Conventions and Change in Semantics

A Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Arts in the Graduate School of The Ohio State University

By

Giorgio Sbardolini, MA

Graduate Program in Linguistics

The Ohio State University

2019

Master’s Examination Committee:

Craige Roberts, Advisor
Ashwini Deo
Abstract

Meanings are in some sense conventional, but they change over time. If conventions of meaning are game-theoretic equilibria of signaling games, a diachronic study of meaning can be formulated in full generality by investigating how equilibria change over time.

Lewisian conventions, whether from a rational choice perspective or from an evolutionary perspective, are too resistant to change: conformity among coordinators drives a population to equilibrium, establishing (semantic) order spontaneously. Some variation may be introduced within Lewis’s framework by allowing for randomly occurring learning mistakes. However, there is a class of phenomena, characterized as semantic social change, on which conventions of meaning undergo change not by chance, but because of shifts in the attitudes of language users. Because of the agents’ ever-changing practical goals, new linguistic policies may be adopted. This is a central force driving semantic change, albeit somewhat overlooked: semantic social change is a reflection of social change that becomes visible in the lexicon.

In order to account for semantic social change, dynamic conventions may be defined, as a temporal generalization of Lewisian conventions. Agents who synchronically behave as coordinators, are diachronically in conflict with each others. A population of speakers is divided in different groups, or profiles, determined by different practices and goals. In particular, despite the existence of a Lewisian convention about
the meaning of a word, a small group of innovators might appear, who start a new convention for its use. The new convention may then spread to the rest of the population, provided the innovators are in a position to exercise enough societal pressure on linguistic peers who belong to their *social network*, to force them to comply. On the resulting model, linguistic agents are playing a game of conflict, in which the balance of competing forces moves a community of speakers from an equilibrium to the next.
Vita

April 21, 1988 .............................. Born - Monza, Italy

2010 ........................................B.A. Philosophy,
Università degli Studi di Milano

2012 ........................................M.A. Philosophy,
Università degli Studi di Milano

2013-present ............................. Graduate Teaching Associate,
Ohio State University.

Publications

Research Publications


Fields of Study

Studies in Linguistics: Prof. Craige Roberts
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Vita</td>
<td>iv</td>
</tr>
<tr>
<td>1. Conventions and Change in Semantics</td>
<td>1</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>2</td>
</tr>
<tr>
<td>2. Conventions and diachronic variation</td>
<td>6</td>
</tr>
<tr>
<td>2.1 Dimensions of social change</td>
<td>9</td>
</tr>
<tr>
<td>2.2 Grammatical change</td>
<td>12</td>
</tr>
<tr>
<td>3. Signaling games</td>
<td>14</td>
</tr>
<tr>
<td>3.1 Strategy profiles</td>
<td>20</td>
</tr>
<tr>
<td>3.2 Evolution and Drift</td>
<td>23</td>
</tr>
<tr>
<td>4. Dynamic conventions</td>
<td>29</td>
</tr>
<tr>
<td>4.1 Social networks</td>
<td>40</td>
</tr>
<tr>
<td>5. Concluding remarks</td>
<td>43</td>
</tr>
<tr>
<td>Bibliography</td>
<td>49</td>
</tr>
</tbody>
</table>
Chapter 1: Conventions and Change in Semantics

Suppose you and I want to meet once a week. At the beginning, perhaps by chance, we might see each other at the library cafeteria on a Thursday morning. It does not matter to us where and when we meet, but it does matter that we do. As the weeks go by, we keep meeting there and then, and thus our arrangement becomes conventional: we don’t have to promise to be there in seven days, and we just keep meeting unless the external circumstances get in the way (say, because the library is closed).

According to David Lewis (1969), semantic regularities may be regarded as more or less analogous to our meeting convention. If you and I want to communicate, it doesn’t matter to us how expressions are paired with meanings, so long as we use them in (roughly) the same way. Like meeting conventions, conventions of meaning are somewhat arbitrary.

The centrality of conventions in an account of communication has recently been reasserted by Lepore and Stone (2014), and lies at the basis of Lewis’s (1979) influential “scorekeeping” account of conversation. At the foundations of Pragmatic theory, some notion of conventional what is said is central to the interlocutors’ derivation of what is meant or implicated on a given occasion of use (Grice 1989). However, there is an important disanalogy between conventions of meaning and meeting conventions: words used to have meanings that they no longer have, and new meanings arise all
the time, regardless of the external circumstances, and despite the standing desire to communicate. Unlike meeting arrangements, language is naturally dynamic.

