David Lewis on Convention

Ernie Lepore and Matthew Stone
Center for Cognitive Science
Rutgers University

David Lewis's landmark *Convention* starts its exploration of the notion of a convention with a brilliant insight: we need a distinctive social competence to solve coordination problems. Convention, for Lewis, is the canonical form that this social competence takes when it is grounded in agents' knowledge and experience of one another's self-consciously flexible behavior. Lewis meant for his theory to describe a wide range of cultural devices we use to act together effectively; but he was particularly concerned in applying this notion to make sense of our knowledge of meaning. In this chapter, we give an overview of Lewis's theory of convention, and explore its implications for linguistic theory, and especially for problems at the interface of the semantics and pragmatics of natural language. In §1, we discuss Lewis's understanding of coordination problems, emphasizing how coordination allows for a uniform characterization of practical activity and of signaling in communication. In §2, we introduce Lewis's account of convention and show how he uses it to make sense of the idea that a linguistic expression can come to be associated with its meaning by a convention. Lewis's account has come in for a lot of criticism, and we close in §3 by addressing some of the key difficulties in thinking of meaning as conventional in Lewis's sense. The critical literature on Lewis's account of convention is much wider than we can fully survey in this chapter, and so we recommend for a discussion of convention as a more general phenomenon Rescorla (2011).

§1 Coordination in the Social World

Lewis regarded *coordination problems* as providing a distinctive insight into the social world. Coordination problems name a subset of strategic situations where agents must choose actions that agree with one another to achieve mutually acceptable outcomes. In coordination problems, agents have to make matching choices from among multiple candidate strategies, but there's no intrinsic reason to prefer one over another. So, agents must act solely based on their expectations about one another. Coordination problems, we shall see, require a distinctive kind of reasoning and knowledge. Uniquely, in coordination problems, agents' expectations about one another are both decisive in settling what they should do and independent of the practical logic of their situation. The phenomenon of coordination thus brings powerful constraints to anchor our thinking about social competence.

The examples that Lewis gives show the scope of this phenomenon. One example is holding a meeting. If we're arranging as a group where to hold a

meeting, we could hold it at your place, in Princeton perhaps. Or we could hold it at our place, in New Brunswick. Considering our joint effort, it's equally efficient for us to meet in either place. We have multiple options, and it's up to us what we do. Practical logic doesn't dictate where we go. Nevertheless, what we do has to match. If half of us end up in Princeton and half in New Brunswick, we won't have a meeting at all. You share with us our interest in going to the same place. So, you should go where you expect we will be waiting to meet you; and we will do likewise. Our expectations about one another are decisive.

Once one recognizes coordination as a phenomenon, examples proliferate. A vivid and dramatic case is the choice whether to drive on the right side of the road, as Americans and Continental Europeans do, or whether to drive on the left side of the road, as they do in the UK and Japan. In choosing to go right, we are acting on our expectations of how others will choose to go. We think they will also go right. If we thought they would all go left, then we would have gone left as well; we all wish to avoid congestion or collision. The coordinated expectations that send us all to the right—as in the United States—do not make reference to anything in the world that settles preferences or strategies for us. Going left is just as good. Either strategy serves its participants equally well, just so long as all of the participants agree on the same solution.

Lewis also considers resuming a dropped phone call between two parties. It's in their interest to pick up the conversation and continue where they left off. But if both dial immediately, the connection fails. Exactly one party must attempt to initiate it. You might have a variety of ways of stating a rule or policy for settling who does what. According to Lewis, the operative one in Oberlin, back when he was growing up, was that the person who dialed the original call was the one who must redial when the call is dropped. The person who received the original call must wait. We have acquaintances who still follow this strategy today, in the days of cell phones, where even the receiver has a button to redial the last incoming caller. The case illustrates a point Lewis emphasizes: to solve coordination problems, agents must do their part; it may not be enough for them simply to act the same way.

The point extends to coordination among multiple agents. Lewis considers a group that must carry out a search together. They maximize their success when each person in the party stakes out a different area to search. When everyone searches the same area, territory remains unexplored. Their search may fail. So, again, everybody has to choose his action depending on what everybody else chooses.

It may be that the coordination activity is something hard to describe, implicit or continuous. Lewis uses the example of two people in a rowboat from Hume (1740: §3.2.2). The two people have to row in concert in time for their strokes to be effective and the motion of the boat smooth. But it doesn't particularly matter how fast they row within certain limits; they can adjust the force they use so that

it's a reasonable pace no matter how fast their interval is. Normally, they can get into sync without ever making an explicit agreement or even discussing their pace, simply by making small adjustments to one another. In the end, they wind up not only matching their motions, but also matching their expectations of one another. Each chooses to row at a certain speed because he expects the other to do so.

