Convention, Causation, and Grounding

Daniel W. Harris

March 31, 2018

Abstract

Theories of linguistic convention attempt to explain the states of populations that allow them to communicate efficiently with language. I argue that whereas some major theories are best understood as causal explanations, others are best understood as grounding explanations. It follows that some debates have pitted theories with orthogonal explanatory aims against one another. I argue, for example, that the theories of David Lewis and Ruth Garrett Millikan, which have been generally understood as competitors, are compatible.

1 Semantic Coordination

Some populations of humans find themselves in states that allow them to use language to reliably and efficiently communicate rich and precise information with one another. I will say that these populations are *semantically coordinated*, while remaining neutral, for now, about just what being in such a state amounts to.

Here's an example. I recently made baked beans from a family recipe that my mother explained to me. This is step six:

(1) Bake at 275° for seven hours, adding liquid occasionally if the beans look dry, and leaving them uncovered to brown for the last hour.

The beans turned out well, and this was a direct result of the complex plan that I formed upon hearing (1). If my mother and I hadn't been semantically coordinated, she wouldn't have been able to get me to form this plan—at least, not nearly as reliably and efficiently.

To be semantically coordinated is to be in a state that solves what David Lewis (1969) calls a coordination problem. There are indefinitely many ways that the human propensity to pair expressions with meanings could be systematized. A population's communicative interests could be served equally well by coordinating on

any of these systems of pairings. To solve the coordination problem presented by this space of options is to enter into a state of semantic coordination.

Such states are neither inevitable nor accidental, and they call out for explanation. Philosophers have approached this explanatory task by developing theories of linguistic convention. According to David Lewis, for example, linguistic conventions are solutions to semantic coordination problems (1969, ch.1) that are "sustained by an interest in communication" (1975, 167). Ruth Millikan (2005, 25) says that "…linguistic conventions…have an essentially communicative function", and that they serve this function by getting communities of speakers to coordinate on "semantic mapping functions" that systematically pair forms with meanings. Most other theorists of linguistic convention have embraced similar goals.[1]

2 Causal Explanation versus Grounding Explanation

So theories of linguistic convention aim to explain the efficiency of linguistic communication by explaining states of semantic coordination. But there are different kinds of explanation. For example, one could try to explain the fact that my mother and I are presently coordinated by saying what it is about us at this very moment—what it is about our present states of mind, say—that makes it the case that we are in a position to communicate with language. On the other hand, one could try to explain our present state of coordination by telling the story of how we came to be this way—the story of how we became competent speakers of a similar dialect of English, for example. Explanations of these two kinds might turn out to be more or less orthogonal. The state of mind that underlies my command of English may have come about in various ways; I might have acquired it as a child through interaction with my mother, for example, or she and I could have come to be competent users of English separately, as adults. And either of these stories about how we became coordinated might be compatible with a variety of theories of the psychological states underlying human linguistic competence. There are, in other words, at least two orthogonal dimensions along which we might try to explain states of coordination. Metaphysicians have recently taken pains to distinguish explanations of these two kinds, which they dub *grounding* explanations and *causal* explanations, respectively.

Roughly speaking, a causal explanation accounts for a phenomenon by spelling out the events that led to it and saying how they brought it about. If the phenomenon to be explained is a persistent state, a causal explanation of it might also be concerned with the factors that sustain it. One can also give a causal explanation of a *type* of phenomenon by spelling out the kinds of events, processes, or states of affairs

[1] See, for example, Bennett (1976, ch.6); Loar (1981, §10.4, 253–260); Miller (2001, 21, 23); Schiffer (1972, ch.5).

that will tend to give rise to or sustain tokens of it. For example, a causal explanation of high obesity rates in populations of humans might work by pointing to the kinds of dietary, environmental, social, and cultural factors that, if they prevail in a human population, will tend to cause obesity. And a causal explanation of a high rate of obesity in a particular human population will point out which of these particular factors are at work there.

To give a grounding explanation of a fact is to spell out the more fundamental facts in virtue of which it obtains—i.e., the facts that ground it, that *make it the case*, or *in virtue of which* it obtains. A grounding explanation of the fact that a population has a high obesity rate at a given time, for instance, might show how that statistical fact is grounded in a collection of physiological and social facts about the members of the population at that time. It is helpful to think of grounding explanations as accounting for a fact about one level of reality by pointing to a lower-level fact that underlies it.[2]

Grounding and causation have plenty in common: they're both explanation-supporting strict partial orders—what some have called 'explanatory relations'.[3] They give rise to some of the same controversies; debates over the principle of sufficient reason can be construed as debates over the well-foundedness of either causation or grounding, for example. But nearly everyone who is not a grounding skeptic accepts some version of the causation–grounding distinction, and, by extension, the distinction between the two corresponding genres of explanation, and I will follow this metaphysical mainstream.[4,5]

Although the causation–grounding distinction is relatively uncontroversial, the criteria that mark it are not. A common metaphorical characterization is that causation and grounding yield different dimensions of explanation, with causation some-

[2]Interest in the nature of grounding and grounding explanation has recently been (re)awakened, in part by showing that philosophers had previously relied on some such notion unreflectively in many different literatures. Some influential discussions of grounding include Audi (2012); Fine (2001; 2012); Rosen (2010); and Schaffer (2009). One open question about the grounding relation concerns the kind of relata it connects. I will sometimes talk about facts grounding other facts, and sometimes talk about states grounding other states. I could rephrase the latter in terms of the former, or both in terms of properties grounding properties, if that were deemed preferable.

[3]Fine (2012, 39). Schaffer (2012) argues that grounding is not transitive, and so not a partial order, but he holds the same view about causation, and elsewhere (2015) he argues that grounding is the image of causation.

[4]Bennett (2017) argues that grounding and causation fall into a broader family of "building relations" which also includes relations like realization and constitution, some of which (such as the relation between a cake and its ingredients) are explanation supporting. I don't think any of Bennett's other relations could be at work in explanations of semantic coordination.

