

ARTICLE

## Special Issue: Canadian Philosophical Association 2024 Prize Winning Papers

This Paper Won a Student Essay in English Prize and a Congress Graduate Merit Award (CGMA) Prize at the 2024 Canadian Philosophical Association Conference

# New Solutions, Old Problems: Agreement and Novelty in Dynamic Conventions

Nathan Malcomson

Department of Philosophy, York University, Toronto, ON, Canada  
Email: [nathanmalcomson88@gmail.com](mailto:nathanmalcomson88@gmail.com)

### Abstract

Social conventionalism — the position that social conventions governing language use determine or constitute the meanings of our words — has faced two major problems. The first is the Agreement Problem: how could speakers agree to use words in certain ways without already speaking meaningfully? The second is the Novelty Problem: how can conventions fix the meanings of innovative uses of words? David Lewis famously responded to the Agreement Problem but his account flounders on the Novelty Problem. Josh Armstrong emends Lewis' account to solve the Novelty Problem. I argue that Armstrong's emendation fails and that neither he nor Lewis has an adequate response to the Agreement Problem.

### Résumé

Le conventionnalisme social — selon lequel les conventions sociales régissant l'usage de la langue déterminent ou constituent le sens de nos mots — s'est heurté à deux problèmes majeurs. Le premier est le problème de l'accord : comment des locuteurs peuvent-ils accepter d'utiliser des mots d'une certaine manière s'ils ne parlent pas déjà de manière significative ? Le deuxième est le problème de la nouveauté : comment les conventions peuvent-elles fixer le sens des usages innovants des mots ? La solution de David Lewis au problème de l'accord est bien connue, mais son analyse ne résout pas le problème de la nouveauté. Josh Armstrong modifie cette analyse afin de résoudre ce dernier. Je soutiens que la correction d'Armstrong échoue et que ni lui ni Lewis n'offrent de réponse adéquate au problème de l'accord.

**Keywords:** linguistic conventions; the Agreement Problem; the Novelty Problem; dynamic conventions; communication; Josh Armstrong; David Lewis

© The Author(s), 2025. Published by Cambridge University Press on behalf of the Canadian Philosophical Association / Publié par Cambridge University Press au nom de l'Association canadienne de philosophie. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

## 1. Introduction

As Bertrand Russell observes, “[i]t is natural to think of the meaning of a word as something conventional” (Russell, 1921/1968, p. 189). Indeed, the idea that the meanings of a speaker’s words are fixed by the conventions of their linguistic community seems to be how many non-philosophers think of meaning.<sup>1</sup> In the philosophy of language, the position that conventions fix the meanings of our words — that the conventions in our linguistic community governing the use of a word, ‘p,’ determine what we mean by ‘p’ — and therefore that they must play an essential role in an account of meaning is called ‘social conventionalism.’ The opposing view, that conventions do not fix the meanings of our words, and thus that they have no essential role in a philosophical account of meaning, is called ‘anti-conventionalism.’ Despite enjoying some *prima facie* plausibility, social conventionalism was mostly discredited among philosophers in the early 20<sup>th</sup> century, in large part due to a problem famously expressed by Russell — what I will call the ‘Agreement Problem.’<sup>2</sup> However, social conventionalism was later rehabilitated by David Lewis in his book-length treatment of the position, *Convention* (Lewis, 1969/2002). Many philosophers took Lewis to have adequately responded to Russell’s concern, making social conventionalism a plausible view again, but a new problem cropped up for Lewis’ kind of social conventionalism, most forcefully expressed by Donald Davidson (1986/2005a) — what I will call the ‘Novelty Problem.’ Since Lewis then, social conventionalists have turned their focus from the Agreement Problem to the Novelty Problem, though their strategy has often been to argue that the Novelty Problem is not, in fact, a serious problem for social conventionalism.<sup>3</sup> It is thus significant that Josh Armstrong, a social conventionalist, takes the Novelty Problem seriously. In response to it, Armstrong has suggested a substantial emendation to Lewis’ account, which — if it goes through — provides social conventionalism with all of the tools it needs to address the Novelty Problem head-on (see Armstrong, 2016). In what follows, I contend that Armstrong’s emendation of Lewis’ social conventionalism does not work; the account Armstrong develops cannot overcome the Novelty Problem and ultimately falls to the Agreement Problem as well. More importantly, though, through Armstrong’s extension of Lewis, we can see that *Lewis* never actually succeeded in responding to the Agreement Problem. Thus, Lewis’ brand of social conventionalism flounders on two foundational problems, leaving us with no good reason to be social conventionalists.

## 2. The Agreement Problem

In short, the Agreement Problem is how one could come to participate in a linguistic convention without already having a language. Russell expressed the problem as follows:

<sup>1</sup> See Ludlow (2014, p. 2) and Quine (1969/2002, p. xi) for similar analyses of how non-philosophers generally conceive of language.

<sup>2</sup> I do not think that this problem *originated* with Russell. At the very least, Armstrong finds an early version of it in Rousseau (1755/1984; as cited in Armstrong, 2016, p. 86). However, Russell’s pithy expression of the problem is the version that is most often quoted in these debates (along with Quine’s later expression of it).

<sup>3</sup> See Camp (2016) and Lepore and Stone (2017) for two recent cases of social conventionalists arguing that the Novelty Problem does not pose a significant problem for their view.