Lewis develops a formal approach to coordination and communication by drawing on game theoretic ideas from von Neumann & Morgenstern (1947) and Schelling (1960). We review his formalism briefly here; we recommend Clark (2011) for a gentle introduction to game-theoretic analysis of coordination and communication; a broader technical overview of the subject from a philosophical perspective can be found in Skyrms (2010).

A game is a formal way of describing a situation in which multiple agents interact, make choices and achieve outcomes. Games generalize the decision-theoretic models that characterize the choices of individual agents acting alone. In decision-theoretic models, agents make choices by anticipating the expected outcome of their actions in virtue of the laws of nature. In games, multiple agents must make choices, perhaps without being able to observe one another's actions. To anticipate the outcomes of actions in games, agents must reason not only about the natural course of events but also about the choices other agents will make. This gives an inherently strategic character to ideal reasoning in game-theoretic models.

The simplest formalization of games is simply to map out the choices that are available to each agent and the payoffs that result for each combination of actions on the part of the players. This is known as the *extensive form* of a game. The game in Diagram (1) typifies the situations that Lewis associates with coordination.

	c_1	c_2
r_1	$1,1$	$0,0$
r_2	$0,0$	$1,1$

(1)

It describes an interaction between two agents, the *row player* and the *column player*. They each select one of two possible actions; then they get a payoff. Coordination is required here because the row player and the column player both get a good payoff if they choose their first action, and they both get a good payoff if they choose their second action. But they get bad outcomes if their choices don't match. Their preferences are aligned in this case, and their preferences dictate that they make matching choices, but their preferences don't dictate which choice they will make or must make or should make. Either matching strategy is just as good.

What this shows is that the intuitive notion of coordination that Lewis has identified in his vignettes from everyday life has a mathematical basis. Formalizing coordination in terms of games highlights what's distinctively social about coordination. The right action is inherently underdetermined by agents' interests and the practical logic of the situation—that is, the way actions lead to outcomes. Agents' expectations about one another are decisive and, in a certain sense, arbitrary.

We can make this more precise by appealing to the natural game-theoretic account of rational choice. Consider an agent in a game. If she chooses her actions rationally, based on their expected outcomes, she needs to consider her expectations about the other agents' actions, because those actions are going to determine her payoffs in part. If the row player knows the column player will do c_1, she ought to do r_1. If she knows that the column player will do c_2, she ought to do r_2. But the same is true of the column player. If he knows that the row player will do r_1, he ought to do c_1. If he knows that the row player will do r_2, he ought to do c_2. Thus, a good way to think about a satisfactory outcome in a game like this—what is it that agents should aim for, and what is it that their behavior should achieve—is as a *Nash equilibrium* (Nash 1950). A Nash equilibrium is a set of strategies, chosen simultaneously by all the players, in such a way that each player achieves the best outcome possible *given* the strategies that all the other players are following. As we have seen, in (1), both c_1–r_1 and c_2–r_2 are equilibria. The formalism captures the intuitive arbitrariness of agents' behavior by the fact that the game is associated with these two different equlibria.

Lewis defines coordination problems as situations of inter-dependent decision by at least two agents where coincidence of interest predominates and where there are at least two equilibria. In these cases, agents that are coordinating successfully must have additional knowledge of one another—some social knowledge or competence—that allows them to find the particular equilibrium they use as opposed to the others.

For Lewis, *communication* involves particular kinds of coordination problems. The relationship between communication and coordination is clearest in simple idealized situations known as sender–receiver games. In these situations, one person produces a *signal* knowing privately some aspect of the state of the world. The other person, the recipient, acts having seen the signal. What we are calling signals in these games are just arbitrary actions with no real-world effects that agents intrinsically prefer. Signals get their power through the value of information, as the receiver makes better decisions by recognizing the strategy the sender uses in presenting the signal they do.

The story of Paul Revere's Ride, as portrayed in Longfellow's historically-embellished 1860 poem, provides the canonical example. Paul Revere and the

sexton of Boston's Old North Church must coordinate the militia's defenses against British troops. Revere must determine where the defenses will be mounted, but only the sexton has the key information about how the attack will proceed. The sexton hangs a pattern of lanterns in the church belfry, which can be seen from far away, as a warning signal. He is to use one light if the British are coming by land, and two lights if by sea. In themselves, the lanterns are little more than decorative. But because of the sexton's plan, the pattern of lights is contingent on what the British are doing, and Revere knows this. So, Revere orchestrates his pattern of defense contingent on the lights he sees. If he sees one lantern hanging in the belfry, he prepares against a land attack; if he sees two, he prepares against a sea attack. The joint strategy of Revere and the sexton allows them to respond effectively no matter how the attack comes. In fact, it is another equilibrium. The sexton has no reason to change his strategy, given what Revere will do, and Revere has no reason to change his strategy, given what the sexton will do.