The purpose of the present work is twofold: first, to isolate an important way in which language is dynamic, that has to do with its social nature, which I will call semantic social change; secondly, to present a dynamic account of semantic conventions, which is more adequate in explaining semantic social change. Dynamic conventions rest on more realistic and plausible assumptions about competence and the nature of conventional communication systems.

1 Introduction

In Sudanese Arabic, the most widely used word to talk about uncircumsized women, *ghalfa*, is strongly pejorative, suggestive of promiscuity and sexual slavery. In an effort to eradicate the practice of female circumcision (female genital mutilation), UNICEF campaigned to appropriate *saleema*, “a word that means whole, healthy in body and mind, unharmed, intact, pristine, and untouched, in a God-given condition”:

1

> The broad objective is to change the way that people talk about female genital cutting by promoting wide usage of new positive terminology to describe the natural bodies of girls and women.

On an earlier convention, *saleema* used to mean “whole, healthy in body and mind, unharmed”, later, the word has come to mean “uncircumcised woman” in a positive light. The success of the UNICEF project depends on many factors, but recent studies support a cautious optimism (Johnson et al., 2018). In this case, change in word meaning is driven by a shift in the expectations of a community of Sudanese speakers,

> https://www.unicef.org/sudan/protection_6092.html
as a result of their increased awareness of the consequences of certain traditional practices.

Change in the meaning of *saleema*, in other words, is not precipitated by the external circumstances: it is not the case that the non-linguistic environment in Sudan changed first, and that speakers adjusted accordingly. At least in the intentions of UNICEF, semantic change itself would contribute to initiating a change in the social life of Sudanese communities. Indeed, conventions of meaning may change in ways that are intimately related to societal changes. As the case of *saleema* illustrates, the phenomenon of *social change in semantics* shows that conventions of meaning are dynamic: the relation between an expression and its meaning may be affected by a shift in the attitudes shared by a community of linguistic agents.

According to David Lewis (1969, 1972), conventional regularities, including semantic regularities, are sustained over time by a convergence of interests among agents. Our interest in meeting supports a convention to meet at the library cafeteria every week. Our interest in understanding each other supports a convention to use an expression $\psi$ with a meaning $p$. Crucially however, more might be at stake in using language than mutual understanding. Speakers might have multiple interests, defined by their different goals, which are typically extralinguistic: for example, the goal of some Sudanese speakers to eschew the practice of female genital mutilation. If mutual understanding were the only interest speakers had in using language, that was relevant for our understanding of communication, Sudanese speakers would continue to use *ghalfa* to talk about uncut girls—despite its negative implications, that’s the word we have been using, and with which communication succeeds: to speak otherwise is, almost certainly, just to be misunderstood.
Of course, it has always been clear that a narrow focus on an interest for mutual understanding, in investigating the conventionality of linguistic meaning, was an idealization. Lewis’s influential account idealizes too much though, and we can relax some of his assumptions. In particular, speakers have a multiplicity of interests, one of which is mutual understanding, while others are given by whatever practical goals they might have. On this natural generalization of Lewis’s picture, it’s competition that drives semantic evolution: semantic change of the social kind, as for *saleema*, is the result of a conflict-driven interaction among speakers with possibly misaligned interests: those who stick to the traditional use, and those who innovate.

Lewisian conventions are solutions to coordination games. Dynamic conventions will be defined as solutions to mixed-motive games, i.e. games in which players have multiple interests—such as the Prisoner’s Dilemma or the Stag Hunt. Importantly, an equilibrium in a mixed-motive game does not obtain in perpetuity, as it were, but naturally shifts to another equilibrium. On the account presented here, linguistic agents can still be regarded, at any given time, as playing a coordination game that defines the set of semantic rules for their language; diachronically however, they are playing a competitive game, in which the balance of forces among interest groups shifts a Lewisian convention to the next. My proposal is thus to treat semantic social change as a kind of social change, and to apply current theories of social change, in particular [Bicchieri (2006)](Bicchieri2006), to the study of language as a social phenomenon.

A multiplicity of interests entails that a population of speakers may be divided in different groups, defined by those interests. In the *saleema* case, the two groups are defined by different attitudes toward the practice of female genital mutilation (i.e. roughly, acceptance and rejection). Plausibly, innovative behavior initially appears
and takes hold of a small group of individuals: the innovation then spreads, under certain conditions. But why would one group prevail, shifting the population to a new equilibrium? Why not instead suppose, for example, that the population simply fragments into subcommunities, speaking different sublanguages of the initial language?