[5]Bennett (2011, 93–94) and Schaffer (2015, §4.5) take seriously the possibility that grounding and causation are the same relation, but both ultimately reject it. Kivatinos (2017) gives the only sustained defense of grounding-causation unity that I know of.

times represented horizontally and grounding represented vertically.[6] What does this metaphor amount to? A tempting answer is that causation is a diachronic relation, unfolding over time, whereas grounding is a synchronic relation, unfolding at a time across levels of fundamentality.[7] Indeed, I do wish to claim that whereas causation is always diachronic, grounding is at least sometimes—and perhaps always or at least characteristically—synchronic, and that this is one property that distinguishes the two relations. But there are some controversial cases that may lead us to think that grounding can sometimes be diachronic. It is tempting to read some philosophers as arguing that the fact that I am a human and the fact that I have intentional mental states are grounded in the fact that I have a certain kind of evolutionary or personal history, for example. On these views, facts about species membership and mental representation are grounded in causal-historical facts, and the properties of belonging to species and being in intentional-mental states are historical kinds.[8]

The view that there are any such historical kinds is controversial, but I will concede their existence for the sake of argument, in part because doing otherwise would unnecessarily cloud the dialectic of this essay. My main point in §3 will be that Millikan's theory of linguistic convention is best construed as offering a purely causal explanation of states of semantic coordination, one that is orthogonal to questions of grounding. Since Millikan is the most prominent proponent of historical-kind theories in other domains, the causation–grounding distinction is particularly tricky to draw in her neighborhood of logical space. Still, I think that even Millikan should recognize the distinction, and I think that her theory of convention should be counted as a purely causal one, even by her own standards. Since my plan is to demonstrate this in §3, I will need to get as clear as possible on the causation–grounding distinction first, which will be my goal in the remainder of this section.

I'll take the synchronicity of an explanation to be sufficient for its being a grounding explanation, then, but I won't assume that diachronicity is sufficient for an explanation to be causal. What else distinguishes causal from grounding explanations? Other popular—though by no means universally accepted—criteria include the following. (i) Causation and grounding have different modal profiles: causes nomologically necessitate their effects whereas grounds metaphysically necessitate what they ground.[9] (ii) Grounding is always deterministic whereas causation is

[6] E.g., Bennett (2017, 67).

[7] To my knowledge, there has been no sustained defense of the claim that grounding is always synchronic, but some have made (or at least entertained) the claim: Leuenberger (2014, 156); Rosen (2010, 118n8); Skiles (2015, §5.3). The claim is at least hinted at or implied by others; e.g., Schaffer (2012, 122): "Grounding is something like metaphysical causation. Roughly speaking, just as causation links the world across time, grounding links the world across levels".

[8] Although she does not frame her views in terms of grounding, it is tempting to read this view back into Millikan (1984; 2005, ch.6).

[9] This criterion is endorsed by many authors, including Audi (2012); Bennett (2011); Rosen (2010);

sometimes nondeterministic, and so, in the words of Schaffer, whereas "grounding implies an associated (metaphysical) supervenience, causation does not imply an associated (nomological) supervenience" (2015, §4.5). (iii) Grounding is an internal relation, in the sense that it holds wholly in virtue of the nature of the ground; by contrast, causation is an external relation, in the sense that nothing about the natures of either relatum guarantees that they are causally connected.[10] (iv) Some have argued that grounding, unlike causation, must be well-founded.[11] (v) In at least some cases of grounding—and, some would say, typically or always—it makes sense to say that what is grounded is "nothing over and above" what grounds it; but it never makes sense to say that an effect is nothing over and above its cause.[12] In addition to being somewhat controversial, these criteria often aren't easily wielded in order to adjudicate controversial cases. Nonetheless, in arguing that Millikan's and Lewis's theories of convention are best construed as attempts at purely causal and purely constitutive explanation, respectively, I will steer around these controversies. All of the criteria agree about the cases at issue here, and so it doesn't matter which we accept.

In practice, many have taken a know-it-when-I-see-it approach to distinguishing grounding from causation. This is unsurprising given that the literature has developed by moving from paradigm cases to competing attempts at characterizing them. Even if we countenance historical kinds, distinguishing explanations of the two kinds is usually easy and intuitive. According to teleofunctionalism (spelled out in terms of grounding), my belief that dogs are friendlier than cats is grounded in facts about how psychological states of the same type acquired their proper function in my and other humans' psychological economies; the same belief is causally explained by the fact that dogs have usually been more welcoming to me than cats.

As I suggested earlier, explanations of semantic publicity may likewise deal in either causation or grounding. Take the fact that my mother and I are semantically coordinated on the meaning of 'if the beans look dry'. A grounding explanation of this fact would say what it is about me and my mother in virtue of which we are coordinated on the meaning of 'if the beans look dry'. A causal explanation would tell the story of how we got that way.

Part of the point of the next two sections will be to show that attending to the causation–grounding distinction can bring greater clarity to debates about the nature of linguistic convention. There is some risk of anachronism here. Each of the theories of convention that I'll discuss was formulated prior to the recent resur-

Trogdon (2013). Skiles (2015) and Leuenberger (2014) argue that there are counterexamples to the idea that grounds metaphysically necessitate what they ground.

[10] Bennett (2011, 32); Rosen (2010, §4.5); Schaffer (2015, §4.5).

[11] E.g., Schaffer (2015, §4.5).

[12] For a discussion of this argument in the literature, see Skiles (2015, §4.5).

gence of work on grounding. It is therefore plausible that the authors of these theories were not sensitive to the distinction. Indeed, much work on grounding has been motivated, at least in part, by the perceived limitations of the toolkit of modal metaphysics—a toolkit that was most adeptly deployed in the work of David Lewis, whose theory of convention is the most influential contemporary option, and one of my subjects here.[13] I will proceed on the assumption that the causation–grounding distinction is a genuine and fruitful one, and try to show that we stand to gain understanding by projecting it back onto the major theories of linguistic convention. Whereas some of these theories are most charitably construed as theories of what ground states of semantic coordination, others are best construed as theories of how those states arise and persist.

3 Causal Theories of Linguistic Convention

My plan in this section is to say what a causal theory of linguistic convention would look like, using the theories of Ruth Millikan and Brian Skyrms as examples. If I am right, then these theories are best understood as causal explanations of how states of semantic coordination arise and persist, while remaining relatively neutral about what grounds these states.