We can schematize key elements of the Revere story in a formal game. The formalism highlights key features of Lewis's understanding of signaling. For Lewis, coordination is fundamental to signaling. And, for Lewis, what we call the meanings of signals is just agents' expectations about when signals are to be used and what responses they will elicit. Say, the sexton has two lanterns, and will hang them differently depending on his knowledge of the British plans. That makes for two possibilities: one if by land, two if by sea (L_1S_2); and one if by sea, two if by land (S_1L_2). Let's assume that Revere, too, will act differently depending on what he sees: that makes for two possibilities: land defense for one light, sea for two (L_1S_2); and sea defense for one light, land for two (S_1L_2). Here are the payoffs in Diagram (2).

	L_1S_2	S_1L_2
L_1S_2	1, 1	0, 0
S_1L_2	0, 0	1, 1

(2)

So, Revere and the sexton reach an equilibrium when both play L_1S_2. And they also reach one when both play S_1L_2. Either works, but they must coordinate. It is their mutual expectations that are decisive in what they actually do. Mathematically, we see the commonality with straightforward coordination problems, like what side of the road to drive on. The fact that the sexton's actions here have no intrinsic effects gives an added nuance to the case. What Revere and the sexton's mutual expectations amount to here is an *interpretation* for the actions of the sexton as signals of what the British are doing.

Although this abstraction allows Lewis to characterize the *meanings* of signals directly in terms of agents' information, choices, and outcomes, the characterization of meaning that results is very coarse. A signal that indicates that a proposition *P* is true, and that therefore prompts the recipient to choose an action *A*, means just as much *P* as it does "Do *A*". In the Revere story, hanging a single lantern in the belfry could mean the British are coming by land, but it could also be an instruction to prepare the land defenses, or it could be both. All the theory allows us to do is to characterize the probabilistic inference that comes from combining the expectation the sexton will act according to a certain conditional policy and the observation of a specific instance of that policy. The information carried by signals is even further removed from their intuitive meanings in cases where agents' interests do not entirely coincide, so agents have interests in hiding information as well as sharing it, may send unreliable signals and may interpret signals as unreliable. (This has proved problematic not only for human communication but also for understanding the meaning of animal signaling; see Lachmann, Számadó & Bergstrom 2001 for example.) We'll return to this weakness of Lewis's approach in §3.

§2 Convention

Coordination is a powerful concept for theorizing about social competence. Suppose we observe agents reliably solving a particular coordination problem according to a particular joint strategy. Then they *must* share knowledge or skills that let them choose this pattern over its alternatives. Here's *the* place where agents' knowledge of one another and alignment with one another seems to be doing special work in their abilities to interact—over and above what they know about rationality and over and above what they know about how the world works.

In principle, the mechanisms behind successful coordination could take many forms. Coordination could be the result of agents' innate architecture. This seems likely to be the case, for example, for the famous waggle dance of honeybees (von Frisch 1967). Bees signal the distance and direction of sources of nectar by particular patterns of movement. Other bees can interpret these movements, so they can retrieve the nectar in turn. These movements, like the sexton's signal in the story of Paul Revere, have neither intrinsic meaning nor

substantive practical effects. They get their meaning from the strategies of the bees, and in particular from the match between the circumstances in which bees that have found nectar produce the signals and the behaviors with which bees in the audience respond.

If bees had evolved differently, the waggle dance might very well have involved a different set of joint strategies. Indeed, there is some evidence that different species of bees do produce and interpret their waggle dances differently (Su, Cai, Si, Zhang, Tautz & Chen 2008). So, the formalism of coordination provides a useful construct for characterizing the social competence of the bees. On the other hand, we needn't suppose that individual bees are rational agents solving this coordination problem on the basis of expectations about one another. Individual bees may have little alternative but to act as they do. Their social competence may be hardwired, or perhaps may allow for only very limited kinds of learning or modulation from interactions with other bees.

Indeed, when biologists approach the evolution of social behavior, they often start from the modeling assumption that agents' choices are simply dictated by their genes. Each individual acts not based on rational calculations but based on a predefined set of rules. However, agents act repeatedly; they have lots of chances to interact with one another across lots of situations. Agents that coordinate often in these interactions are more successful, and thus, are more likely to propagate their patterns of choices to their descendants. Under these assumptions, populations will converge to fixed points where the distribution of strategies is under no pressure to change. Those fixed points turn out to be a subset of the familiar Nash equilibria, so-called *evolutionarily stable strategies* (Smith & Price 1973). The analysis confirms the relevance of the game-theoretic analysis for the theory of social competence, even if agents' individual decisions are heuristic or even fixed in advance.

Here's a somewhat different kind of case. Even if agents have choices to make, reliable coordination could simply be the result of psychological mechanisms that make particular actions salient. In this case, agents don't need to have specific knowledge or expectations about one another to make their choices. They simply do what seems natural to them. Schelling (1960) showed that in many decision problems, one course of action does seem the simplest or the most natural. It naturally springs to mind or attracts the attention. He called such strategies *focal points*. In many cases, even in novel coordination problems with strangers, people can identify aligned focal points, and so work out how to coordinate successfully much more frequently than one would expect by chance.