These questions are about the causal mechanism underlying change. On the conception of language use and semantic change presented here, neither an explicit agreement between parties, nor consent accorded on the basis of deliberation, nor conscious metalinguistic reflection, are required for the establishment of a semantic regularity: though possibly present, it would be implausible and psychologically unrealistic to maintain that such intellectually sophisticated and cognitively demanding processes necessarily play a causal role in semantic change. By and large, semantic change is not chosen or decided by speakers—despite what Humpty Dumpty says to Alice. Instead, change derives from the normative pressure exercised on speakers by social networks. Linguistic agents are embedded in a network of relations with their peers, some of whom they might imitate. Pressure for conformity does not require an agent’s conscious awareness of the process of change, nor cognitive capacities more sophisticated than a limited memory of past interactions and some sensitivity to feedback.

In the first part of the thesis, I will define and provide evidence for the phenomenon of semantic social change, present a brief review Lewisian conventions, and explain

2Lewis Carroll, Through the Looking Glass:

“I don’t know what you mean by ‘glory’,” Alice said. Humpty Dumpty smiled contemptuously. “Of course you don’t – till I tell you. I meant ‘there’s a nice knock-down argument for you!’ ” “But ‘glory’ doesn’t mean ‘a nice knock-down argument’,” Alice objected. “When I use a word,” Humpty Dumpty said, in rather a scornful tone, “it means just what I choose it to mean – neither more nor less.”
why social change is not accounted for under Lewis’s assumptions. In the second part, I will introduce dynamic conventions as solutions to mixed-motive games, and discuss the role of social networks. I will conclude with some remarks on the notion of competence the present model requires on speakers, and with some philosophical considerations on the notion of open texture and on the indeterminacy of meaning.

2 Conventions and diachronic variation

In each instance of change, at least two properties may be distinguished, that are assigned before and after the change. Aging is a type of change, and we assign the properties being young and being old to the same person before and after the change. Semantic change is no different.

Two alternative meanings are assigned to saleema before and after the change, i.e. “whole and healthy” and uncut girl. It does not matter that two alternative linguistic forms exist in the lexicon (i.e., ghalfa vs. saleema), for this is not in general the case. Consider some appropriated slurs. In the course of a televised debate during the 1968 presidential election, the GOP strategist William F. Buckley Jr. famously addressed his progressive counterpart Gore Vidal by saying:

Now listen, you queer, stop calling me a crypto-Nazi or I’ll sock you in your goddam face, and you’ll stay plastered

It is quite clear that Buckley’s intention was to offend Vidal, a homosexual, with the appellative you queer. Indeed, the scandal was sensational, issuing in legal controversies protracted for several years. Concerning queer, the OED reports:

Although originally chiefly derogatory, since the late 1980s it has been used as a neutral or positive term, originally by some homosexuals.
Appropriated slurs like *queer* used to carry a pejorative meaning, activating negative stereotypes, and are now neutral or even positive terms to designate their target group. In the case of *queer*, the change must have occurred between the late 60ies and the late 80ies. As in the *saleema* case, the attitudes of a particular group of linguistic agents (in this case, the slur’s target group) are particularly relevant for an explanation of the observed semantic variation. Unlike the *saleema*/*ghalfa* contrast, the alternative meanings of *queer* do not correspond to a lexical alternative between syntactically distinct forms.

Arguably, *queer* has the same extension on Buckley’s use as on our own, i.e. the set of homosexual people. According to most accounts of slurs, the derogatory component is separated from the semantic component ([Camp, 2013](#)), and so change in the *queer* case is about the non-semantic content of the word. Notice however that this is not the case with *saleema*, where a change in extension is noticeable: from having the set of “whole and healthy” things as its extension, the word came to have a set of women as extension.

Some instances of change in extension have been recently discussed in the philosophical literature. Consider the following passage from Herman Melville:

> I take the good old fashioned ground that the whale is a fish, and call upon holy Jonah to back me. This fundamental thing settled, the next point is, in what internal respect does the whale differ from other fish. Above, Linnaeus has given you those items. But in brief they are these: lungs and warm blood; whereas all other fish are lungless and cold blooded. ([Moby Dick, Chapter 32](#))

According to the speaker, Ishmael, whales are fish—it’s just that they have lungs and warm blood. As Chalmers (2011, 519) points out, Ishmael needn’t be ignorant of

---

3The importance of the attitudes of the targeted individuals in the appropriation of slurs is supported by much psychological evidence—cf. Galinski et al., 2013.
any biological facts: we might suppose he knows exactly which animals have which biological features. However, he uses the word in such a way that the extension of *fish* includes whales.

