibria (Fig. 1b), or they may both prefer one equilibrium over the other (Fig. 1c), but in both cases players cannot unilaterally improve their payoffs when they are in one of the equilibria, even if it is a suboptimal one. Games with multiple coordination equilibria in which coincidence of interests predominates are called 'coordination problems', and Lewis conventions solve these coordination problems by directing all players to the same coordination equilibrium.

If this is what conventions are for, the question remains how they are established. Lewis considers three possible mechanisms:

|  | Driver B | |
| --- | --- | --- |
|  | Keep left | Keep right |
| **Driver A** Keep left | 1, 1 | 0, 0 |
| Keep right | 0, 0 | 1, 1 |

(a) Two drivers approach each other on the same road.

|  | First receiver | |
| --- | --- | --- |
|  | Call back | Wait |
| **First caller** Call back | 0, 0 | 1, 2 |
| Wait | 2, 1 | 0, 0 |

(b) A telephone call is cut off.

|  | Guest B | |
| --- | --- | --- |
|  | Bring a gift | Bring nothing |
| **Guest A** Bring a gift | 2, 2 | 1, 0 |
| Bring nothing | 0, 1 | 3, 3 |

(c) Two guests attending a dinner party.

Figure 1: Payoff matrices for three coordination games.

(i) Players explicitly agree to adopt a convention.
(ii) Players select a salient coordination equilibrium that somehow 'stands out' for all of them.
(iii) Players copy successful precedents.

Since the first language conventions obviously cannot be explained by appeal to explicit agreement, and we are mainly concerned with language conventions, we are left with salience and successful precedents (Lewis treats the latter as a special case of the former, but that is as it may be). Successful precedents are the crucial ingredient in Skyrms's (1996) account of the evolution of language conventions, which we turn to now. Our discussion of Skyrms's account will be brief and informal; we refer to Mühlenbernd and Baumann (this volume) for a more thorough presentation.

Skyrms studies the evolution of so-called 'signalling games', which are a subclass of the coordination games invented by Lewis. In its simplest form, a signalling game involves two players: a sender and a receiver. The sender observes a state of affairs and produces a signal so as to enable the receiver to act appropriately, given that state of affairs. To make this more concrete, imagine a simplified version of the alarm call system used by vervet monkeys (Price et al., 2015; Seyfarth, Cheney, & Marler, 1980). There are two possible states of affairs: one in which the sender sees a leopard and one in which the sender sees an eagle. There are two signals for the sender to choose from: chuttering and barking. There are two actions for the receiver to choose from: climbing a tree and hiding in a bush. The appropriate action when encountering a leopard is to climb a tree; when encountering an eagle it is hiding in a bush. In this signalling game, the sender and the receiver each have four possible strategies:

*Sender strategies*
   S1: Chutter if you see a leopard; bark if you see an eagle.
   S2: Bark if you see a leopard; chutter if you see an eagle.
   S3: Chutter if you see a leopard or an eagle.
   S4: Bark if you see a leopard or an eagle.

*Receiver strategies*
   R1: Climb a tree if you hear a chutter; hide if you hear a bark.
   R2: Hide if you hear a chutter; climb a tree if you hear a bark.
   R3: Climb a tree if you hear a chutter or a bark.
   R4: Hide if you hear a chutter or a bark.

Assuming that both the sender and the receiver obtain the highest payoff when the receiver carries out the appropriate action, this signalling game has two optimal solutions: (S1, R1) and (S2, R2). In the first case, the sender selects S1 and the receiver selects R1; in the second case, the sender selects S2 and the receiver selects R2. These two combinations of strategies are obviously the best; they are called 'signalling systems'.

Given that, by definition, there are at least two signalling systems for any signalling game, how do players manage to settle on the same signalling system? Skyrms addresses this question using evolutionary game theory. Evolutionary game theory offers tools for studying the dynamics of strategies emerging from interactions within a population. The key concept is that of an 'evolutionarily stable strategy'. A strategy adopted by a population is

evolutionarily stable if it cannot be overturned by a small number of players using a different strategy (Maynard-Smith & Price, 1973).

Skyrms seeks to determine (i) whether signalling systems are likely to evolve at all, and (ii) if a signalling system emerges, whether it is able to sustain itself, i.e., whether signalling systems are evolutionarily stable. In order to address these questions Skyrms uses analytical methods, especially evolutionary game theory, and computational models. A computer simulation starts with a population in which strategies may be distributed randomly, proportionally, or otherwise. On each round of play, pairs of individuals engage in sender-receiver interactions, which shape the distribution of strategies in the following round, as winning strategies become more frequent while losing strategies become rarer. By running many such simulations, it can be seen that certain strategies tend go extinct while others become dominant.