Consider Millikan's theory. For a practice to be conventional, Millikan argues, is for it to come about in a way that meets a pair of conditions. First: conventions have been reproduced; they have been "handed down from one person, pair, or group of persons to others" (2005, 56). Second, conventions have been reproduced at least in part "by weight of precedent". This second condition ensures the arbitrariness of conventions—in Millikan's words, "that if the pattern [of behavior] has a function, then it is not the only pattern that might have served that function about as well" (2005, 56). To label a practice conventional in this sense is to explain it by articulating a causal story of how it gained currency in a population. Within the general category of conventions, according to Millikan, are the special case of coordinating conventions, which are conventions whose proper function is to coordinate the activities of the members of a group. According to her own background commitments, to say that a practice has a certain kind of outcome as its proper function is to say that the fact that it has caused this outcome in the past has played a significant role in causing it to be reproduced. And so, to label a convention a coor-

[13] Some other authors have recently attempted to reconstruct some of Lewis's metaphysical positions in terms of grounding (e.g. Rosen 2010; Schaffer 2009). However, Lewis seems to have pointedly avoided making grounding claims. Although he argues that events have only causal explanations, he avoids commitment as to other kinds of explananda (Lewis, 1986). It seems clear, however, that Lewis either countenanced no concept of grounding or was skeptical of the concept's intelligibility (Daly, 2012).

dinating convention is to offer a more precise causal explanation of its origins than if one had merely labeled it a convention: its prevalence in the population is due to the coordination-related benefits that it has previously conferred, together with he fact that the population members have some mechanism for reproducing beneficial practices of this kind. Linguistic conventions, according to Millikan, are systems of coordinating conventions whose proper function is to facilitate communication. A linguistic convention serves this purpose by being organized according to a "semantic mapping function"—a mapping from linguistic forms to meanings that is sensitive to syntactic structure (Millikan, 2005, ch.3).

Millikan's theory of linguistic convention thus explains states of semantic coordination by showing that they will tend to arise and persist in populations whose members meet a trio of conditions: (i) they have a shared interest in reliably communicating, (ii) they have the tendency to reproduce behaviors that serve their interests, and (iii) their natures allow them to develop practices that recursively map syntactically complex linguistic forms to meanings. Natural-language users meet these three conditions, and so Millikan argues that her account goes some way toward explaining how we come to have and maintain a shared stock of expressions with shared meanings that allow us to communicate efficiently—i.e., how we come to be semantically coordinated.

Millikan's explanation of semantic publicity has the same structure as epidemiological explanations of other traits in human populations—explanations in which standing characteristics of the population are identified as causal factors that reliably lead to and maintain the trait to be explained. Compare, for example, a theory that explains obesity in humans by identifying the risk factors that tend to bring it about. Epidemeological explanations of this kind are causal explanations *par excellence*. This suggests that we should think of Millikan as offering a causal explanation as well.

The criteria that I discussed in the last section support this interpretation. The relation between semantic coordination and Millikan's three conditions—like the relation between obesity and its risk factors—is diachronic, metaphysically contingent, probabilistic, and external, and there is no temptation to say that semantic publicity is "nothing over and above" the constellation of factors that lead to it, any more than to say that obesity is "nothing over and above" its risk factors.

Even if we countenance historical kinds in general, as Millikan does, it is implausible that being in a state of semantic coordination is among them. Consider some time-slice of a well-functioning speech community with a common stock of publicly meaningful expressions—say, myself and my mother when she communicated to me via her baked-beans recipe. The fact that 'liquid' had the meaning it had for us—and, in turn, the fact that we could efficiently use the word to communicate about liquid—depended on some constellation of our psychological characteristics.

Whatever the relevant characteristics are, if we had lacked any of them (and any substitute characteristics that might have played the same psychological roles), then we would not have been semantically coordinated on a shared meaning for 'liquid', and we would not have been able to communicate as efficiently with it. Plausibly, the meaningfulness of 'liquid' for me and my mother was grounded in in the fact that we possessed these characteristics. But both the fact that 'liquid' had a shared meaning for me and my mother and the psychological facts that grounded this fact could have come about in numerous ways. Perhaps we reached our mental states via the kind of process that Millikan describes. But we can imagine a variety of other ways in which a pair of speakers might get into the same, coordinated psychological state: their linguistic abilities could be the result of an explicitly worked-out agreement, a series of extremely unlikely but causally unconnected accidents, a devious, large-scale neurological experiment, or a millennia-long evolutionary history that has genetically predetermined the way in which they speak.[14] But no matter how my mother and I reached our psychological states, those states explain the fact that 'liquid' had a shared meaning for us, and thus our ability to communicate with the word.

Millikan is aware of these possibilities, and she makes it clear that the aim of her theory of convention is merely to describe one kind of process by which semantic publicity could, and often or characteristically does, come about (2005, 1–2):

> I do not claim that this kind of conventionality is the only kind of conventionality there is. I do not claim there are no other senses in which even languages, nonnatural ones, might be 'conventional'; and I do not claim that it is definitional of all language that it has to be conventional. (Perhaps there are animal 'languages', languages that are inherited rather than conventional in the sense I shall describe.)

It follows from what Millikan says in this passage that a pair of individuals could have been semantically coordinated on all the same meanings they are actually coordinated on, even if their state of coordination had come about by an unconventional process. But despite this difference in etiology, their state of coordination could depend on the same psychological facts in either case: in one world, these psychological states came about by a process in keeping with Millikan's theory of convention; in other possible worlds, the same states came about by way of biological evolution, explicit agreement, or unlikely accident. The best way to understand all of this, I submit, is to say that Millikan's theory of convention is not a theory of what grounds states of semantic coordination.[15] What the theory offers us is a

[14] Cf. Chomsky (1980, 93); Peacocke (1976, 168–170)).

[15] Another way to put the point: to participate in a convention, in Millikan's sense, is to be a member

plausible empirical account of how states of semantic coordination often or normally come about and persist.