However, not all cases of coordination can be explained by such mechanisms of architecture and alignment. Coordination is sometimes a matter of learned expectations about agents' strategies. These cases, for Lewis, are cases of *convention*. A convention, for Lewis, involves the

deliberative choice of one regularity *R* over another regularity *R'* in a recurrent situation *S*, in virtue of expectations shared among a population *P*. In the ideal case (which Lewis allows to be satisfied only to a limited degree), a regularity *R* in behavior is a convention for *P* if it's common knowledge among the members of *P* that they follow *R*, that they expect one another to follow *R*, that *R* is a solution to the coordination problem they face in *S*, and that there is another solution *R'* that they could have conformed to instead of *R* (Lewis 1969: 76).

Note that Lewis's account of conventions does not seem to fit cases of alignment based on hardwired strategies or innate focal points. In these cases the agents' *preferences* in situation *S* might be satisfied just as well by a regularity other than the one they exhibit; the agents might even know this. But it wouldn't be *possible* for them to conform to any other regularity than the one they actually exhibit. Only if agents have somehow learned or negotiated strategy *R* does it seem they really could have exhibited some other strategy *R'*.

Drawing on this framework, Lewis proposes that meaning in language is a matter of convention. His understanding generalizes the simple signaling conventions of the Revere story, where one agent produces an action conventionally under certain circumstances, and other agents respond to that action conventionally in specified ways. For Lewis, those who make utterances, conventionally do so only under particular circumstances: namely, when those utterances are true. Conversely, those who interpret utterances, conventionally act as though the world is a certain way: namely, as though the utterances are true. Such matching strategies on the part of speakers and their audiences constitute equilibria, Lewis suggests. The group prefers the outcomes that follow from effective communication. And there are many such equilibria, as the diversity of human languages attests.

Lewis's definition allows that conventions could be instituted explicitly. Agents might agree to defer to an authority that establishes how they should act together, on the model of the *Académie Française*. Agents might continue to defer to conventions that date from a particular agreement they regard as binding, like the Geneva Conventions. But Russell (1921) and Quine in the foreword to Lewis (1969) observed that any conventions of language could not all have such explicit origins, on pain of regress. Consider making the first such agreement. The parties would need a language to carry out their negotiations. But since this is the first agreement there could not be any conventions of language yet.

However, Lewis's notion of convention allows for conventions that evolve gradually and tacitly as a side effect of agents' independent need to make their way in the social world. The first time we face a coordination problem, we have no choice but to do something. So, we do whatever occurs to us. But

salience gives us a certain probability of success. It may be that what occurs to us matches what others do, because of our natural affinities. Successful coordination creates a precedent that we can rely on in future interactions. Eventually we have enough confidence from our successful experience coordinating that we have developed the mutual expectations that constitute a convention. As long as agents are able to coordinate in particular kinds of ways, then the conventions are implicit in the choices they make and the knowledge they have of one another. Because Lewis's definition does not call for explicit agreement in cases of convention, his definition escapes the regress objection.

§3 Conventions of Meaning in Critical Perspective

Lewis's understanding of convention provides a beautiful way of relating social competence to underlying experience, knowledge and rationality. Lewis's theory of the conventions of language is a brilliant application of the ideas in an attempt to reduce knowledge of language in general, and knowledge of meaning in particular, to our expectations about one another. Moreover, Lewis's account has much in common with H.P. Grice's general view of meaning as a product of interlocutors' collaborative intentions (Grice 1957). To coordinate by convention entails recognizing the intentions of other agents. In particular, if meaning amounts to some conventional regularity R, as Lewis suggests, then in each utterance, the speaker A intends what he does to exemplify R, and his audience B knows this. Moreover, A and B both expect everyone to do their part in R, and do their part conditional on the others doing so. By Lewis's definition of convention, if B thought A intended to participate in R' and expected B to do so as well, B would participate in R' instead. Another way to say this is that if B recognized A's intention differently, B would not make the response that B makes to A's utterance. In this sense, recognizing A's intention is why B does what B does. Lewis's account of convention thus dovetails with Grice's analysis of meaning.

Despite these attractive features of Lewis's appeal to convention, the understanding of language that emerges from it is problematic in many respects. In the next few subsections, we will explore some of these problems.

Improvised Meaning

Lewis is working, at least implicitly, with a view of language as relatively stable and in common currency. On this view, semantics reveals widespread regularities that hold across people, space and time. The meanings of words might be a prototypical example. Because such regularities are widespread and stable, there won't be anything tendentious about thinking about language use in terms of recurring coordination problems and conventional solutions to them.

This way of thinking about language is harder to maintain if improvisation is an essential part of meaning in language. When we coin new words, modulate established meanings, and arrive at creative understandings of one another, we do not rely entirely on conventional semantic properties. Indeed, we may not even take up any expectations that we are going to continue to use words in these new ways. If we want to characterize meanings in such cases, it seems unlikely that we can do so just with shared solutions to recurrent coordination problems.