We can hardly suppose a parliament of hitherto speechless elders meeting together and agreeing to call a cow a cow and a wolf a wolf.<sup>4</sup> (Russell, 1921/1968, p. 190)

The key here is Russell's use of the concept of *agreement*. An agreed upon convention could constitute or establish many social activities, but not language — for, how could a group of people agree to use words in certain ways if they do not have any words with which to verbalize such agreement? W. V. Quine later expressed the problem as follows: “What is convention when there can be no thought of convening?” (Quine, 1969/2002, p. xi). As long as we operate with a conception of conventions that relies upon the antecedent existence of linguistic communication for a convention to become established, then the Agreement Problem is crippling to any attempt to argue that conventions fix the meanings of our utterances.<sup>5</sup>

### 3. Lewis' Solution

Lewis' debate-shifting move was to provide an account of conventions that does not rely on the notion of agreement to explain how conventions get established. Instead, Lewis builds his account within a framework taken from decision theory. He recasts social conventions as widely expected solutions to coordination problems. Coordination problems are, roughly, social situations where a group of agents each want to reach the same solution to a shared problem but where there are multiple equally good solutions available; so, the focus of each agent is upon coordinating their behaviour with their fellows so all arrive at the same solution, rather than upon achieving any one particular solution. We can arrive at a solution to a coordination problem in a variety of ways, such as by predicting each other's beliefs and desires or by sheer luck. However, relying on sheer luck is an undependable strategy and it is thought that mind-reading is too cognitively demanding to do all the time, so rational agents look for reliable, efficient ways to arrive at solutions to recurring coordination problems. If two agents have already solved some kind of coordination problem once before, they will be able to look back to that first resolution in any future encounters with the same problem, and what they did in the first situation will stand out as salient. Precedents set by previous solutions to coordination problems thus pave the way for agents to use the same methods to seek the same kind of solution in subsequent coordination problems of the same kind.<sup>6</sup> When most members of a community adopt an arbitrary solution to a recurring coordination problem and expect others to do so, then a convention has formed.<sup>7</sup> Thus, Lewis' account appears

---

<sup>4</sup> This quote is one of the instances where Russell ignores the use/mention distinction. So, to be clear, the end of the quote should read: “to call a cow a ‘cow’ and a wolf a ‘wolf.’”

<sup>5</sup> The seriousness of the Agreement Problem is explicitly acknowledged by Armstrong (2016, p. 95) and seems to be something that Lewis was aware of as well (see Lewis, 1969/2002, pp. 2, 87–88).

<sup>6</sup> See Lewis (1969/2002, pp. 35–42) for his discussion of this feature of his account.

<sup>7</sup> Lewis' official list of necessary and sufficient conditions for conventions is much more involved. See Lewis (1969/2002, pp. 78) for his “final definition” of conventions. However, Lewis also provides briefer explanations of conventions throughout his work, such as “Convention turns out to be a general sense of common interest; which sense all the members of the society express to one another, and which induces

to explain what constitutes conventions and how they become established in a community without relying on explicit verbal agreements between agents.

To apply the above general account of conventions to language, Lewis begins by trying to explain how non-linguistic signalling systems arise as conventional solutions to problems of communication. Crucially, he thinks that he can tell this story without appealing to a signal's meaning (see Lewis, 1969/2002, p. 122). Lewis believes that a communicative coordination problem arises when a communicating agent wants to get another agent to do some action by the receiver's recognition of the communicating agent's intention in making some signal. Lewis' flagship case to illustrate how this works involves a sexton trying to communicate with Paul Revere, who is some distance away. As Lewis describes things, Revere expects a signal from the sexton (see Lewis, 1969/2002, p. 125); the signal will either convey that the British are coming by land, that the British are coming by sea, or that the British are not coming.<sup>8</sup> There are all sorts of different signals that the sexton could use to convey the relevant state of affairs that ends up holding, and neither he nor Revere cares which signalling system is used, so long as they each understand the given signals in the same way. Once the sexton and Revere coordinate on a particular mapping of states of affairs to signals, then that signalling system will be salient in future cases of the same communicative coordination problem, and as it comes to be used again and expected to be so used by the various parties, then we can say that the content of the signals in that system becomes fixed by the fact that it is expected that each signal will be "conventionally associated with some readily observable state of affairs, or with some definite responsive action, or with both" (Lewis, 1969/2002, pp. 142–143).

Lewis extends the same strategy to explain how words get their meanings, as he thinks that linguistic communication is simply an expansion of our capacity for non-linguistic signalling: "our hypothetical verbal signalers do not do anything we do not do. We just do more. Their use of language duplicates a fragment of ours" (Lewis, 1969/2002, p. 143). His idea seems to be that linguistic signs, such as words or sentences, are initially connected to representational states of mind by some agent who uses that sign to convey to another agent the representational content they have in mind.<sup>9</sup> That sign will then become salient in other cases where one wants to convey that same content. As more speakers come to use the same sign to express the same representational content and expect each other to so use the sign, the sign will come to be understood as *meaning* whatever content it is generally used to express. Eventually, a community of speakers will come to use a stable set of signs to communicate, and this set constitutes the lexicon of that community's language, fixing the meanings of any utterance made in that language (see Lewis, 1969/2002, p. 198).

---

them to regulate their conduct by certain rules" (Lewis, 1969/2002, pp. 3–4). For my purposes, the specific details of his more extensive definition are not necessary.

<sup>8</sup> Really, the signals convey that Revere ought to prepare *as if* the British are coming by land, sea, or not at all. The signal conveys a prescription. This is important only because it helps to illustrate that the contents of non-linguistic signals are inherently coarse-grained. See Lewis (1969/2002, pp. 143–145).

<sup>9</sup> Lewis skips steps in his account of linguistic systems, as I will argue in later sections. Armstrong gives the most natural elucidation of what the Lewisian position would be on the formation of linguistic conventions, and so I follow his account in my description of Lewis here (see Armstrong, 2016, pp. 82–86).

#### 4. The Novelty Problem

Despite various contentions with the details of his account,<sup>10</sup> many philosophers take Lewis to have adequately responded to the Agreement Problem, as he presents a conception of what constitutes a convention that doesn't rely on explicit verbal agreement between the agents participating in the convention. However, a new problem arises for Lewis' account, as it is committed to the implausible conclusion that the lexicon of a community must be relatively *static* over time. If the meanings of our utterances are fixed by conventions, and conventions are formed gradually as more agents come to expect the same signs to be connected to the same representational contents, then new or innovative uses of words will take time (and repetition across a community<sup>11</sup>) to become meaningful. However, such a view on how the meanings of our words get fixed flounders when it comes to explaining how successful communication between agents occurs. For, it fails to explain how agents understand one another when they use novel words or old words in novel ways, both of which are common phenomena among language users. This problem was developed by Davidson (1984, 1986/2005a, 1994/2005b), who argues that since novel and non-standard uses of words are ubiquitous in daily life, and are often understood perfectly well, linguistic conventions are neither necessary nor sufficient for successful linguistic communication. Ernie Lepore and Matthew Stone have nicely expressed the problem with a static view of language as follows:

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. (Lepore & Stone, 2015, p. 321)

This is the Novelty Problem; in short, how can social conventionalists account for novel, creative, or malapropic uses of language when — by definition — they don't follow an established convention of use? This is where Armstrong comes in.