The passage suggests that at or before Melville’s time Ishmael’s use was more common than current use: perhaps, any animal that can be “fished” counted as fish. After all, whaling was an important industry at that time, only later to become gradually irrelevant. At that point, the practice of “fishing whales” no longer stood in the way of speakers adjusting their linguistic policies to a more accurate reflection of the scientific consensus (no one, I think, wants to appear ignorant for no reason). It’s plausible that such combination of factors conspired to, as it were, push whales outside the extension of *fish*.

As these remarks indicate, the scope of social change in semantics encompasses all convention-based types of content, whether this is at-issue (like the extension of a predicate), or not. The at-issue/not-at-issue distinction depends on whether a type of content contributes directly to addressing the relevant Question Under Discussion in the discourse (Roberts, 2012, Tonhauser et al., 2013). Thus, which type of content is undergoing change in a specific instance is to be decided on a case by case basis, given independently motivated structural constraints on the lexicon and on conversation (i.e., constraints such as compositionality and projectivity). For the cases discussed so far, meaning-shifts may be located in the at-issue component (for *fish* and *saleema*, where we observe a change in extension), or in the not-at-issue component (for *queer*). Furthermore, it may be a matter of controversy to decide among types of content: different theories of slurs draw the line between semantics and pragmatics in different places. For an account of semantic change, however, there is no need to take sides on
these questions: distinctions within meaning are downstream with respect to speakers’ actions, and there is no reason to suppose that semantic social change should track distinctions, such as between at-issue and not-at-issue content, that are “invisible” to speakers.

2.1 Dimensions of social change

Semantic change is not only a social phenomenon. A classic contribution to the discussion on conventionality and linguistic variation is Davidson’s work (1984, 1986, 1991). Davidson notices that we often do get the message across in spite of the lack of conventional patterns of use and interpretation, for example with the use of metaphors and malapropisms. According to Davidson, it follows that conventions aren’t necessary for communication. This argument is controversial, but regardless of its merits, Davidson’s criticism of Lewisian conventions is very different from the one advanced here: Davidson focuses on local instances of linguistic innovation, typically limited to one-shot exchanges. Semantic social change does not take place at the individual level, but at population level, and not in the span of a conversation but in evolutionary time.

The population-size and temporal scale of semantic social change may be vividly illustrated by the following example, though the previous cases lend themselves to the same point. Adam’s farewells to the angel Raphael, as narrated by John Milton, may sound quite odd to us:

Go heavenly Guest, Ethereal Messenger,
Sent from whose sovran goodness I adore.

4For a critical discussion of Davidson on semantic conventions, see Green (2001), Camp (2016), and especially Armstrong (2016), who points out that conventions can and should be understood dynamically in order to account for diachronic phenomena in semantics. I think this is absolutely right. A formal theory of dynamic conventions is presented below.
Gentle to me and affable hath been
Thy condescension, and shall be honour’d ever
With grateful Memorie ... (Paradise Lost, Book 7, 1283-1287)

How could someone’s condescension be gentle and affable? Why is Adam thanking Raphael for being condescending? The OED entry for *condescension*, quite unexpectedly perhaps, reads as follows:

Voluntary abnegation for the nonce of the privileges of a superior; affability to one’s inferiors, with courteous disregard of difference of rank or position.

Interestingly, this entry appears to have been written by James A. H. Murray at the end of the 19th century, and not revised since [Hanks, 2013: 161](#). Until then, condescension used to be a somewhat pleasant (“curteous”) type of behavior, one for which it might be appropriate to be thankful. Today, there is nothing courteous in being condescending. According to the Merriam-Webster, condescension is characterized as a “patronizing attitude”, with examples clearly indicating an offensive, haughty, and perhaps contemptuous demeanor on part of the condescender.