Millikan's theory remains relatively neutral about what grounds states of coordination: it abstracts away from questions about which kinds of mental states underlie semantic competence. For example—and foreshadowing a bit—it is compatible with her view that the grounds of semantic coordination are personal-level intentional mental states (e.g. beliefs), or non-propositional states of knowledge-how, or sub-personal level mental states, such as the state of speakers' language modules or faculties of language. Indeed, Millikan's theory of linguistic convention is neutral as to questions of human cognitive architecture. The theory is even neutral as to the nature of the mechanisms by which successful linguistic practices are reproduced. Perhaps we consciously copy actions that, we recognize, have helped us in the past; or perhaps our reproduction of successful behaviors is largely due to psychological mechanisms that are automatic, associative, subpersonal or otherwise inaccessible to conscious control. Again, these considerations bolster the case for thinking of Millikan's theory as a relatively pure example of causal explanation—one that is relatively orthogonal to issues of grounding.

Aside from Millikan's theory, the most prominent family of purely causal theories of linguistic convention has grown out of work on replicator dynamics, which itself grew out of game-theoretic models of natural selection (Harms, 2004; Skyrms, 2010). In theories of this kind, the processes that lead to states of semantic publicity are modeled as iterated signaling games—series of situations in which a sender issues a signal, and a receiver takes some action on the basis of the signal. By making modest and abstract assumptions about the rate at which successfully coordinated strategies for sending and receiving signals are replicated, it is possible to show with mathematical precision how these strategies can spread through a population until they reach stable equilibria. Like Millikan's theory, replicator-dynamic models of convention are purely causal, in that they work by showing how states of semantic publicity are produced and maintained over time in populations, without making any assumptions about the underlying states of senders and receivers that ground their strategies. This lack of grounding commitments can be seen from the fact that replicator-dynamic models typically include no assumptions about the underlying natures of the sender and receiver themselves. Some theorists have even argued that the same models they propose for human communication can also be applied to communication between other mammals, slime moulds, and bacteria—entities whose states of coordination are grounded in what are presumably rather different ways (Skyrms, 2010, 3–6).

---

of historical kind (after all: it is just to have gotten that way by a certain kind of process); but to be in a state of coordination is not to be a member of a historical kind.

We needn't sign on to any particular strategy for causally explaining states of semantic coordination in order to recognize that such explanations leave open the orthogonal dimension of explanation concerned with grounding. Indeed, there is some reason to think that purely causal explanations are consigned to a certain kind of shallowness until they are paired with grounding explanations. Although it may be possible to model communication between slime moulds and between humans in the same way at some level of abstraction, a full understanding of either will surely take us into grounding territory, since we'll want to know about the implementation details. In any case, even if Millikan or the replicator-dynamic theorists are entirely correct, there is more explaining to be done.

4 Grounding Theories of Linguistic Convention

I now turn my attention to David Lewis's theory of convention, and to a genre of theories of convention that have grown out of his work. What unites these theories is that they seek to explain states of semantic coordination in terms of the shared propositional attitudes of coordinated populations. My main goal in this section is to argue that theories in this genre are best construed as grounding explanations of states of semantic coordination.

I begin with Lewis, who makes the innocent-sounding claim that a state of semantic coordination can be explained by the fact that those coordinated participate in a convention to use a language. For Lewis, however, 'language' and 'convention' are technical terms. A language, according to Lewis (1975, 163), is:

> a function, a set of ordered pairs of strings and meanings. The entities in the domain of the function are certain finite sequences of types of vocal sounds, or of types of inscribable marks; if σ is in the domain of a language $\mathcal{L}$, let us call σ a sentence of $\mathcal{L}$. The entities in the range of the function are meanings: if σ is a sentence of $\mathcal{L}$, let us call $\mathcal{L}(\sigma)$ the meaning of σ in $\mathcal{L}$.

A convention, according to Lewis, is a regularity R in a group G that meets the following conditions (1975):

(i) everyone in G conforms to R;

(ii) everyone in G believes that everyone else in G conforms to R;

(iii) the belief that others conform to R gives the members of G "a good and decisive reason to conform to R";

(iv) there is a general preference among individuals in G for conformity to R;

(v) there is at least one other possible regularity R′ that would serve the purposes of the members of G just as well R; and

(vi) it is common knowledge among members of G that (i)–(v) obtain.

Lewis tells us that for any sentence S, proposition P, and group G, S means P for G just in case, for some language L that maps S to P, there prevails in G a convention to utter sentences of L only if L maps them to true propositions and to expect others to do the same (1975, 164–171). Lewis calls this a "convention of truthfulness and trust in L". In response to an objection put to him by Schiffer, Lewis also requires that genuine linguistic conventions be sustained by a shared interest in communication, here meaning that the "purposes" mentioned in clause (v) are communicative purposes (Schiffer, 1993, 233).

According to Lewis, the fact that the members of a group are semantically coordinated with respect to an expression *e*, and so the fact that they can communicate efficiently with *e*, are explained by the fact that the group participates in a convention to use a language that maps *e* to a common meaning. What sort of explanation is at issue in this claim? I think that we interpret Lewis most charitably if we reconstruct his theory of convention as being backed by a wholly synchronic relation between explanandum and explanans. If this is right, then his theory of convention should be construed as a grounding explanation.

We can arguably find just two diachronic elements in Lewis's theory. Condition (i) requires a convention to be a regularity, which entails that it has been going on for a while by the time a convention can be said to exist. (Moreover, condition (vi) requires that the participants in a convention have common knowledge of this fact.) Secondly, Lewis's condition (iii) might be interpreted as requiring that participants conform to a convention because they have conformed in the past (and realize this), and it might be tempting to interpret this 'because' in causal terms. (As we'll see below, this is how Millikan interprets Lewis.)

These diachronic elements of Lewis's theory can be removed without any loss of explanatory power. First, replace Lewis's (i) with (i*):[16]

(i*) Everyone in G believes that they conform to R.

Second, clarify (iii) to make it clear that it requires only that G-members' present (and possibly inaccurate) belief that they conform to R is what gives them a decisive reason to conform. This requires that we understand the reasons involved in

[16] Millikan (2005, 11–14) argues at length that conventional practices needn't be genuine regularities. See also Gilbert (1989, §4.5) and Davis (2003, §9.5).

a specific way—as 'internal' or 'motivating' reasons, in Williams' (1980) sense. If the reasons in question are to do any work in explaining why agents who possess them will normally be capable of efficient communication, I think we have to read the clause this way in any case.