There are at least two (not necessarily incompatible) ways of conceptually grounding this 'replicator dynamics': biological and cultural. On a biological construal, payoffs represent reproductive fitness, so that players using a winning strategy have more offspring than players using a losing one, with their offspring being genetically predisposed to use the same strategy as their parents. On a cultural construal, the replicator dynamics is driven by social learning: players learn to adapt their strategies based on the payoffs they obtain in repeated interactions. In the case of real-world vervets, both biological and cultural factors seem to be at play.

With respect to our simplified vervet model, Skyrms (1996) found that, no matter how the signalling strategies are initially distributed, sender-receiver populations practically always converge on one signalling system or other:

> Almost every state of the population is carried by the dynamics to one signaling system or another. *The emergence of meaning is a moral certainty.* Which signaling system is selected depends on the initial proportions. (Skyrms, 1996, p. 93)

This result accords with our intuitions, but subsequent research has qualified the picture substantially (Hofbauer & Huttegger, 2008; Huttegger, Skyrms, Smead, & Zollman, 2010; Wagner, 2009). It has been shown that the probability of signalling systems emerging depends on various features of the signalling game, including the number of possible signals, the prior probabilities of states of the world, the payoffs for senders and receivers, and the structure of interactions, like who is more or less likely to interact with whom (see Mühlenbernd & Baumann, this volume, for further discussion).

This line of research has been groundbreaking, first, because it shows how the emergence and persistence of signalling systems can be modelled to high degrees of precision; secondly, because it proves that communicative conventions can evolve even when the relations between signals and their meanings are completely arbitrary; and thirdly, because it demonstrates how such systems can evolve in populations of creatures with minimal cognitive capacities, since Skyrms-style conventions don't require strategic reasoning, mental-state attribution, or any other form of higher cognition.

*3. Towards an inclusive approach*

In the foregoing, we reviewed two theories of convention that, on the face of it, seem to be worlds apart. Whereas Lewis's account requires rational decision making, reasoning about others' mental states, and common ground, Skyrms doesn't need any of that, and makes do with quite basic forms of interaction between individuals with quite basic cognitive capacities, if indeed they have cognition at all. Compared to this approach, Lewis's proposal seems positively baroque. What's more, due in large measure to its austerity, Skyrms's account has a much wider scope than Lewis's, as it applies not only to adult humans, but also to their children, to non-human species (including vervet monkeys and chickens), and even to inanimate signallers, like neurons, for example. In brief, Skyrms's account is hands down the more frugal as well as the more general of the two.

However, this generality comes at a cost. The generality of a theory is inversely proportional to its ability to account for distinctive features of sub-classes of phenomena in its purview. Therefore, it doesn't come as a surprise that, although its wider scope is certainly a virtue, Skyrms's framework is ill-equipped to deal with the complexities of human conventions, language conventions in particular. For example, although it seems generally true that primitive mechanisms can enable conventions to emerge and persist, in human societies there are more sophisticated mechanisms, as well, like sanctioning, censuring, deliberation, and others, and it is fairly obvious that such mechanisms are relevant for the dynamics of social conventions in our lineage, even if they may not tell the whole story.

Therefore, the two approaches are better seen as complementing rather than excluding one another, with Skyrms providing a framework that applies quite generally, while the Lewisian framework is tailored to social conven-

tions in our lineage, including language. Moreover, since Lewis conventions are richly structured and their various components are not necessarily an evolutionary package deal, it is natural to suppose that they evolved piecemeal. For example, it might be hypothesised (plausibly, in our view) that incipient forms of normativity, like sanctioning, predated the capacity for having beliefs about others' beliefs.

Let's elaborate this picture somewhat by having another look at signalling games. The concept of a signalling game is due to Lewis (1969); evolutionary theories of signalling started taking off in the 1990s, with the work of Skyrms and others. It is important to note that signalling games are a lot more central to these later developments than they were to Lewis's project. This is understandable, since Lewis was concerned with human communication only, and the 'sample signals' he prefaced his discussion with are at best of marginal interest to that topic, involving as they do lanterns, flags, impromptu gesturing, trail marks, and railroad signs (Lewis, 1969, pp. 122-130).