Indeed, even for Lewis, conventions are established gradually. At first, coordination succeeds through other mechanisms, like salience, good luck, or partial or tentative precedents. In these cases, speakers don't yet have common knowledge of an equilibrium that would settle the meanings of their words. We cannot appeal to convention to characterize what their meanings are in these cases.

Indeed, the more you think about it, the more problematic it seems to try to *reduce* knowledge of meaning to knowledge of other agents' strategies for using language, as Lewis seems inclined to do. This becomes particularly clear if we consider the improvised meaning that Grice was most concerned with, so-called pragmatic 'conversational implicatures' (Grice 1989). We can separate semantics from other aspects of utterance interpretation only if we can distinguish what an utterance means from everything that the audience can infer from the fact that the speaker has said it. But Lewis is out to explain meaning in terms of our expectations about one another. For Lewis, utterances get their meanings from the prevailing mutual expectations about what everyone is going to do. In the signaling games which capture the essence of Lewis's idea, what an agent learns when she receives a signal is exactly that the world is the type of place and the speaker is the kind of person who would send that signal now. That's never going to distinguish semantics from pragmatics in an intuitive way.

Take a concrete case. We want to assign different meanings to "There's a bug on your back!" and "Swat the bug!" But each would be uttered when there's a bug on the hearer's back, and each calls for the hearer to swat that bug. So, each has an equal claim to mean, there's a bug, and each has an equal claim to mean, swat it!

Lewis recognizes this problem, and has a particular project for how to deal with it. He thinks that his conventions of truthfulness and trust in natural language aren't subject to exactly these criticisms. The conventions for meaning in natural language, he argues, are productive and systematic in ways that resolve much of the indeterminacy we see in the Paul Revere case. In other words, "There's a bug on your back!" and "Swat the bug!" may be used in the same circumstances, but we will ultimately have to give them

different conventional meanings, because we will ultimately have to have a compositional semantics, and compositionality will distinguish them.

Compositionality certainly places profound constraints on semantics. But if much of meaning is in fact improvised, then we will always need to describe institutions that let us make meaning, not just institutions that let us use established meanings. Even if we keep Lewis's central place for coordination and social competence in delimiting semantics, we will need some additional tools. And this broader account is likely to fundamentally shape the way we see the boundary between semantics and pragmatics. We can describe the difference between "There's a bug on your back!" and "Swat the bug!" *not* in terms of general expectations of when speakers use utterances and how their audiences react, but in terms of the specific dynamics and institutions of meaning that make one an utterance about how things are and the other an utterance about what the audience should do.

Universal Grammar

Lewis is also working, at least implicitly, with a view of language as entirely learned. This shows up in the very strong sense in which conventions are arbitrary in Lewis's account. When we have a convention to solve a coordination problem with a certain strategy, we've learned that people around us follow that strategy—or we've worked to establish it as an equilibrium with the people around us. Moreover, if we have a convention, then we could have arrived at a different joint strategy had our experiences and choices been different. This also shows up in the strong assumptions that Lewis makes about the mutual knowledge and rationality associated with linguistic conventions. Because Lewis views linguistic conventions, at least implicitly, as learned manifestations of people's general social competence, Lewis says that a convention is established only when agents have common knowledge of one another's preferences, one another's rationality, one another's likely choices, and hence the convention itself. These are very strong assumptions, indeed; but they are necessary to explain why the convention gives agents general and decisive reasons to choose the actions they do.

We've seen that this is not the only reason why one might observe a particular equilibrium in a coordination game. In particular, in some circumstances, we expect equilibria to arise through *arbitrary psychological mechanisms*, rather than through arbitrary experience. Linguists in the Chomskyan tradition take precisely this sort of approach to linguistic communication. On this view, significant aspects of linguistic structure and meaning are universal across human languages. They are determined by the language faculty, the underlying cognitive module that enables people to learn and use their native language. The operation of the language faculty helps to explain how children are able to acquire their native language so quickly, simply by participating in the highly limited linguistic interactions of early life. What children need to learn is just

relatively superficial particulars about the language spoken in their community. See Pesetsky (1999) for a brief overview or Baker (2002) for a more detailed exposition of this perspective.

Linguists' understanding of the language faculty fits poorly with the assumption that linguistic meaning is conventional in Lewis's sense. Speakers' linguistic choices are not always the product of their deliberative rationality; they are often automatic consequences of the operation of language faculty.

If speakers' choices match, it may be because the same mechanisms apply across individuals, not because speakers make their choices by taking learned expectations about one another into account. In this case, speakers could act in accordance with the rules of language without even *knowing* what those rules are—much less having common knowledge of the rules. In particular, we might question whether speakers need have any beliefs, even tacitly, about how their own language faculty operates; the language faculty simply does its thing. But suppose we do acknowledge an attitude that speakers bear to the information that tacitly characterizes the language faculty. If that information simply reflects the innate architecture of the system, we have little reason to characterize speakers' attitudes epistemically, in terms of justified conclusions informed by relevant evidence. See for example Chomsky (1980) or Devitt (2006).