(iii*) the belief that others conform to R gives the members of G a decisive internal reason to conform to R

If meeting Lewis's original conditions is sufficient to allow a group to communicate efficiently with language, then so is meeting these new conditions.[17] What I have removed are merely some assumptions about how a group with beliefs and dispositions that instantiate Lewis's conditions would have reached this state. One way that a group of agents could come to commonly believe that they conform to a certain linguistic regularity is that they actually do conform to it, and have observed this. But even if these beliefs were false—if, say, they had been implanted by aliens, or by a massive coincidence, or if they were the product of self-deception—the beliefs would guide group members' future communicative actions in just the same ways as if they were true. In other words: we lose nothing of Lewis's explanation of semantic coordination by stripping out its diachronic elements. The resulting theory posits a purely synchronic relation between explanandum and explanans, and is therefore a grounding explanation.

It might be objected that we could (and Lewis would) consider the synchronic relation in question to be mere supervenience. However, one of the reasons for countenancing grounding in the first place is that supervenience and related modal notions are too course-grained to back explanations (Bliss and Trogdon, 2016; Rosen, 2010). Given that my aim here is not to answer generalized grouding skepticism, and given the explanatory aims of theories of convention, I think that my adjusted version of Lewis's theory is best understood as a grounding explanation—one whose centtal claim is that states of semantic coordination are grounded in complex systems of propositional attitudes.

There have been many criticisms of Lewis's theory. Some have objected to the role that Lewis gives to truthfulness and trust, in some cases arguing that linguistic conventions are built on regularities of other kinds.[18] Others have argued that well-functioning speech communities sometimes fail to meet some of Lewis's conditions (i)–(v).[19] Others have taken issue with Lewis's condition (vi), which requires

[17] Stotts (2015; 2016) suggests altering Lewis's theory in similar ways.

[18] Davis (2003); Kölbel (1998); Laurence (1996); Schiffer (1972, 1993, 2006); Williams (2002); Williamson (2000); Wilson and Sperber (2002).

[19] Burge (1975) imagines isolated speech communities from whose religious beliefs it follows either that their language is the only possible one (contra iii+vi) or that they would continue to to speak their

speech communities to possess common knowledge about their linguistic practices or policies. Common knowledge is normally defined as follows.

(2) COMMON KNOWLEDGE

(ia) A knows that p.
(ib) B knows that p.
(iia) A knows that B knows that p.
(iib) B knows that A knows that p.
And so on, *ad infinitum*.

The claim that semantically coordinated language users must stand in this relation to propositions about how they use language is almost certainly too strong. Some who have stayed within Lewis's overall framework have taken issue with Lewis's appeal to *knowledge*, as opposed to some weaker propositional attitude. Bach and Harnish (1979) substitute a notion of mutual belief. Another alternative is Stalnaker's (1984; 2014) notion of *acceptance*, which is a state of treating a proposition as true, though possibly only for some provisional purpose. And several authors have suggested that the mere potential for common knowledge or common belief could do the job (Clark and Marshall, 1992; Schiffer, 1972; Sperber and Wilson, 2002). Another way of weakening (2) is to reduce the degree of iteration required. Some have argued that infinitely intersubjectively iterated attitudes are either impossible (Lederman, 2017) or just overkill for real-world purposes (Bach and Harnish, 1979). A useful generalization of common knowledge, which is spelled out by Clark and Marshall (1992), is shared knowledge$_n$, which a pair of agents possess just in case they satisfy conditions (ia)–(nb) of (2). (Common knowledge is shared knowledge$_\infty$.) In general, we can say that a group bears a shared attitude to p just in case they have shared knowledge$_n$, shared belief$_n$, shared acceptance$_n$, or the potential for one of these attitudes to p (for some $n \geqslant 1$).

Several major theories of convention can be seen as alterations to Lewis's theory to deal with some of these issues. What they all share is that they all attempt to explain semantic coordination in terms of the fact that the coordinated agents bear some shared attitude (perhaps weaker than common knowledge) to a proposition about their linguistic regularities, practices, or policies (perhaps weaker than Lewis's (i)–(v)).[20] It should be clear that none of these ways of altering Lewis's theory (as I have reconstructed it) would turn it from a grounding explanation into a causal explanation.

language even if they could no longer communicate with it (contra v+vi). If their religious beliefs are powerful enough, these scenarios might constitute violations of (iii) and (v), independently of (vi). See also Stotts (2015, 2016).

[20] E.g. Gilbert (1989); Miller (2001); Schiffer (1972, 1993).

5 Rational and Arational Causal Explanations

It may be objected that by removing the diachronic elements from Lewis's theory and turning it unambiguously into a grounding explanation, I have in fact elided some of Lewis's explanatory ambitions.

Certainly, Lewis's theory *suggests* a kind of causal explanation. The thought is that agents who conform to a linguistic regularity will normally observe that they do so, and this observation will normally lead them to form beliefs and preferences about themselves and each other that meet Lewis's conditions. And once a group of agents forms beliefs that instantiate Lewis's conditions, these beliefs become rationally self-reinforcing, because it is rational to continue believing as long as others do the same. This is what might be called a *rational-causal* explanation: it describes an essentially rational process by which the mental states underlying semantic coordination arise and persist.

It is tempting to pair Lewis's theory with this sort of rational-causal explanation for several reasons. First, his theory's main theoretical posits are propositional attitudes of the kind that canonically arise and persist through rational processes. If all is going well, we get our beliefs as a result of perceptual or inferential processes that lend them justificatory support. It is therefore natural to think of Lewisian constellations of beliefs as having come about in this way.

Second, although Lewis's official theory of convention is spelled out in the necessary and sufficient conditions discussed in the previous section, he surrounds his statements of this theory with informal remarks that suggest a kind of rational-causal explanation. For example, he notes Schelling's (1960) finding that groups of agents often manage to solve novel coordination problems without communicating by focusing on the same, highly salient coordination equilibrium. Lewis then speculates, in rational-causal terms, about how this process works (1969, 35–36):

> How can we explain coordination by salience? The subjects might all tend to pick the salient as a last resort, when they have no stronger ground for choice. Or they might expect each other to have that tendency, and act accordingly; or they might expect each other to expect each other to have that tendency and act accordingly, and act accordingly; and so on. Or—more likely—there might be a mixture of these. Their first- and higher-order expectations of a tendency to pick the salient as a last resort would be a system of concordant expectations capable or producing coordination at the salient equilibrium.