It hardly needs arguing that our communicative practices go way beyond signalling, and in many dimensions, too. To name only three: languages are vastly more complex than any other medium of communication reported so far, language and language use are thoroughly normative (Geurts, this volume), and our communicative exchanges rely on the attribution of mental states (Harris, this volume, Moore, this volume). But at the same time, our linguistic exchanges retain relatively primitive conventional features that are naturally viewed as pure Skyrms-style signalling. For example, it is well established that, in face-to-face settings, addressees are active contributors to the discourse even when it isn't their turn to speak. Their contributions come in the form of various kinds of vocal cues ('okay', 'yes', 'right', 'yeah', 'mhm') and body language (nodding, shrugging, eye gaze, facial expressions). This 'back-channelling', as it is called (Yngve, 1970), is not just an idle accompaniment to the speaker's performance. As was shown by Bavelas, Coates, and Johnson (2000), speakers *need* these responses, and may become seriously disoriented if they cease coming. In their study, Bavelas et al. asked one participant to tell a story from their own experience, while a second participant was instructed, unbeknown to the other, to perform an irrelevant task while listening to the story. This additional task diminished the hearers' back-channel responses, which turned out to have a deleterious effect on the quality of the story-telling, causing speakers to hesitate, go in circles, and become needlessly specific.

Since, for the most part, back-channelling goes under the awareness radar of senders and receivers alike, it is not a form of intentional communication. There are exceptions, of course. A speaker may very deliberately and emphatically utter 'mhm' to express disbelief, for example. But in the normal run of events, a muttering of 'mhm' is not a speech act on any account. Instead, back-channel cues are better viewed as signals whose main purpose is to ensure the speaker that everything is clear and encourage her to carry on. And unlike the signalling systems discussed by Lewis, back-channelling is anything but a marginal phenomenon.

To sum up, Skyrms conventions and Lewis conventions are very different in that the latter are several steps up in complexity from the former. But both are needed to understand social interactions between humans, in general, and their communicative interactions, in particular.

*4. Coordination and cooperation*

Following the general trend in the literature, this chapter is primarily concerned with conventions that serve to solve coordination problems, that is to say, scenarios in which players need to make interdependent choices that will yield an optimal outcome for all. Thus understood, coordination doesn't require that players have a joint goal; it may suffice that all pursue individual goals, taking into account that all others are doing the same thing.

To illustrate the distinction between joint goals and individual goals that are merely shared, compare the following scenarios:

(i) Every morning, A and B walk to the train station for their commute. Since they live in the same area and always take the same train, they sometimes end up walking next to each other. A and B share the same goal, and this may even be common ground between them. Nevertheless, we wouldn't say that they're walking to the train station *together*.

(ii) Now suppose that A and B come to agree that, henceforth, they will take their morning walks together. Now they have a *joint* goal, and that clearly makes a difference.

But what is the difference? According to Gilbert (1990) and many others, it is that in (ii), but not in (i), A and B undertake the *obligation* to pursue a shared goal. Concretely, this might mean that they will wait if the other falls behind, discuss the possibility of taking a different route, notify the

other when they have a day off, and so on. By the same token, A and B are now both entitled to censure the other if these obligations are not fulfilled. For example, if A fails to show up one morning, B is entitled to complain when they meet again. In other words, by undertaking a joint goal, A and B become mutually *committed* to doing what they can to ensure that their goal is achieved. No such commitment is in play if they merely share a goal.

'Cooperation' may be defined as coordinated action aimed at achieving a joint goal. Human communication is generally a cooperative enterprise. As Grice (1975, p. 45) famously put it, we expect each other's utterances to be 'such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.' As a consequence, when making an utterance, we undertake certain commitments, for example, that what we say is true. It is widely agreed that commitment making is a central aspect of human communication (Geurts, 2019; this volume).

It has been argued that the claim that human communication is cooperative is at odds with cases in which speakers' goals are diametrically opposed (e.g., Asher & Lascarides, 2013). A much-discussed example is the following, which is taken from a cross-examination between a prosecutor (P) and a defendant (D):

P: Do you have any bank accounts in Swiss banks?
D: No, sir.
P: Have you ever?
D: The company had an account there for about six months, in Zurich.