#### 5. Armstrong's Solution

Armstrong agrees with the above assessment of the Novelty Problem:

I agree with Davidson that, insofar as philosophers of language have labored under this static assumption, their accounts of language are seriously threatened by facts about innovation and language change. (Armstrong, 2016, p. 95)

However, Armstrong does not think that a static conception of language is thrust upon the social conventionalist. He argues that social conventionalism can allow

---

<sup>10</sup> See Burge (1975), Gilbert (1983), and Millikan (1998) for complaints that have less to do with the Agreement or Novelty Problems and more to do with the details of Lewis' necessary and sufficient conditions for the formation of a convention.

<sup>11</sup> Note that this issue arises even if we restrict the relevant community to a few members, or even two members.

for facts about linguistic innovation by viewing linguistic conventions as *dynamic*, as being able to “develop on the fly without necessarily being established in advance” (Armstrong, 2016, p. 96). Thus, to avoid the Novelty Problem, Armstrong redefines what constitutes a linguistic convention, writing:

the claim is that a group of agents have established a convention if, having found an arbitrary solution to a coordination problem, the members of the group are committed (and expect that the other members of the group are likewise committed) to reverting to that solution should that problem reoccur. (Armstrong, 2016, pp. 96–97)

And again:

a group of agents have established a novel semantic convention if they successfully coordinate on a pair of rules mapping representational states of mind to signs and signs to representational states of mind, and expect that those same rules will be utilized again if the sign comes to be redeployed. (Armstrong, 2016, p. 97)

These redefinitions of linguistic conventions rely on Lewis’ later claim (Lewis, 1975) that it is not how social regularities *begin* but how they *persist* that makes them conventional.

Armstrong’s response to the Novelty Problem is thus that neologisms and innovative uses of words and the like are governed by conventions because they immediately generate expectations among speakers on the future use of the relevant word (expectations that fix what those words mean going forward). To illustrate his point, Armstrong provides the example of an agent who encounters a stranger walking their dog, and the dog-walker utters ‘Mupsy loves to greet strangers with a lick.’ We expect our agent to immediately grasp that ‘Mupsy’ refers to the dog, and not, say, the dog’s owner, despite never encountering the name ‘Mupsy’ nor the dog Mupsy before — and this fact is taken by many to be an indictment of social conventionalism, showing that conventions play no essential role in an account of successful communication. However, Armstrong analyzes the case differently:

Such a case is no threat to conventionalism, for the conventionalist can maintain that, while speaker and audience didn’t have common knowledge of the conventions governing “Mupsy” prior to the speaker’s utterance, the speaker’s utterance served to transmit common knowledge of the relevant linguistic conventions to the audience. (Armstrong, 2016, p. 96)

This should give us a sense of just how dynamic Armstrong’s dynamic conventions really are; he is clear that “[a] *single use of a novel expression in a linguistic utterance is [...] sufficient to establish an entirely new linguistic convention among the group of agents engaged in a communicative exchange*” (Armstrong, 2016, p. 96, emphasis added). Mere *understanding* passing between speakers is sufficient to establish a linguistic convention. Armstrong’s dynamic conventionalism locates the conventionality of meaning in an agent’s commitments to revert to the same solution to future

coordination problems of the same kind, that is, to commitments to follow the same rules in connecting signs to representational states of mind in future uses (and interpretations) of those signs (see Armstrong, 2016, p. 97). Armstrong differs from Lewis, however, in maintaining that such commitments are generated on the fly, after a single use of or encounter with a word.<sup>12</sup>

## 6. Dynamic Word Use and Linguistic Commitments

Key to Armstrong's re-reading of Lewis are the expectations that uses of words (particularly novel or innovative uses of words) generate among speakers. However, he does not spell out the nature of these commitments in the way one would like. Armstrong initially sticks close to the concepts and framework operative in *Convention* to redefine what constitutes a linguistic convention, and so the idea seems to be that in encountering a novel use of a word, speakers must take each other to be committed to using that word in the same way in all subsequent coordination problems of the same kind (see Armstrong, 2016, pp. 96–97). But, what is the kind of coordination problem that we are committing to solving in the same way going forward once we respond to the stranger, using the word 'Mupsy' for the first time? (Perhaps replying: "Thanks for the warning, it's nice to meet Mupsy.") Could the relevant coordination problem be that of trying to speak of the dog Mupsy? This would entail, on Armstrong's account, that in first using the word 'Mupsy' to refer to the stranger's dog, we must be committed to always referring to that dog, and no other, as 'Mupsy.' That can't be the case: we may well meet another dog named 'Mupsy' or come to use 'mupsy' as an unrelated adjective. Furthermore, we could easily use 'Smith's dog' or 'that dog' or 'my sworn nemesis' to refer to Mupsy or — importantly — we could call the dog by a new name each time we encounter the dog, out of sheer spite.

To move away from any implausibly restrictive description of the relevant coordination problem, one could suggest a disjunctive solution to coordination problems involving talking about the dog Mupsy. For example, in entering the convention concerning the dog Mupsy, I am committing to calling the dog 'Mupsy' or one of another of an appropriate set of names for or descriptions of the dog. But, this abandons what was right in Lewis' conception of conventions, as too many solutions to the relevant coordination problem are then available and so no one solution can reasonably be salient or mutually expected. Being committed to calling Mupsy by her name or some other name or description that picks her out is an empty commitment: it is simply too broad.