And once a solution has been reached once, Lewis argues, this precedent will make a repeat solution salient—again, by leading to agents' mutual expectations in a rational-

causal way (1969, 36):[21]

> An easier, and more common, case is that of a *familiar* coordination problem without communication. Here the agents' source of mutual expectations is precedent: acquaintance with past solved instances of their present coordination problem.

So, although the theory of convention that Lewis spells out in detail posits (what I have reconstructed as) a synchronic, grounding relation between states of coordination and systems of propositional attitudes, he also had in mind a rational-causal explanation of how these systems of propositional attitudes typically arise and persist. The crucial element in this explanation is *mindreading*—the capacity to infer others' mental states on the basis of their behavior together with whatever other background information that we might have about them. This rational-causal story is intended to apply to all sorts of conventions, including linguistic conventions.

Still, we should keep in mind that the rational-causal explanation sketched above is merely *suggested* by Lewis's theory of linguistic convention, not entailed by it. It is not necessary for agents who conform to Lewis's conditions to have reached this state by way of inferring each other's beliefs or expectations on the basis of salience and precedent. There are other ways in which the relevant beliefs could, at least in principle, have been formed. Some of these are likewise rational-causal: a group of agents could have acquired these beliefs by being explicitly taught in a language class conducted in some other language, for example. But shared beliefs could also arise in *arational* ways: via biological evolution, a mad scientist's experiments, etc. A further option that I will briefly explore below is that humans are equipped with dedicated, arational mechanisms for forming our metalinguistic beliefs.[22] And so, insofar as Lewis should be read as offering a causal explanation of semantic coordination, this explanation is independent of the grounding explanation that I have reconstructed from the theory he spells out explicitly.

How plausible is the kind of rational-causal explanation suggested by Lewis? To answer this question, we should seek empirical evidence. Lewis's explanation is, after all, an empirical hypothesis. In particular, we should look for evidence about whether becoming a user of a language (and staying that way) is a process that generally depends on mindreading and other rational-causal cognitive processes. Our

[21] Cf. Schiffer (1972, 122–124).

[22] Those who posit arational language-acquisition mechanisms also tend to think that these mechanisms lead to nondoxastic mental states, and that linguistic competence is not grounded in propositional attitudes. I will consider this possibility in the next section. For now, I wish to consider the possibility that the Lewisian grounding explanation is correct but that his rational-causal explanation is wrong.

main source of evidence about how humans enter into states of semantic coordination comes from the psycholinguistic study of language acquisition. There, we find some support for Lewis's account, but also some challenges for it.

The most favorable evidence for Lewis comes from the study of lexical acquisition, and, in particular, the study of how humans quickly and often permanently become competent with common nouns after minimal exposure—what psycholinguists call "fast-mapping" (Carey, 1978). For example, Markson and Bloom (1997) have found that fast-mapping is correlated across all age groups with the capacity to learn arbitrary information when that information is presented linguistically. And Bloom (2000, ch.3) has drawn on a wide range of empirical evidence to argue that our capacity to learn the meanings of many lexical items mainly boils down to mindreading and other domain-general forms of inference.

On the other hand, there is a considerable body of evidence that many aspects of language acquisition depend on dedicated and domain-specific mechanisms that operate independently of the kind of general-purpose mindreading and learning mechanisms suggested by Lewis. This view has been widely defended by generative linguists, many of whom posit a dedicated Language Acquisition Device (LAD) whose operations are dissociated from domain-general learning capacities.[23] Importantly for present purposes, the operations of the LAD is not typically understood in terms of the posits of personal-level intentional psychology—posits like belief, intention, inference, and so on. Instead, the LAD is normally treated as a computational system whose informational resources are cut off from language learners' beliefs, desires, intentions, and other personal-level intentional states.

There are several reasons to take this idea seriously. One is that many aspects of language acquisition are dissociated from domain-general learning: the capacity to acquire the syntactically rich aspects of human language peaks during a critical period in childhood but then significantly declines, whereas domain-general learning follows no such pattern (Newport, 1990). And acquiring a new lexical item requires representing it as having a range of syntactically and compositionally relevant properties that most speakers apparently do not have beliefs about—properties like syntactic category and argument structure—suggesting that it must be some domain-specific acquisition mechanism doing the representing (Pinker and Jackendoff, 2005, §2.1).

My point is not to argue that Lewis's rational-causal explanation is clearly false, or that some arational-causal explanation is clearly superior. In fact, I am inclined to an intermediate position, on which acquiring competence in a given language requires some combination of general-purpose learning, domain-specific language-

[23] The essence of this idea, and the term 'language acquisition device', are due to Chomsky (1965, 32). For a recent book-length overview of acquisition research in this tradition, see Guasti (2017).

acquisition mechanisms, and innate competence. The interesting question, and the question that occupies contemporary acquisition theorists, is how these three components are weighted, and how they interact. This is a decades-old debate, and settling it would go far beyond the scope of this essay.

Rather, there are two important takeaways from this section. First, even if we assume that some Lewisian theory of convention provides us with the correct grounding explanation for states of semantic coordination, this leaves open the possibility of a range of possible causal explanations. The kind of rational-causal explanation suggested by Lewis is among the options, but so are arational-causal explanations that appeal to a domain-specific language-acquisition device, as well as intermediate explanations that appeal to both kinds of learning capacities.

Second—and perhaps more surprisingly—*any* of these accounts is compatible with the causal theories of convention discussed in §3. To see this, recall that both Millikan and Skyrms offer causal explanations that operate at a very high level of abstraction. For example, Millikan's theory causally explains a population's state of semantic coordination by saying that the population members have (i) a shared interest in communicating, (ii) a tendency to reproduce behaviors that serve their interests, and (iii) a capacity to recursively map linguistic forms to meanings. The debate about whether we acquire language by means of general-purpose or language-specific mechanisms is a debate about the implementation details of condition (ii). If Lewis's remarks are correct, then we reproduce successful linguistic practices by means of a series of inferences from salience, precedent, and our expectations about others' beliefs and desires. If generative linguists are right, then we integrate innate information with data present in our environment with the help of a dedicated language-acquisition device. Neither of these accounts conflicts with Millikan's theory of convention. They merely suggest different ways in which populations of humans may instantiate her theory.