The last utterance implies, but does not explicitly state, that the defendant did not have a personal bank account in Switzerland, which turned out to be false. Since the defendant was under oath, he was accused of perjury, though ultimately not convicted (Solan & Tiersma, 2005, p. 213ff.). In cases like this, it is crucial to distinguish between joint and individual goals. Presumably, both the prosecutor and the defendant were trying to stay within the confines of the law; this was a joint goal. But whereas the prosecutor sought to obtain all evidence relevant to the case, the defendant shared that goal only in so far as it didn't conflict with his personal interests. Apparently, the prosecutor missed this crucial proviso, and the defendant was banking on that oversight, but that didn't make him uncooperative in the Gricean sense of the word. Conflicting interests need not prevent us from pursuing joint goals, but whenever there is a conflict of interests, it is paramount to separate between joint goals and individual goals.

The same holds for many other forms of antagonistic interaction. For example, consider a game of chess (Geurts, this volume). Two chess players have diametrically opposite individual goals: each wants to win and the other to lose. This is what makes chess a competitive game. However, at the same time, chess is a form of cooperation, whose joint goal is to produce a winner, following the rules of the game. Most of our competitive activities are like chess in that they require a backdrop of cooperativity and willingness to play by the rules.

In summary, human conventions generally subserve cooperation rather than mere coordination. Relatedly, many if not most human conventions seem to be normative in that they commit agents to certain courses of action, and that these agents may be held accountable if they fail to act accordingly. In these respects, too, human conventions differ from conventions in other species.

## 5. Iconicity and arbitrariness

Among the features commonly attributed to conventions, arbitrariness is one of the most prominent. Conventions of word meaning are a standard case in point. In English, 'cats' and 'dogs' are used to refer to cats and dogs, respectively, but it might as well have been the other way round, and other languages use different words altogether. Though these observations may seem rather trite, they are being questioned by a swell of recent publications which claim that iconicity is a paramount source of non-arbitrariness in language (e.g., Ćwiek et al., 2021; Dingemanse et al., 2015; Sidhu et al., 2021).

Iconicity is a notion whose popularity has been waxing and waning over the ages. After a prolonged period during which it was widely held that the relation between linguistic form and meaning is mostly arbitrary, iconicity has been regaining ground, and its importance is now being asserted with considerable force:

> For decades, theoretical approaches and empirical data on the evolution of spoken languages have been dominated by the dogma of arbitrariness, according to which the forms of words do not resemble their meanings. [...] In line with the idea that arbitrariness prevails in spoken language, iconicity—the resemblance between form and meaning—has been thought to be largely confined to onomatopoeias, such as words like 'bang' and 'peep' which imitate the sounds they

15

denote. In recent years, however, more and more research shows that such iconicity plays important roles in the evolution, acquisition and use of spoken language. (Ćwiek et al., 2021, p. 2)

According to Ćwiek et al., the 'dogma' that the relation between words and their meanings is arbitrary has by now given way to the insight that iconicity is central to language and facilitates its acquisition as well as its evolution. These are strong claims, which echo even stronger claims made elsewhere in the literature; for example, Dingemanse et al. (2015, p. 603) speak of a 'paradigm change' and an 'upheaval [that] is underway in current thinking about the arbitrary nature of linguistic signs.' In our opinion, these claims are greatly overstated.

Although it is rarely discussed in much detail how iconicity could have helped language to evolve, it seems clear that it could have played several, non-exclusive roles: iconicity may have affected the invention of new words, it may have facilitated the interpretation of new words, and it may have made some word/meaning pairs easier to learn than others. Still, the mere fact that iconicity could have played any or all of these roles does not imply that it did.

Iconicity is commonly defined in terms of similarity; as Ćwiek et al. put it, a sign is iconic if its form resembles its meaning. Given that similarity is a relative notion (every thing resembles any other thing in some respects) as well as a graded one, it is inevitable that iconicity has the same features, which entails that iconicity is heavily dependent on the context. That does not necessarily render it useless, as long as these features are firmly borne in mind. Unfortunately, however, the epithet 'iconic' is routinely used as an absolute term, and more often than not, its context-dependence is disregarded.

Ćwiek et al.'s paper illustrates the second point. Extending a line of research that began with Ramachandran and Hubbard (2001), Ćwiek et al. recruited native speakers of 25 languages across the world, and invited each participant to decide which of two shapes corresponded to a spoken nonce word. One of the shapes was rounded, the other spiky; the nonce word was either 'bouba' or 'kiki'. The main finding of Ćwiek et al.'s survey was that native speakers of all but three languages in their sample showed a more or less pronounced tendency to associate 'bouba' with the rounded shape and 'kiki' with the spiky shape.