Armstrong is going to have a hard time outlining the relevant kind of coordination problem that we are apparently committing to solving by using 'Mupsy' to refer to Mupsy in all future cases. This is perhaps why Armstrong redescribes his account of linguistic conventions in non-Lewisian terms (that is, without reference to coordination problems, decision theory, etc.). He recasts his account of dynamic

---

<sup>12</sup> Armstrong thinks that his account is the *proper* reading of Lewis, i.e., that Lewis' account already allows for dynamic conventions (see Armstrong, 2016, p. 96). I disagree. I think Armstrong's account is certainly Lewisian, but I think it represents a substantive (and much-needed) *extension* of Lewis' account.

conventions as the broader observation that in using a sign to express a representational state of mind, speakers are committed to using that sign to express the same representational state of mind *should they come to use that sign again*, and hearers are committed, upon first mapping that sign to a representational state of mind, to understanding that sign to express that representational content should the speaker use that sign again (see Armstrong, 2016, p. 97).<sup>13</sup> In short, in *first* using the name ‘Mupsy’ to refer to Mupsy, I am committed to meaning *Mupsy* by ‘Mupsy’ in all future uses of the word ‘Mupsy.’

While recasting the nature of linguistic conventions in non-Lewisian terms allows Armstrong to avoid challenging questions about the correct descriptions of the relevant coordination problems and solutions, his recast conclusion is either empirically dubious or trivial, depending on how strongly it is cashed out. It should be clear that speakers can and do modulate or change what they mean by their words over the course of their lives, if not over the course of individual conversations. In his book-length treatment of meaning modulation (Ludlow, 2014), Peter Ludlow provides us with the example of the word ‘sandwich.’ We can think of all sorts of different things one might mean by the word ‘sandwich’ over the course of one’s life. Kids talk, non-metaphorically I think, of Oreos and S’mores as sandwiches, while by ‘sandwich,’ their parents often mean something more restrictive, like *food item consisting of meats and/or vegetables placed between two slices of bread*. Other times, people mean something fairly general by ‘sandwich,’ including hot dogs and lettuce wraps in its extension (but excluding Oreos). In one instance, the meaning of ‘sandwich’ was litigated in court, with one party presumably changing what they generally meant by the word at the end of the proceedings (see Ludlow, 2014, p. 10). It is just a small step from these observations to conceiving of a speaker who modulates what they mean by ‘sandwich’ depending on whether they are talking with a child, a waiter, or a judge. We could repeat this analysis with many other common terms (Ludlow also discusses ‘book,’ ‘doll,’ and ‘journalist,’ among other words).<sup>14</sup> Thus, it is perfectly coherent to believe (and not merely imagine, via a thought experiment) that speakers can change or modulate what they mean by a word across time and conversational contexts; it is similarly coherent to believe that speakers may shift back and forth between what they mean by their words (e.g., ‘philosophy’ or ‘decimate’) in different conversational contexts and perhaps even over the course of a single conversation. Furthermore, as malaprops make evident, speakers can mean things by their words that they do not intend to mean by them in the future. So, it would be wrong to state that in using a word, ‘p,’ to mean *p* for the first time, a speaker is committing to using ‘p’ to mean *p* in *all* future uses of ‘p’ or in *all* future cases where they want to convey *p*.

In response to clear instances of a speaker modulating what they mean by some word or changing what they mean by a word over time, the most plausible thing

<sup>13</sup> To be charitable, Lewis himself thinks that his account of conventions could be explained apart from the decision-theoretic framework (see Lewis, 1969/2002, p. 3). However, abandoning talk of coordination problems and solutions robs the Lewisian account of its most tractable tools.

<sup>14</sup> In addition to the stock of words discussed in Ludlow (2014), see Buchanan and Schiller (2021), Davidson (1986/2005a), Donnellan (1968), Recanati (2004), and Richard (2019) for discussions of other cases of meaning change/modulation.



Armstrong could do is weaken his conclusion to the more palatable claim that in using some word 'p' to mean *p*, speakers are committed to meaning *p* by 'p' in their future uses of 'p' *unless they have mitigating desires or reasons otherwise or there is some causal explanation for the change* (such as memory loss). This seems to be an accurate description of how speakers use their words, but it certainly does not amount to anything like the social conventionalist position that Lewis is concerned to defend and that Davidson is concerned to argue against. For, such commitments do not fix what one means by one's words. Armstrong admits that in instances where a speaker is using their words innovatively, the hearer's understanding of the speaker's utterance is *not* derived from any antecedent knowledge of conventions: "novel linguistic conventions are partly the result of successful linguistic communication, rather than the other way around" (Armstrong, 2016, p. 97). He explains that this is because "[w]hen agents communicate without conventions, they cannot utilize their prior experiences with one another to guide their signaling choices" (Armstrong, 2016, p. 97). Innovative uses of words are instances where knowledge of conventions does not yield the correct meaning of our interlocutor's utterance. But, if speakers are only committed to using their words to mean the same things in future uses of them unless they have mitigating reasons (or causes) otherwise, then *any* conversational context is potentially one where a speaker is using their words innovatively. From the interpreter's point of view (as Buchanan & Schiller, 2021 and Davidson, 1986/2005a have pointed out), *every* conventional or expected interpretation of a word is defeasible: there is always the possibility that the speaker across from them is now using their words in a non-standard, innovative, or malapropic way. This entails that conventions, cashed out as commitments to use our words in consistent ways over time, do not fix the meaning of any *given* utterance of a speaker in advance. Rather, it is the speaker's communicative intentions in making their utterance that determines how we ought to interpret their words in that moment.<sup>15</sup>

We might think that, since speakers *often* use words to mean the same things over time, relying on expectations based on inferences from past behaviour is going to *often* yield the correct interpretation of our interlocutor. Again, this much seems right, but it is not something that anti-conventionalists deny. Davidson, for one, maintains that, for pragmatic reasons, speakers often develop dispositions to use (at least many of) their words in the same ways over time and develop expectations (what Davidson calls a 'prior theory'<sup>16</sup>) about how their neighbours will use their words, all things considered.<sup>17</sup> The anti-conventionalist can agree that linguistic

<sup>15</sup> I do not intend this brief note on communicative intentions to be taken as a complete account of meaning. There is surely much more that needs to be said to present a full account of meaning and communication; I just want to be clear that conventions are not essential to such an account and that the role they are often forced into is more accurately fulfilled by a discussion of a speaker's communicative intentions.