This line of thought brings the dialectic between Lewis and Millikan into much sharper focus. Millikan's main explicit criticism of Lewis consists in pointing out that conventions needn't arise and persist in a rational-causal way (2005, 56):

> …for those [conventions] that spread because they [solve coordination problems], the 'because' is almost never a reasoned because but some more mundane kind of causal because. The rest of us conform to linguistic conventions in exactly the same unreasoned way that the idiot and the child do.

This passage does not amount to a criticism of my reconstruction of Lewis's theory of convention, which tells us how states of semantic coordination are grounded but is neutral about how those states come about. It could be that states of semantic

coordination are grounded in systems of propositional attitudes as Lewis's theory tells us, but that these attitudes come about and persist via arational mechanisms. Millikan's criticism does not address this possibility. Rather, she is criticizing the rational-causal theory that Lewis suggests but does not actually build into his theory. Ultimately, whether Millikan is right is largely a matter for empirical psycholinguistics to decide.

6 Non-Lewisian Grounding Explanations

In the previous section, I considered the possibility that Lewis is right to think that states of semantic coordination are typically grounded in systems of propositional attitudes, but wrong to think that these attitudes normally arise in rational-causal ways. But here is another possibility worth considering—that states of semantic coordination aren't grounded in propositional attitudes at all, but in psychological states of some other kind.

Several variations on this idea have been defended. A common assumption of such views is that the semantic coordination of a community can be explained by the fact that community members possess suitably related grammars, and that possession of a grammar is best understood in terms of mental states other than belief or other personal-level intentional mental states. For example, Stich (1978, 508) argues that "the states which store grammatical information" are "subdoxastic states" which "are largely isolated from the body of a subject's beliefs" but "can plausibly be assumed to play a role in the formation of beliefs about what has been said to a person".[24] Some have followed Chomsky in positing a faculty of language—an organ of the mind/brain possession of which should not be understood in terms of intentional mental states.[25] Others have followed Fodor in taking possession of a grammar to consist in the possession of an informationally encapsulated input–output system—a system that causally mediates speech perception and production in a way that is insensitive to language users' beliefs.[26] And Devitt (2006; 2012) has argued that semantic competence is a kind of non-propositional knowledge-how. Intermediate positions are possible, as is the view that semantic coordination is grounded in mental states of more than one of these kinds.

I won't examine the arguments for and against these views here. My point is only

[24] See Pereplyotchik (2017) for a more detailed account of grammatical competence in terms of subpersonal-level mental states.

[25] Chomsky (1965, 1980, 1986). On the point that possession of a faculty of language is not a personal-level intentional mental state, see Chomsky (1980, 93; 2000, 23) and Collins (2004).

[26] Borg (2004); Fodor (1983); Fodor et al. (1974); Frazier (1987); Pettit (2005); Sperber and Wilson (1995). Some have understood Fodor's view as a version of the Chomskian idea of a faculty of language (e.g. Laurence 1996), but Collins (2004) persuasively argues that this is not Chomsky's own view.

that to the extent that any of these views is right, Lewis is doubly wrong. If being a party to a state of semantic coordination is not grounded in one's propositional attitudes, but in some other kind of mental state, Lewis's theory of linguistic convention (construed as a grounding claim) is incorrect. And since non-intentional mental states don't come about via the sort of rational inferences from salience, precedent, and meta-expectations that Lewis imagines, but via some arational processes, we shouldn't expect Lewis's rational-causal explanation of semantic coordination to apply either.[27]

On the other hand, the purely causal theories of Millikan and others are fully compatible with a these alternative grounding explanations. We need only discover arational mechanisms by means of which subpersonal states, or states of the faculty of language, or states of modular input-output systems, or states of non-propositional knowledge-how, tend to arise and persist in populations of related human minds. If we can find such mechanisms, we will have uncovered one of the crucial components of Millikan's sort of causal explanation of semantic coordination.

7 Conclusion

My contention has been that we can make better sense of debates about the nature of linguistic convention if we distinguish causal explanations from grounding explanations and recognize some theories as aiming at the former and others at the latter. One important upshot of this line of thought is that some theories that have previously been thought of as competitor views—notably Lewis's and Millikan's theories—turn out to be offering orthogonal kinds of explanation that that could, in principle, both be correct.

References

Audi, P. (2012). Grounding: Toward a theory of the in-virtue-of relation. *Journal of Philosophy*, 109(12):685–711.

Bach, K. and Harnish, R. M. (1979). *Linguistic Communication and Speech Acts*. MIT Press, Cambridge, Mass.

Bennett, J. (1976). *Linguistic Behaviour*. Cambridge University Press, Cambridge, UK.

Bennett, K. (2011). Construction area (no hard had required). *Philosophical Studies*, 154:79–104.

Bennett, K. (2017). *Making Things Up*. Oxford University Press, Oxford, UK.

Bliss, R. and Trogdon, K. (2016). Metaphysical grounding. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Fall 2007 ed. edition.

Bloom, P. (2000). *How Children Learn the Meanings of Words*. MIT Press.

Borg, E. (2004). *Minimal Semantics*. Oxford University Press.

[27] See Chomsky (1980, 93–95), who argues that one of the differences between propositional knowledge and knowledge of language is that the latter is not formed by the kinds of epistemic processes by which we normally form beliefs, but is instead caused by triggering experiences in an arational way.

Burge, T. (1975). On knowledge and convention. *Philosophical Review*, 84(2):249–255.

Carey, S. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, 15:17–29.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA.

Chomsky, N. (1980). *Rules and Representations*. Columbia University Press, New York.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. Convergence. Praeger, Westport, CT and London.

Chomsky, N. (2000). *New Horizons in the Studyof Language and Mind*. Cambridge University Press.