Note, however, that this understanding of the language faculty may still make crucial appeal to the more general notion of social competence involved in successful coordination. Children must acquire the specific facts that distinguish their native language from other possible languages. These facts are not determined just by the practical interests of members of the community, but crucially depend on the alignment that underwrites coordination. This fundamental arbitrariness constrains the evidence that children should use and the hypotheses that they should entertain in coming to understand the language spoken around them. The reasoning involved may well be different from what children need to understand actions that accomplish practical goals by more straightforward means.

The Chomskyan view provides an interesting foil to Quine's objections about language and convention. If we say, with Chomsky, that some rules of language are a manifestation of an innate social competence, there can be no fear of regress in how these rules arise. Our innate social competence does not have to be agreed on in words; indeed, there is no way it could be. Our innate social competence simply evolves. As new features are introduced into the underlying architecture, they trigger corresponding universals in the linguistic capacities that the architecture manifests.

We needn't go as far as the Chomskyan picture to worry about applying a notion of convention as strong as Lewis's to linguistic knowledge. Burge

(1975) offers a simple case. Consider an isolated population, whose members mistakenly subscribe to the view that their language is the only possible language. This no longer satisfies Lewis's definition of a convention. To satisfy that definition, the group has to have mutual knowledge that there's another alternative equilibrium that they could have adopted. This is exactly what Burge's isolated population denies.

This discussion again suggests Lewis is asking too much in reducing meaning to convention. His notion of coordination provides a useful handle on what's distinctively social about language, particularly meaning in language. When agents face coordination problems, they can't succeed just by practical rationality, because practical rationality doesn't dictate which of the many equilibrium strategies is the right one. So, agents that do coordinate successfully must have additional processes at play—innate, heuristic or deliberative—that allow them to align their choices to one another. Speakers of language do coordinate. Explaining meaning in language thus means getting clear on the processes that the agents must have—innate, heuristic or deliberative—that lets them match their signals and their interpretations.

Conventions, as Lewis considers them, are one possible explanation. But it seems likely that there are crucial ingredients of meaning in language that can't be captured this way. The problem crosscuts the problem of improvised meaning that we considered earlier. Improvised meanings are not conventional because they are not yet established; innate meanings are not conventional because they are not established through the right mechanisms. We draw a similar moral about Lewis's approach from the two cases, however. We believe (and argue as much in our forthcoming book (Lepore and Stone, 2014) that once we say more about the kind of coordination involved in semantics, we will feel less pressure to go further still, and demand that this coordination necessarily be based in convention.

Other Perspectives on Convention and Meaning

Lewis's discussion of coordination and convention has sparked a wide-ranging philosophical inquiry into how far rationality, experience and mutual expectations actually go in explaining what people do together. For example, philosophers have pointed out social customs of various kinds that do not fit Lewis's notion of convention. People often seem to stick to customary patterns of behavior, but more out of habit than out of preferences that actively require conformity (Gilbert 1989, Millikan 1998). At home, the family table normally accommodates a few guests. Still, when the family sits down to eat, everyone may routinely pick the same customary place to sit and leave the same seats empty. It's open to everyone to mix things up, and it might be interesting or even preferable to do so from time to time. So when everyone persists in doing what they usually do, it's not a convention in Lewis's sense. It's possible that uses of language could likewise be routine but not conventional, in theoretically significant ways. For example, you might

think this way about sticking to the familiar formulation in indirect speech acts like “Can you pass the salt?”

Philosophers have also argued that some customs continue to figure in the cultural repertoire despite the fact that people act in accordance with them only rarely. An example is handing out cigars at the birth of a baby. This is not, let us suppose, what most new parents actually do nowadays. So, there are no prevailing mutual expectations that new parents will celebrate by handing out cigars. And parents’ choices to hand out cigars, if they do so, cannot respond to the expectations of the community that they will. Handing out cigars is not a convention, in Lewis’s sense. But it is still a customary option, whose cultural meaning is shaped by the weight of precedent. We might want to accommodate such options, and explain their relationship to other kinds of customary behavior, in a broader theory of the social world. The category might prove just as useful in accounting for elements of language, such as rarely used verbal clichés..