<sup>16</sup> A prior theory, for Davidson, is a theory of meaning that a speaker brings into a conversation, that is, it is the lexicon (plus grammar) that a speaker is prepared to use to interpret their interlocutor. This differs from a 'passing theory,' which tracks how speakers *actually* use their words as the conversation flows, assigning meanings to words within the context of utterance. See Davidson (1986/2005a).

<sup>17</sup> "Of course I d[o] not deny that in practice people usually depend on a supply of words and syntactic devices which they have learned to employ in similar ways" (Davidson, 1994/2005b, p. 110) and "The less we know about the speaker, [...] the more nearly our prior theory will simply be the theory we expect someone who hears our unguarded speech to use" (Davidson, 1986/2005a, p. 103).

conventions serve as a useful crutch in communication because, on its own, such a conclusion does not entail anything substantive about the nature of meaning. Some anti-conventionalists have even gone further, arguing that speakers have good reason to use their words in ways that they expect their interlocutors to understand, and thus that speakers often have good reason to use their words as they have in the past or as their fellows have (see, e.g., Buchanan & Schiller, 2021, p. 66; Davidson, 1986/2005a, p. 98, 1994/2005b, p. 122). But, such an admission still does not vindicate Armstrong's thesis, for two reasons. The first is that for the anti-conventionalist, conventional meanings (viewed as an agent's expectations about a given interlocutor's uses of words or an agent's beliefs about the general linguistic tendencies of others in their community) are ostensibly theories agents bring to conversations concerning what their interlocutors might *now* mean by their words; theories which can be falsified, for any given use of a word, by how the speaker is actually using their words in that particular context. The anti-conventionalist holds that knowledge of linguistic conventions (cashed out in either the more communal, Lewisian way or the more interpersonal, Armstrongian way) is not sufficient for understanding what a given speaker means by their words — we must know, in rough, how it is they intend their words to be understood. The second reason the anti-conventionalist can allow that speakers often do speak as their fellows do (and have reason to do so), is that even still, it is the beliefs and communicative intentions of individual speakers that are doing the real work in our explanation of communication. The anti-conventionalist makes clear that speakers have reason to speak *as they believe* their fellows do; a speaker can have a mistaken belief of what most of their community means by a word. Thus, in communication, we should be after an understanding of what the speaker intends to mean by their words in the present context — not what those words are generally used to mean in their community nor even what the speaker meant by those words in past instances of use.

Insofar as Armstrong's dynamic conventionalism is plausible, it is too weak of a thesis to support social conventionalism, and insofar as the thesis supports an essential role for conventions in account of communication, it is implausible. Thus, positing dynamic linguistic conventions may help explain various features of language use, but it does not help the social conventionalist overcome the Novelty Problem.<sup>18</sup>

## 7. The Agreement Problem Revisited

In addition to the issues it encounters trying to overcome the Novelty Problem, Armstrong's account has a more fundamental flaw — it falls to the Agreement

---

<sup>18</sup> Indeed, see Ludlow (2014) for an approach to language similar to Armstrong's, but couched in what I take to be an anti-conventionalist framework. Like Armstrong, Ludlow hopes to "develop an alternative *dynamic* theory of the nature of language and the lexicon" (Ludlow, 2014, p. 2); however, Ludlow's dynamicity exceeds Armstrong's, for Ludlow ultimately thinks that "human languages are things that we build on a conversation-by-conversation basis" (Ludlow, 2014, p. 3). Ludlow also touches on cooperative coordination in conversation, using the notion of 'lexical entrainment,' but there is no sense in Ludlow that linguistic conventions *fix* what speakers mean; rather, they arise as speakers intentionally (and unintentionally) modulate their uses of words to achieve communicative success.

Problem. Armstrong needs linguistic conventions to be able to be established with one instance of coordination. But, to have a conventional meaning after one instance of some act, Armstrong must say that the initial, convention-initiating act is an act of meaning *p* by ‘*p*.’ This is because he requires an agent, after encountering a novel use of some word, to take the speaker to be committed to using that word in the same way in all future coordination problems of the relevant sort. This requires each speaker to take the initial use of the word as meaningful and to take it to be so *independently* of any established conventions for that word. But, this makes it clear that meaning must precede conventions: we must be able to produce and interpret meaningful utterances without the aid of conventions for linguistic conventions to get going.

As mentioned above, Armstrong acknowledges that his account cannot explain how speakers acquire a language or a new word:

It is undoubtedly true that the newly constructed linguistic conventions cannot serve as part of the explanation of how successful linguistic communication occurred in the first place. (Armstrong, 2016, p. 97)

This is a striking admission from a social conventionalist, as Armstrong seems to be admitting that his account cannot answer the Agreement Problem. Now, one might think that because Armstrong’s account still does not rely on *explicit* agreement between speakers, it avoids the Agreement Problem. But, this would miss the force of the problem. Take Quine’s formulation of the problem; Quine asks, “[w]hat is convention when there can be no thought of convening?” (Quine, 1969/2002, p. xi), making clear that the issue is not about explicit verbal agreements but the necessary priority of meaningful thought and communication over linguistic conventions. One *must* be able to speak meaningfully before one can conform one’s speech to some social convention. The question at hand is whether conventions fix the meanings of our words — if we mean things and can understand others *independently of any established convention*, then the answer to our question is a definitive “no.”

Armstrong — or someone sympathetic to his position — might object that I have characterized social conventionalism uncharitably, making it out to address a question (i.e., what fixes the meanings of our words) that it is not trying to answer. One might want to claim that social conventionalists are only explaining how linguistic conventions *do* arise (as a descriptive matter of fact). I will return to this reading of conventionalism at the end of the article. For now, it’s sufficient to see that this is *not* how social conventionalists frame their arguments, and this is *not* the thesis that anti-conventionalists are concerned with arguing against. Compare how Lewis frames the issue:

[C]onsider this argument, given by Quine and others. The first convention of language to be established could not originate by an agreement conducted in a convention-governed language. So even if *any* convention of language could originate by such agreement, not *all* of them could. [...] I offer this rejoinder: an agreement sufficient to create a convention need not be a transaction involving language or any other conventional activity. All it takes is an exchange

of manifestations of a propensity to conform to a regularity. (Lewis, 1969/2002, pp. 87–88)

I have been arguing that once we capture the conventional aspect [of communication], we are done. We have captured the intentional aspect as well. (Lewis, 1969/2002, p. 159)

In the first of the above quotes, Lewis cites, in a footnote, Russell’s expression of the Agreement Problem, showing that he has the problem — as expressed by Russell and Quine — in mind. Furthermore, in that same quote, Lewis makes clear that he is not merely arguing that some languages are governed by convention or that speakers abide by conventions once they are in a language, but that he is arguing the more substantive point that even the very first language could have come about by convention. When this conclusion is coupled with the content of the second of the above quotes — that the conventional aspect of language subsumes the intentional aspect of language — then we can see that Lewis is after an ambitious, substantive claim about the nature of linguistic communication and an answer to the question of what fixes the meanings of our utterances.