Clark, H. H. and Marshall, C. R. (1992). Definite reference and mutual knowledge. In Clark, H. H., editor, *Arenas of Language Use*, chapter 1, pages 9–59. University of Chicago Press and Center for the Study of Language and Information.

Collins, J. (2004). Faculty disputes. *Mind and Language*, 19(5):503–533.

Daly, C. (2012). Scepticism about grounding. In Correia, F. and Schneider, B., editors, *Metaphysical Grounding: Understanding the Structure of Reality*. Cambridge University Press.

Davis, W. (2003). *Meaning, Expression, and Thought*. Cambridge University Press.

Devitt, M. (2006). *Ignorance of Language*. Oxford University Press.

Devitt, M. (2012). Linguistic knowledge. In Bengson, J. and Moffett, M. A., editors, *Knowing How: Essays on Knowledge, Mind, and Action*. Oxford University Press.

Fine, K. (2001). The question of realism. *Philosophers' Imprint*, 1(1):1–30.

Fine, K. (2012). A guide to ground. In Correia, F. and Schneider, B., editors, *Metaphysical Grounding: Understanding the Structure of Reality*. Cambridge University Press.

Fodor, J. (1983). *The Modularity of Mind*. MIT Press, Cambridge, Mass.

Fodor, J. A., Bever, T. G., and Garrett, M. F. (1974). *The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar*. McGraw-Hill.

Frazier, L. (1987). Theories of sentence processing. In Garfield, J., editor, *Modularity in Knowledge Representation and Natural-Language Understanding*, chapter 15, pages 291–307. MIT Press, Cambridge, MA.

Gilbert, M. (1989). *On Social Facts*. Routledge.

Guasti, M. T. (2017). *Language Acquisition: The Growth of Grammar*. MIT Press, Cambridge, MA, second edition.

Harms, W. F. (2004). *Information and Meaning in Evolutionary Processes*. Cambridge University Press, Cambridge, UK.

Kivatinos, T. (2017). *Grounding, Causation, and the Unity of Ontological Structure*. PhD thesis, The Graduate Center, CUNY, New York.

Kölbel, M. (1998). Lewis, language, lust and lies. *Inquiry*, 41(3):301–15.

Laurence, S. (1996). A chomskian alternative to convention-based semantics. *Mind*, 105:269–301.

Lederman, H. (2017). Uncommon knowledge. *Mind*, forthcoming.

Leuenberger, S. (2014). Grounding and necessity. *Inquiry*, 57(2):151–174.

Lewis, D. K. (1969). *Convention: A Philosophical Study*. Harvard University Press.

Lewis, D. K. (1975). Languages and language. *Minnesota Studies in the Philosophy of Science*, VII:3–35.

Lewis, D. K. (1986). Causal explanation. In *Philosophical Papers, Volume II*, pages 214–240. Oxford University Press, Oxford.

Loar, B. (1981). *Mind and Meaning*. Cambridge University Press, Cambridge, UK.

Markson, L. and Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, 385(27):813–815.

Miller, S. (2001). *Social Action: A Teleological Account*. Cambridge University Press.

Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories*. MIT Press.

Millikan, R. G. (2005). *Language: A Biological Model*. Oxford University Press.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, 14:11–28.

Peacocke, C. (1976). Truth definitions and actual languages. In Evans, G. and McDowell, J., editors, *Truth and Meaning: Essays in Semantics*. Oxford University Press.

Pereplyotchik, D. (2017). *Psychosyntax: The Nature of Grammar and its Place in the Mind*. Springer.

Pettit, D. (2005). Belief and understanding: A rejoinder to gross. *Mind*, 114(453):67–74.

Pinker, S. and Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition*, 95:201–236.

Rosen, G. (2010). Metaphysical dependence: Grounding and reduction. In Hale, B. and Hoffmann, A., editors, *Modality, Metaphysics, Logic, and Epistemology*, pages 109–136. Oxford University Press, Oxford.

Schaffer, J. (2009). On what grounds what. In Manley, D., Chalmers, D., and Wasserman, R., editors, *Metametaphysics: New Essays on the Foundations of Ontology*, pages 347–383. Oxford University Press, Oxford, UK.

Schaffer, J. (2012). Grounding, transitivity, and contrastivity. In Correia, F. and Schneider, B., editors, *Metaphysical Grounding: Understanding the Structure of Reality*, pages 122–138. Cambridge University Press, Oxford, UK.

Schaffer, J. (2015). Grounding in the image of causation. *Philosophical Studies*, Online First.

Schelling, T. C. (1960). *The Strategy of Conflict*. Harvard University Press, Cambridge, MA.

Schiffer, S. (1972). *Meaning*. Oxford University Press, Oxford.

Schiffer, S. (1993). Actual-language relations. *Philosophical Perspectives*, 7:231–258.

Schiffer, S. (2006). Two perspectives on knowledge of language. *Philosophical Perspectives*, 16:275–287.

Skiles, A. (2015). Against grounding necessitarianism. *Erkenntnis*, 80(4):717–751.

Skyrms, B. (2010). *Signals*. Oxford University Press.

Sperber, D. and Wilson, D. (1995). *Relevance: Communication and Cognition*. Blackwell, Oxford, 2 edition.

Sperber, D. and Wilson, D. (2002). Pragmatics, modularity, and mindreading. *Mind and Language*, 17:3–23.

Stalnaker, R. (1984). *Inquiry*. MIT Press, Cambridge, Mass.

Stalnaker, R. (2014). *Context*. Context and Content. Oxford University Press, Oxford.

Stich, S. P. (1978). Belief and subdoxastic states. *Philosophy of Science*, 45(4):499–518.

Stotts, M. (2015). *Conventions and Expression Meaning*. PhD thesis, University of California, Riverside, Riverside, CA.

Stotts, M. H. (2016). Walking the tightrope: unrecognized conventions and arbitrariness. *Inquiry*, published online.

Trogdon, K. (2013). Grounding: Necessary or contingent. *Pacific Philosophical Quarterly*, 94:465–85.

Williams, B. (2002). *Truth and Truthfulness*. Princeton University Press.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford University Press.

Wilson, D. and Sperber, D. (2002). Truthfulness and relevance. *Mind*, 111(443):583–632.