A different sort of worry has to do with Lewis’s requirement that the interests of parties to conventions coincide. There seem to be exceptions, including many of the cases Hume was originally interested in, such as accounting for the institution of property as a convention (Vanderschraaf 1998a). Considering more general situations exposes two weaknesses in Lewis’s account of conventions of meaning. The first has to do with Lewis’s characterization of the mutual expectations involved in convention. In general games, as in Lewis’s coordination games, there can be multiple alternative Nash equilibria. Again, in general games, when an agent chooses the strategy involved in one Nash equilibrium R rather than another R’, the agent’s expectations about what others will do are decisive. The agent prefers to participate in R rather than R’ given that others do. However, in general games, it may no longer be the case that the agent playing R prefers that others play R as well. Perhaps, the agent would be better off if others capitulated and simply abandoned their own interests. (Of course, that won’t happen.) A more general account of convention must therefore adjust Lewis’s definition not only to drop the requirement that the interests of the parties to the convention coincide, but also to characterize agents’ mutual expectations in appropriate game-theoretic terms. See Sugden (2004) or Vanderschraaf (1998b).

More specific problems arise with signaling in the presence of conflict. If signals are cheap to produce and the interests of the sender may differ from those of the recipient, then the equilibrium strategy of the receiver is often not to trust the signal, but simply to ignore it. The recipient may have good reason to expect that she would get the same message from the sender no matter what the world was like. In one-shot games, adversarial communication seems to depend on the use of costly signals, which carry penalties that outweigh the benefits deceptive senders could possibly get from using them. Recipients can trust costly signals because no sender has

an interest in using them unreliably. Seminal work on costly signals includes Spence (1973) in economics and Grafen (1990) in biology.

Utterances do not seem to be costly signals, so we would require a different tack to link meaningfulness to the strategies of language users with conflicting interests. One way to salvage meaningful cheap signals is to consider the interactions of senders and receivers who must talk and act repeatedly over time. When agents interact indefinitely with one another, a wide range of equilibria are possible. This is known as the *general feasibility theorem* or *folk theorem* in economics. These equilibria work because repeated games allow for the threat of punishment. Agents stick to the expected course of action because they know that if they deviate in any one round to pursue their short-term self-interest, other agents will retaliate in subsequent rounds and impose long-term penalties that outweigh their short-term gains.

Suppose that conflicting interests in society make this the right way to think about communication. Then not only would we need to adjust Lewis's definition of convention, perhaps in line with Sugden (2004) or Vanderschraaf (1998b), we'd also need to adjust our understanding of what the conventions that underwrite meaning actually amount to. They'd be not just conventions of truthfulness and trust, following Lewis, but conventions of truthfulness, trust, and punishing liars.

People certainly do retaliate against liars. But lies are particularly problematic only if, like Lewis, we attempt to characterize meaning exclusively through interlocutors' expectations about when speakers will use utterances and how addressees will respond to them—that is, directly in psychological terms. If we think more abstractly in terms of activities of making ideas public or reaching agreements about how things are, we don't put ourselves in the position of having to account for the content of a lie in terms of the circumstances in which a liar uses it, or of explaining the joint interest that interlocutors have in a productive conversation in terms of interests they share in practical outcomes. So, while we acknowledge that conflict has an important place in a broader account of conventions and social competence, we think that this alone does not prevent us from linking meaning to coordination, following Lewis.

Finally, social competence seems to play a role not just in explaining the strategies that we follow when we undertake joint activities but in explaining how many of those activities are constituted in the first place. Think of a game like chess. Players have to coordinate when they play chess. The satisfaction they derive from the game comes in part from the shared standards they must apply to interpret the moves and track the flow of play. Moreover, these rules seem arbitrary in many respects. We could agree to change the rules in a wide variety of ways. We would get different games, perhaps all quite entertaining, with varying degrees of similarity to official

chess. In this sense, when players agree to play a game of chess, they implicitly adopt the convention of interpreting their moves according to the rules of chess. So, Lewis's account of social competence is an important part of explaining how chess works.

However, the example of chess also shows the limits of Lewis's model of the origins of conventions (Marmor 2009). Lewis's model seems suited to cases like driving where agents face a coordination problem that exists antecedently and must converge on a convention for solving the problem. They muddle through, eventually relying on the weight of salience and precedent to secure reliable mutual expectations about what to do.

Imagine trying this with chess. Players start playing a game. They improvise their moves and their meanings, sometimes succeeding in arriving at a common understanding of the state of play through the weight of salience and precedent. Eventually, they have a shared understanding that they are playing chess. Explaining the conventions this way seems to miss the point: the problem of playing chess is itself the product of agents' interaction. Players need to converge on the rules of the game, not just on their strategies for playing. They can do this only if they have the right antecedent relationships and institutions, so that they can flesh out, refine and litigate their inchoate rules to cover fairly the cases that arise as they play. Similarly, to the extent that interlocutors are engaged not just in using existing meanings but in establishing new meanings, they too will require not only conventions governing existing meanings but also the appropriate relationships and institutions for pursuing new meanings.