The word *condescension* and its cognates have thus undergone a shift in meaning. A compelling explanation is offered by Daniel Siegel in his study of *condescension* in Victorian English:

condescension was traditionally used as an argument for paternalism, a model of government in which the legitimacy of empowered groups rested on the ability and disposition of those groups to provide for the less fortunate. When in the early Victorian decades paternalism was eroded by such liberal formations as self-help, separate spheres, voting reform, and the deregulation of markets, condescension became an object of critique. It was reappropriated and disseminated by the liberal opponents of paternalism, in forms meant to expose its inherent deficiencies and danger. [Siegel, 2005: 395]
The history of the meaning of *condescension* is, in part, the history of liberalism in English-speaking societies: the history of democratization and of the rise of the middle class.

An explanatory account of semantic social change takes place at the level of populations, not of individual agents, and over decades, hence well beyond the temporal limits of a conversation. It is not, however, just the phenomena discussed by Davidsons that do not fit the proportions of semantic social change. The methodology of Gricean pragmatics is reliable and well understood, but its scope makes it ill-suited for the study of language evolution: consulting native speakers’ judgments about the use of a sentence in context is no guidance to understanding the history of meaning. Some remarks about methodology are in order.

I have provided some qualitative evidence for social change in semantics, with the examples of *saleema, queer, fish*, and *condescension*. More could be given, of course, but it is important to appreciate that speakers’ linguistic preferences are measurable quantitatively: for example, collocation data about a word in a lexical field (i.e., statistically significant co-occurrences of words in corpora) is an indicator of speakers’ patterns of use. For example, Buckley’s use of *queer* as a pejorative is suggested by its co-occurrence with clearly negative expressions (e.g., *goddam*). Quantitative evidence for semantic social change may be extracted from large datasets about actual usage, of the kind that has become increasingly available recently. Patrick Hanks presents some of the desiderata on evidence in lexicography:

> Word meaning is dynamic, but that does not mean that it cannot be measured. The way to measure it is to study large bodies of evidence, recording the ways in which words are used and the inferences that can be based on different patterns of usage. ... Discovering conventions requires searching for regular patterns of word use shared by different speakers
and writers; this is a sociolinguistic task, which can only be achieved by comparing large numbers of uses of each word in different texts. Such a task was not possible until the development of corpus linguistic technology in the closing decades of the twentieth century. (Hanks 2013, 3)

Semantic social change is a phenomenon of a different proportion than the understanding of speaker’s meaning, hence the methodology of traditional pragmatics is inappropriate for its study. There are, however, rigorous methods for the study of the lexicon, including diachronic studies. For limits of time and space, I leave the details for further work, and trust that the examples I presented should give sufficiently clear evidence about the phenomenon under discussion.

2.2 Grammatical change

Social changes are events such as the spread of vegetarianism in Europe and North America during the first decades of the 21st century, or the secularization of Québec during the Quiet Revolution. Semantic social change is a kind of social change. As such, it is relative to a time, a culture, and a group of people defined by shared attitudes. For condescension, for example, the target group are politically liberal speakers of English during the Victorian age.

Social-ladenness entails an apparent lack of cross-linguistic generality: there is no prediction that words similar to condescension in languages other than English should go through a parallel trajectory. Nevertheless, social change in semantics is a wholly general fact about language. That is, the dependency of meaning on the attitudes of a community of speakers is a wholly general dynamic of language evolution.

A large part of diachronic studies in semantics are not focused on societal changes, but on grammatical changes. Expressions are typically classified as functional (i.e.,
operators that take NP- or VP-meanings as arguments) and non-functional (i.e. predicates). Functional expressions are the most frequently occurring in corpora, at least if we are counting tokens, and the less abundant in a lexicon, for a lexicon is largely made up of nouns, adjectives, and verbs. The observation that functional and non-functional expressions pattern differently in time is familiar, as suggested by the “open” and “closed class” distinction. Open-class expressions are non-functional, and may be added at a fast pace to the lexicon of a language with no consequences for the grammar. Closed-class expressions are functional and there is no comparable freedom. The study of the semantic history of non-functional expressions has been relatively overlooked.