Next, consider how Armstrong characterizes his anti-conventionalist opponent, Davidson:

Davidson holds [that] social conventions facilitate language use in many practical respects; yet, *social conventions are in no sense an essential part of what natural languages are* or of explaining how natural languages support interpersonal communication. (Armstrong, 2016, p. 81, emphasis added)

I think Armstrong is right that what an anti-conventionalist like Davidson is concerned with is *the essence of natural languages*, and that the anti-conventionalist has nothing against the idea that “social conventions facilitate language use.” To argue that anti-conventionalism is wrong then, one is committed to defending the claim that conventions are an essential part of natural languages (or rather, the meaningful use of words). By assuming that speakers can mean things by their words apart from conventions — that they *must be able to do so* to initiate linguistic conventions — Armstrong is already departing from Lewis and, I think, ceding substantive ground to the anti-conventionalist position.

## 8. Lewis and the Agreement Problem

Given that Armstrong’s extension of Lewis falls to both the Agreement and Novelty Problems, should we then return to Lewis, assuming that he at least taught us how the social conventionalist could overcome the Agreement Problem? I think not. I think that in examining Armstrong’s treatment (or lack thereof) of the Agreement Problem, we get a glimmer of the same sort of problem in Lewis, despite Lewis’ express intention to overcome the Agreement Problem.

When Lewis opens *Convention*, he casts his project as one of explaining how conventions could arise without any explicit verbal agreement between speakers.

Regardless of whether one agrees with the historical accuracy of some of Lewis' examples, he presents a very compelling case for understanding how conventions such as 'walk on the right-hand side of the road' or 'if a call gets dropped, the older of the two callers will call back' *could* be formed without any explicit agreement. But, note that even in his most basic cases, Lewis assumes that his interacting agents are discursive creatures. Consider how Lewis supposes agents will converge on a convention in the simplest of cases (from the first chapter of his book):

We must each choose where to go. The best place for me to go is the place where you will go, so I try to figure out where you will go and to go there myself. You do the same. Each chooses according to his expectation of the other's choice. (Lewis, 1969/2002, p. 5)

So each must choose whether to drive in the left lane or in the right, according to his expectations about the others: to drive in the left lane if most or all of the others do, to drive in the right lane if most or all of the others do (and to drive where he pleases if the others are more or less equally divided). (Lewis, 1969/2002, p. 6)

In these instances, and the others Lewis raises (see Lewis, 1969/2002, pp. 1–8), coordinating agents are solving their coordination problems using their beliefs about the intentions of their fellows. It seems, then, that in each of Lewis' first 11 test cases, his agents are concept mongers already, thinking of the world, their own intentional actions, and the beliefs of their fellows under certain descriptions. Thus, language is already on the scene in the first pages of Lewis' book.

At times, Lewis claims that the decision-theoretic aspect of his account can fall by the wayside; it is only an explanatory tool for him (Lewis, 1969/2002, p. 3). One might thus think that if we gave up the decision theory aspect of Lewis' account of conventions, then we could avoid the charge that conformity to conventions requires the possession of beliefs about the world and the actions and beliefs of others. I am dubious of such a proposal — not because non-discursive beings cannot converge on regular, arbitrary, behaviours, but because the repeated, arbitrary behaviours of agents who are not acting under any description are only conventions in the loosest of senses. What Lewis gets right is that even if explicit verbal agreement isn't needed to establish a convention, agents participating in a convention should be able to *make explicit* the shape of the convention in which they are participating:

We must have evidence from which we could reach the conclusion that any of our conventions meets the defining conditions for a convention, but we may not have done the reasoning to reach the conclusion. (Lewis, 1969/2002, p. 63)

How are we to describe the convention of a group of non-discursive beings if they, by definition, do not have propositional beliefs about their actions? It would seem that no one description of their behaviour can be picked out as the right one, the one that captures the convention operative in the group. Lewis allows that the specifics of some conventions (e.g., the exact rate at which two rowers row) might not be epistemically

available to the participating agents but there is still the assumption that the participating agents are both acting under the same description in *some* sense (e.g., they each intend to ‘row at the same rate as my fellow’).<sup>19</sup> Even if the rowers happened to be rowing to the beat of ‘God save the King,’ if neither were aware of this (at all), then it would be wrong to suggest that the rowers have a convention of rowing to the beat of ‘God save the King.’<sup>20</sup>

To be clear, contra what others have suggested (see, e.g., Millikan, 1998), the cognitive demands that Lewis’ account of conventions places on agents participating in a convention are not prohibitive of it being a good account of the nature of conventions, but they do pose a problem for his extension of it to language. Consider Lewis’ discussion of signalling systems. He begins by framing his goal for his chapter on signalling systems quite strongly, as being that of exploring those conventions “whereby we give to suitable actions the status of signals” (Lewis, 1969/2002, p. 122). This makes it seem that apart from conventions, those same actions would not achieve the status of signals. However, immediately, in his famous thought experiment concerning Revere and the sexton, Lewis writes:

*In this case, the sexton and Paul Revere agreed upon signals for a single occasion. In other cases, the same signaling system — preeminently analogous coordination equilibrium combinations of contingency plans for a communicator and an audience — may occur repeatedly, without need for fresh agreement every time. (Lewis, 1969/2002, p. 125, emphasis added)*