We have already noted, however, that meaning can be improvised. Improvised meaning is coordinated, but not conventional. Universals of grammar are not conventions either. Focusing on the social competence needed for communication seems like a better way to update Lewis's insights than to respond to a litany of putative counterexamples to Lewis's account of conventions in general and conventions of meaning in particular. Lewis's insight linking coordination and communication remains central to this project. But the dialectic is much changed from Lewis's original one. Gone is the goal of characterizing meaning exclusively in terms of agents' expectations about one another's strategies. Gone is the worry that such expectations might be impossible to establish without a language to express them in. Instead, the question is how we can characterize the enterprise of communication, to naturally highlight the coordinated strategies interlocutors are following in producing and understanding utterances, and the diverse linguistic resources that are implicated in these strategies.

Conclusion

It's hard to imagine in retrospect, but before Lewis' work in *Convention* the notion

of a convention was deeply mysterious. Here is another debt the profession owes Lewis: thanks to his concepts and analytical tools, we not only understand how conventions can be the natural outgrowth of our interactions with one another, but we can in fact pursue sophisticated inquiry into the nature and scope of those conventions. Lewis's contribution is achieved through ideas whose influence has only increased over time – his account of coordination remains crucial to our understanding of strategic interaction, signaling games continue to provide a fundamental tool for understanding communication, and theorists still must acknowledge that the achievement of coordinated outcomes by rational choice depends on the common knowledge that Lewis associates with convention. In this sense, Lewis's contribution stands as a testament of philosophy not only to find the right questions but to answer them as well.

Bibliography

Baker, Mark. 2002. *The atoms of language*. Basic Books.

Burge, Tyler. 1975. On knowledge and convention. *The Philosophical Review* 84. 249–255.

Chomsky, Noam. 1981. *Lectures in government and binding* Studies in generative grammar 9. Dordrecht: Foris.

Clark, Robin. 2012. *Meaningful games: Exploring language with game theory*. MIT.

Devitt, Michael. 2006. *Ignorance of language*. Oxford: Oxford University Press.

von Frisch, Karl. 1967. The dance language and orientation of bees. Cambridge, MA: Harvard University Press.

Gilbert, Margaret. 1989. *On social facts*. London: Routledge.

Grafen, Alan. 1990. Biological signals as handicaps. *Journal of Theoretical Biology* 144(4). 517–546.

Grice, H. P. 1957. Meaning. *The Philosophical Review* 66(3). 377–388.

Grice, H. Paul. 1989. *Studies in the way of words*. Harvard.

Hume, David. 1740. *A treatise of human nature: Being an attempt to introduce the experimental method of reasoning into moral subjects*. London: John Noon. http://www.gutenberg.org/ebooks/4705.

Lachmann, Michael, Szabolcs Számadó & Carl T. Bergstrom. 2001. Cost and conflict in animal signals and human language. *Proceedings of the National Academy of Sciences of the United States of America* 98(23). 13189–13194. http://dx.doi.org/10.1073/pnas.231216498.

Lepore, E. and M. Stone, *Imagination and Convention*, Oxford: Oxford University Press, 2014, forthcoming.

Lewis, David K. 1969. *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.

Marmor, Andrei. 2009. *Social conventions: From language to law*. Princeton: Princeton University Press.

Millikan, Ruth Garrett. 1998. Language conventions made simple. *The Journal of Philosophy* 95(4). 161–180.

Nash, Jr., John F. 1950. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences of the United States of America* 36(1). 48–49. http://www.jstor.org/pss/88031

Pesetsky, David. 1999. Linguistic universals and universal grammar. In *The MIT encyclopedia of cognitive science*, 476–478. MIT.

Rescorla, Michael, "Convention", *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2011/entries/convention/>.

Russell, Bertrand. 1921. *The analysis of mind*. London: Allen and Unwin.

Schelling, Thomas C. 1960. *The strategy of conflict*. Cambridge, Massachusetts: Harvard University Press.
Skyrms, Brian. 2010. *Signals: Evolution, learning and information*. Oxford: Oxford University Press.
Smith, J. Maynard & G. R. Price. 1973. The logic of animal conflict. *Nature* 246. 15–18. http://dx.doi.org/10.1038/246015a0
Spence, Michael. 1973. Job market signaling. *The Quarterly Journal of Economics* 87(3). 355–374.
Su, Songkun, Fang Cai, Aung Si, Shaowu Zhang, Jürgen Tautz & Shenglu Chen. 2008. East learns from west: Asiatic honeybees can understand dance language of european honeybees. *Public Library of Science ONE* http://dx.doi.org/10.1371%2Fjournal.pone.0002365.
Sugden, Robert. 2004. *The economics of rights, co-operation and welfare*. New York: Palgrave 2nd edn.
Vanderschraaf, Peter. 1998a. The informal game theory in Hume's account of convention. *Economics and Philosophy* 14. 215–247.
Vanderschraaf, Peter. 1998b. Knowledge, equilibrium and convention. *Erkenntnis* 49. 337–369.
von Neumann, John & Oskar Morgenstern. 1947. *Theory of games and economic behavior*. Princeton, New Jersey: Princeton University Press 2nd edn.