Intuitively, it is doubtful that the words 'bouba' and 'kiki' are similar to rounded and spiky shapes, respectively, and more generally, it feels wrong to

16

speak of similarities between sounds and objects. Therefore, the 'bouba/kiki' effect does not seem to be a form of iconicity as Ćwiek et al. themselves define it. They acknowledge this point, and then go on to employ the term in a broader sense without clarifying it. Sadly, this is a common practice in the iconicity literature: the standard definition of 'iconicity' is duly cited, and then ignored. Typically, the term 'iconicity' is then alternated with, for instance, 'crossmodal association', 'crossmodal correspondence', 'synaesthetic association', 'sound symbolism', etc. Incidentally, there is evidence that the capacity for having such associations actually relies on the capacity for language, rather than the other way round: great apes and preverbal children are largely insensitive to crossmodal associations as exemplified by the 'bouba/kiki' effect (Fort et al., 2017; Margiotoudi, Allritz, Bohn, & Pulvermüller, 2019; Tomasello & Call, 2018), which casts doubt on the notion that iconicity facilitates language learning. Be that as it may, in order to prevent confusion, we suggest that, in the following, 'iconic' be read as a shorthand for 'exhibiting some statistical correlation between form and meaning'.

Ćwiek et al. interpret their findings as showing that there is 'a strong tendency for people across the globe to associate the spoken word 'bouba' with a round shape and 'kiki' with a spiky one' (p. 10). This interpretation is way too strong. By asking, 'Which shape corresponds to the sound?', Ćwiek et al. effectively *instructed* their participants to accept that the sound they heard 'corresponded' to one and only one of the two shapes. Hence, strictly speaking, Ćwiek et al. observed the effect only in a highly constrained setting, in which an informant was looking at a rounded and a spiky shape, heard a recorded sound, and was then invited to select the shape that 'corresponded' to the sound. Briefly, Ćwiek et al.'s 'bouba/kiki' effect depended massively on the context they created for their informants, and the evolutionary relevance of this context is moot, to say the least.

This point holds more generally: in most experiments that purport to demonstrate iconicity effects, participants are asked to choose between a small number of meanings. In some studies, this is not an issue, but the bottom line remains the same. For instance, Sidhu et al. (2021) found that, in a sample of English nouns, some phonemes are more common in words referring to round objects, while others are more common in words referring to spiky objects. Sidhu et al. represent this finding as 'an instance of iconicity, and thus nonarbitrariness, in human language' (p. 1390), and in their title they go so far as to state that 'sound symbolism shapes the English language.'

However, these are steep exaggerations, for in fact the reported effect was quite weak (the authors call it 'modest', p. 1396). A more sober-minded rendition of Sidhu et al.'s findings is that, in a small corner of the English lexicon, there is a weak correlation between certain phonological features and two features of shape. There is no evidence that this has ever been a major force in shaping the English language, as Sidhu et al.'s title has it.

To return to our main topic, does the wave of recent studies reporting statistical correlations between form and meaning have any bearing on mainstream theories about language conventions, as discussed in this chapter? The short answer is 'no'. As Lewis (1969) puts it, a convention is arbitrary if there is another convention that everyone would have adopted if everyone else did. Even if everyone used the word 'bouba' to refer to ping-pong balls, say, it is evident that all would be willing to switch to using 'kiki' instead if everybody else did. The mere fact that there is a statistical correlation between the form of a word and its meaning does not exclude its being arbitrary in the sense of Lewis, Skyrms, and others. In fact, form-meaning correlations, weak and strong, could easily be incorporated in Skyrms-style models as biases of language users. Doing so might help to assess whether such correlations could have played a substantial part in the evolution of language, but for the time being that issue remains moot, the more so because Skyrms (2010) has shown that signalling conventions can emerge even if the creation of signals is completely random.

## 6. Conclusion

The general picture that emerges from the foregoing discussion is this. Social conventions constrain the ways in which we interact with each other. Languages are systems of social conventions that constrain the sound patterns we emit, their structures, meanings, and so forth. In other species than our own we observe that such behavioural regularities may develop in populations of creatures whose cognitive capacities are quite modest, and Skyrms-style models help to explain how this is possible. In our lineage, however, conventions have grown vastly more complex, as humans steadily became more adept at coordinating their activities, at taking into account each other's goals and beliefs, and at being normative about social conventions in general ('Don't pick your nose!') and language conventions in particular ('Don't split your infinitives!'). Social conventions have deep evolutionary roots, but our

forebears carried them to unprecedented levels of complexity. The evolution of language was part and parcel of that process.