In the context of his larger project, this is a very perplexing admission — this is precisely what Russell and Quine were concerned with showing: namely, agreement is needed to get linguistic conventions off the ground. This admission is not unique to the Revere case either: Lewis goes on to provide five more exemplars of signalling conventions and, in most of them, the relevant signals arrive on the scene already endowed with meaning. Concerning his first example, the International Code of Signals, Lewis admits that “[i]t originated by explicit agreement among representatives of some of those now party to it” (Lewis, 1969/2002, p. 126). Lewis’ fourth and fifth examples involve railway and road signs, respectively. But, of course, such signals are created by rational agents and their meanings are spelled out in operation manuals and laws upon the creation of the physical signals. Lewis’ third example involves hikers following the trail markings of a more experienced guide, presupposing at the start that the agents know that the markings are contentful/meaningful signals (with a small range of possible meanings) and not, say, mere scratches on a tree. (The issue of how a mere physical marking or behaviour comes to be taken as *a contentful signal* in the first place is not taken up by Lewis.) So, almost all of Lewis’ examples of signalling systems involve either explicit agreement between the participating parties or rational agents who (somehow) know in advance to take a given mark, etc., as *a signal*.<sup>21</sup>

<sup>19</sup> See Lewis (1969/2002, pp. 63–64) for this discussion.

<sup>20</sup> This, of course, echoes Anscombe’s discussion of intentional action and the descriptions under which agents act (see Anscombe, 1957/1965, pp. 37–47).

<sup>21</sup> Lewis’ second example of a conventional signalling system involves construction workers using hand gestures to direct the drivers of heavy machinery. As described by Lewis, I think that this example is

The presupposition of meaningful thought and talk is even more acute in Lewis' approach to natural languages. He begins his chapter on natural languages by positing "possible languages," which are fully *interpreted* languages (Lewis, 1969/2002, p. 162). With this framing, Lewis thus shifts the question from how conventions of language could arise in a community of agents that have no language nor propositional thought (Russell's and Quine's concern), to how a given natural language comes to be used in a community *over other possible languages* (see Lewis, 1969/2002, p. 195). Lewis' general answer to his new question is the following:

*L* is an *actual language* of *P* if and only if there prevails in *P* a convention of truthfulness in *L*, sustained by an interest in communication. (Lewis, 1969/2002, p. 194)

But, clearly, in such an answer, there is no hint as to how *L*'s achieving the status of being actual in *P* could happen before the agents in *P* have a language or can communicate meaningfully. Thus, by the time he gets to natural languages, Lewis has provided no explanation of how his account of the conventionality of language avoids the Agreement Problem. This is not to say that there is no possible way for a social conventionalist to address the Agreement Problem, but contrary to what many have thought, Lewis does not provide any viable answer to Russell's and Quine's concern.

## 9. Conclusion

While there is ample reason to believe that Lewis took himself to be responding to the Agreement Problem in *Convention*,<sup>22</sup> one might think that his account could be vindicated if only we gave up our concern with the Agreement Problem. If we focused instead on how linguistic conventions *do* arise in a community, and what makes language use conventional, then perhaps Lewis would have a viable account for us. But, with this suggestion, it is important to be clear about the role linguistic conventions are supposed to be playing in our account of meaning. If the suggestion is simply that linguistic conventions exist and play a (very) helpful role in communication, then the anti-conventionalist has no qualms with this suggestion; such a claim is less of a philosophical thesis about the nature of language and more of an anthropological (or perhaps psychological) thesis about how speakers go about communicating in this world.<sup>23</sup>

---

problematic. For, the hand gestures of construction workers are often less arbitrary than Lewis portrays them to be; they are usually chosen to reflect, in a more-or-less isomorphic fashion, distances, directions, speed, etc. So, construction worker hand signals do not meet all of the criteria of conventional behaviour. However, there is potential for this sort of example, as it presents a case where signals seem to arise spontaneously among a group of agents. But, Lewis does not spell out how the mere behaviours of non-discursive creatures could come to be understood as conveying particular meanings.

<sup>22</sup>In addition to the passages already quoted, there are others, such as: "Thus I have tried to answer Quine's and White's skeptical challenge" (Lewis, 1969/2002, p. 203).

<sup>23</sup>This is not to say that the claim is *uninteresting* or that anthropological theses are second-rate, but simply that we need to be clear that this sort of claim has nothing to say about what fixes the meanings

If the suggestion is that *once an agent is in a community, then the linguistic conventions of their community fix the content of their words*, we have a significant philosophical thesis on our hands. Indeed, regardless of what else he is committed to, Lewis is committed to at least this claim:

The language [one] is using depends on the conventions he is party to. (Lewis, 1969/2002, p. 208)

Here and elsewhere, Lewis seems to support the idea that once one is in a community of speakers, the meanings of one's utterances are fixed by the conventions governing the 'actual' language of that community. But, this is just the point at which Lewis' social conventionalism runs into the Novelty Problem. For, it is a fact that speakers successfully use words that have no conventional meaning (e.g., neologisms) and that speakers successfully use words in conflict with how they are typically used (e.g., malaprops and innovative uses of words). Furthermore, given the pervasiveness of dialects and idiolects, knowledge of the linguistic conventions governing the actual language, *L*, of one's community is insufficient to understand the utterances of others.

Towards the end of *Convention*, Lewis acknowledges some of the facts about language use that generate the Novelty Problem, ending his account of linguistic conventions with the following concessions:

Standardization for the sake of communication is a good thing, but not all-important. (Lewis, 1969/2002, p. 201)

I think we should conclude that a convention of truthfulness in a single possible language is a limiting case — never reached — of something else: a convention of truthfulness in whichever language we choose of a tight cluster of very similar possible languages. (Lewis, 1969/2002, p. 201)

[T]he less restrictive our permanent convention is, the less experience it takes to identify it, catch on, and begin to be party to it. (Lewis, 1969/2002, p. 202)

With these remarks, Lewis gets closer to describing how speakers actually use language, but at the expense of any essential explanatory role for conventions in our account of meaning. So, attempting to vindicate Lewis' account by downplaying the importance of the Agreement Problem does not work, as his account still fails to adequately address the Novelty Problem.

For the reasons I gave earlier, I do not think that Armstrong's recent account can save social conventionalism from the Novelty Problem (nor, for that matter, from the Agreement Problem). However, there is still much good we can draw from the accounts of Lewis and Armstrong. If we remove any metasemantic ambitions from Lewis' account, reading it as a descriptive claim about how linguistic conventions arise in a community of speakers (as suggested above), then I think that Lewis'

---

of our words and that it is thus wrong to offer up this descriptive claim as a rebuttal to anti-conventionalists such as Davidson.



treatment has much to commend it. He certainly moves us past naive conceptions of conventions as social contracts and shows how a linguistic convention could arise without the participants being explicitly cognizant of it. Lewis also gives us tools to explain how and why speakers would modulate the meanings of their words so as to be understood by the interlocutors across from them. Furthermore, Armstrong's contention that diachronic stability in our use of words is a critical aspect of human communication, and a pre-condition for innovative uses of words, seems broadly right. That is, despite no individual word having its meaning necessarily fixed in advance of an utterance, it does seem that a speaker who altered the meanings of *all* of their words (that is, their entire personal lexicon, not just their uttered words) in *each new context of utterance* is perhaps inconceivable — or, at least, it is dubious that they can mean things by their words (for Humpty Dumpty concerns<sup>24</sup>). So, there are important general lessons to draw from Armstrong's analysis of the diachronic stability of word use and Lewis' general account of conventions. But, these lessons are best placed in an *anti-conventionalist* account of meaning. Until the social conventionalist can provide answers to the Agreement Problem and the Novelty Problem, we have no good reason to think that social conventions fix the meanings of our words nor that they must play an *essential* role in our account of meaning. We should thus join Russell in asserting that “the basis of a language is not conventional, either from the point of view of the individual or from that of the community” (Russell, 1921/1968, p. 189).

**Acknowledgements.** I would like to thank Claudine Verheggen for many invaluable discussions on this material; the article was significantly improved by her searching questions. I would also like to thank the anonymous referee for *Dialogue* for their helpful notes and the audience at the University of Waterloo 2024 Graduate Conference for feedback on an earlier version of this article.

**Competing interests.** The author declares none.

## References

- Anscombe, G. E. M. (1965). *Intention* (second edition). Harvard University Press. (Original work published 1957.)
- Armstrong, J. (2016). Coordination, triangulation, and language use. *Inquiry*, 59(1), 80–112. <https://doi.org/10.1080/0020174X.2015.1115270>
- Buchanan, R., & Schiller, H. I. (2021). Pragmatic particularism. *Philosophical and Phenomenological Research*, 105(1), 62–78. <https://doi.org/10.1111/phpr.12801>
- Burge, T. (1975). On knowledge and convention. *The Philosophical Review*, 84(2), 249–255. <https://doi.org/10.2307/2183970>
- Camp, E. (2016). Convention's revenge: Davidson, derangement, and dormativity. *Inquiry*, 59(1), 113–138. <https://doi.org/10.1080/0020174X.2015.1115277>
- Davidson, D. (1984). Communication and convention. *Synthese*, 59(1), 3–17. <https://doi.org/10.1007/BF00873278>

---

<sup>24</sup> ‘Humpty Dumpty concerns’ are concerns that if we claim that the communicative intentions of speakers fix the meanings of their words, then we are committed to the idea that speakers can mean *whatever* they want by their words, *whenever* they want. Anti-conventionalists do not have any real issue with these concerns, for we can point to rational constraints on successful communication that prohibit infelicitous speakers such as Humpty Dumpty from meaning just *anything* by their words; for one such discussion of this, see Buchanan and Schiller (2021, pp. 65–67).

- Davidson, D. (2005a). A nice derangement of epitaphs. In D. Davidson, *Truth, language, and history* (pp. 89–107). Oxford University Press. (Original work published 1986.) <https://doi.org/10.1093/019823757X.003.0007>
- Davidson, D. (2005b). The social aspect of language. In D. Davidson, *Truth, language, and history* (pp. 109–126). Oxford University Press. (Original work published 1994.) <https://doi.org/10.1093/019823757X.003.0008>
- Donnellan, K. (1968). Putting Humpty together again. *The Philosophical Review*, 77(2), 203–215. <https://doi.org/10.2307/2183321>
- Gilbert, M. (1983). Notes on the concept of a social convention. *New Literary History*, 14(2), 225–251. <https://doi.org/10.2307/468683>
- Lepore, E., & Stone, M. (2015). David Lewis on convention. In B. Loewer & J. Schaffer (Eds.), *A companion to David Lewis* (pp. 315–327). Wiley-Blackwell. <https://doi.org/10.1002/9781118398593.ch20>
- Lepore, E., & Stone, M. (2017). Convention before communication. *Philosophical Perspectives*, 31(1), 245–265. <https://doi.org/10.1111/phpe.12103>
- Lewis, D. (1975). Languages and language. In K. Gunderson (Ed.), *Minnesota studies in the philosophy of science* (volume VII) (pp. 3–35). University of Minnesota Press. <https://hdl.handle.net/11299/185227>
- Lewis, D. (2002). *Convention: A philosophical study*. Blackwell Publishers. (Original work published 1969.)
- Ludlow, P. (2014). *Living words: Meaning Underdetermination and the Dynamic Lexicon*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198712053.001.0001>
- Millikan, R. G. (1998). Language conventions made simple. *The Journal of Philosophy*, 95(4), 161–180. <https://doi.org/10.2307/2564683>
- Quine, W. V. (2002). Foreword. In D. Lewis, *Convention: A philosophical study* (pp. xi–xii). Blackwell Publishing. (Original work published 1969.)
- Recanati, F. (2004). *Literal meaning*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511615382>
- Richard, M. (2019). *Meanings as species*. Oxford University Press. <https://doi.org/10.1093/oso/9780198842811.001.0001>
- Rousseau, J. (1984). *A discourse on inequality* (M. Cranston, Trans.). Penguin Books. (Original work published 1755.)
- Russell, B. (1968). *The analysis of mind*. George Allen & Unwin. (Original work published 1921.)

---

**Cite this article:** Malcomson, N. (2024). New Solutions, Old Problems: Agreement and Novelty in Dynamic Conventions. *Dialogue* 63(3), 415–432. <https://doi.org/10.1017/S0012217324000234>