*References*

Arbib, M. A., Liebal, K., & Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. *Current Anthropology*, *49*, 1053–1063.

Asher, N., & Lascarides, A. (2013). Strategic conversation. *Semantics and Pragmatics*, *6*, 1–62.

Bartsch, R. (1987). *Norms of language: Theoretical and practical aspects.* Longman.

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of personality and social psychology*, *79*, 941-952.

Bicchieri, C., Muldoon, R., & Sontuoso, A. (2018). Social norms. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2018 ed.). Metaphysics Research Lab, Stanford University.

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., . . . Winter, B. (2021). The *bouba/kiki* effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1841).

Davis, W. A. (2003). *Meaning, expression, and thought.* Cambridge University Press.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, *19*, 603–615.

Fort, M., Lammertink, I., Peperkamp, S., Guevara-Rukoz, A., Fikkert, P., & Tsuji, S. (2017). Symbouki: A meta-analysis on the emergence of sound symbolism in early language acquisition. *Developmental Science*, *21*, e12659.

Geurts, B. (2018). Convention and common ground. *Mind and Language*, *33*, 115–129.

Geurts, B. (2019). Communication as commitment sharing: Speech acts, implicatures, common ground. *Theoretical Linguistics*, *45*, 1–30.

Gilbert, M. (1989). *On social facts.* Routledge.

Gilbert, M. (1990). Walking together: A paradigmatic social phenomenon. *Midwest Studies in Philosophy*, *15*, 1–14.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Speech acts. Syntax and semantics 3* (pp. 41–58). Academic Press.

Hofbauer, J., & Huttegger, S. M. (2008). Feasibility of communication in binary signaling games. *Journal of Theoretical Biology, 254*, 843–849.

Huttegger, S. M., Skyrms, B., Smead, R., & Zollman, K. J. (2010). Evolutionary dynamics of Lewis signaling games: Signaling systems vs. partial pooling. *Synthese, 172*, 177–191.

Irvine, E. (2016). Method and evidence: Gesture and iconicity in the evolution of language. *Mind and language, 31*, 221-247.

Lewis, D. (1969). *Convention.* Harvard University Press.

Lewis, D. (1975). Languages and language. In K. Gunderson (Ed.), *Minnesota studies in the philosophy of science* (pp. 3–35). University of Minnesota Press.

Margiotoudi, K., Allritz, M., Bohn, M., & Pulvermüller, F. (2019). Sound symbolic congruency detection in humans but not in great apes. *Scientific Reports, 9*, 12705.

Maynard-Smith, J., & Price, G. R. (1973). The logic of animal conflict. *Nature, 246*, 15–18.

Millikan, R. G. (1998). Language conventions made simple. *Journal of Philosophy, 95*, 161–180.

Moore, R. (2013). Imitation and conventional communication. *Biology and Philosophy, 28*, 481–500.

Planer, R. J., & Sterelny, K. (2021). *From signal to symbol: the evolution of language.* MIT Press.

Price, T., Wadewitz, P., Cheney, D., Seyfarth, R., Hammerschmidt, K., & Fischer, J. (2015). Vervets revisited: A quantitative analysis of alarm call structure and context specificity. *Scientific Reports, 5*, 13220.

Ramachandran, V., & Hubbard, E. (2001). Synaesthesia: A window into perception, thought and language. *Journal of Consciousness Studies, 8*, 3–34.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science, 210*, 801–803.

Sidhu, D., Westbury, C., Hollis, G., & Pexman, P. (2021). Sound symbolism shapes the English language: The maluma/takete effect in English nouns. *Psychonomic Bulletin and Review, 28*, 1390-1398.

Skyrms, B. (1996). *Evolution of the social contract.* Cambridge University Press.

Skyrms, B. (2010). *Signals: Evolution, learning, and information.* Oxford University Press.

Solan, L., & Tiersma, P. (2005). *Speaking of crime: The language of criminal justice.* University of Chicago Press.

Sugden, R. (2005). *The economics of rights, co-operations, and welfare* (2nd ed.). Palgrave.

Tomasello, M., & Call, J. (2018). Thirty years of great ape gestures. *Animal Cognition*, *22*, 461–469.

Wagner, E. (2009). Communication and structured correlation. *Erkenntnis*, *71*, 377–393.

Yngve, V. (1970). On getting a word in edgewise. In *Papers from the sixth regional meeting of the Chicago Linguistic Society* (pp. 567–577).