Concerning diachronic semantic studies of functional expressions, Deo distinguishes a structural and a dynamic component:

In explicating the structural component of semantic change, what is needed is a precise characterization of the logical and conceptual relations (i.e., the similarities and differences) between the meanings of the related functional categories, such that the transition from one meaning to another is rendered natural. The dynamic component, on the other hand, draws from theories of language use and language evolution to provide a plausible account of the recruitment of new functional exponents, their categoricalization, and subsequent generalization to a broader meaning under normal conditions of usage and transmission. (Deo, 2015a, 184-5)

The structural component consists at least of a description of “the logical and conceptual relations” of the two meanings (before and after the change). Changes in the meanings of functional expressions are driven (at least in part) by grammatical pressures, such as the competition between meanings on the same scale—see e.g. Deo (2015b) on the grammaticalization of imperfective meanings, Dahl (2001) on emphatic expressions, and Ahern and Clark (2017) on Jespersen’s cycle. These cases of semantic variation in the functional domain are cross-linguistically robust,
and therefore a grammatical explanation of the underlying phenomena is called for. The same is not true of semantic social change. Variation in this case is observed in the meaning of open class terms, and the underlying engine is not the grammar but the attitudes of a community of language users. Often, little can be said concerning logical relations between meanings in the non-functional category. We can therefore expect the structural component in a theory of this kind of change to be relatively thin. A detailed description of the dynamic component is the topic of §4.

3 Signaling games

I shall now review the coordination model of communication, introduced by David Lewis (1969; 1972). I’ll focus in particular on what kinds of diachronic variation the coordination model allows for. It will become apparent that certain assumptions made by Lewis are incompatible with an adequate account of semantic social change. Fortunately, these assumptions are idealizations: they are useful to simplify the discussion, but dropping them will lead to a more realistic model of linguistic interactions.

A Lewisian convention is a solution to the task of communicating by means of signals that have no pre-established or “intrinsic” meaning—no ‘natural meaning’ in the sense of Grice (1989). Formally, consider any scenario $F$ whose abstract features may be modeled as a coordination game: intuitively, all players win if and only if all do what others do. Lewisian conventions are solutions of coordination games. The concept of convention may be captured by the following definition (Bicchieri (2006, 31–38); Lewis’s original definition (1972 5-6) is partly different, for reasons I will discuss in the conclusion):
A behavioral regularity $R$ in a situation of type $F$, where $F$ is a coordination game, is a \textit{convention} in a population $P$ if and only if there is a large enough $A \subseteq P$ such that for any $a \in A$,

1. $a$ believes that $R$ applies to $F$ (\textit{Contingency condition}), and
2. $a$ has an interest in conformity to $R$ provided that $a$ believes that enough individuals in $P$ conform (\textit{Epistemic expectation}).

Conventions are regularities in behavior supported by: (1) the belief that they exist in the given circumstances—this we may call a contingency condition; and (2) a conditional interest in conformity: that is, a preference for conforming whenever one thinks that enough individuals conform—this is an epistemic expectation. How much is “enough” is a vague matter, some exceptions being tolerable. In the interesting case, there could be another behavioral regularity $R'$ besides $R$, which anybody could have had an interest in conforming to, had anybody thought that enough conformed to it. Thus conventions are not counterfactually robust: indeed, they are somewhat arbitrary.

In a simple coordination game between You and Myself, the distribution of payoffs may be represented as follows. (As customary, the first element of an ordered pair is the row player’s payoff; the restriction to 2 players with 2 strategies each is for simplicity.)

Coordination game in normal form:

<table>
<thead>
<tr>
<th></th>
<th>$y$</th>
<th>$x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>$(1, 1)$</td>
<td>$(0, 0)$</td>
</tr>
<tr>
<td>$x$</td>
<td>$(0, 0)$</td>
<td>$(1, 1)$</td>
</tr>
</tbody>
</table>

Payoffs are rewarded according to the players’ conditional interest in conformity: Myself and You both win if and only if both perform the same strategy (a full action plan is called a strategy). For example, in a meeting game, Myself and You win if both go to the library cafeteria, or both go to the main square. Either way, our interests
in meeting are fulfilled (interests are combinations of attitudes formally represented by payoffs). In other (more precise) words, \((x, x)\) and \((y, y)\) are the only strict Nash equilibria of the game, i.e. combinations of strategies such that any player is strictly worse off for unilaterally changing strategy. According to Lewis, conventions are strict Nash equilibria of coordination games.

With regard to semantics, let’s suppose that for each agent in a population \(P\) there is (at least) a binary alternative between strategies \(x\) and \(y\), i.e. using an expression \(\psi\) to mean \(i_1\) vs. using \(\psi\) to mean \(i_2\). For example, speakers may use \(fish\) to mean biological fish or to mean animal-for-fishing, or they may use \(queer\) with negative stereotyping or without. Communication succeeds if and only if all players use the same expressions with the same meanings, in which case they coordinate on one of the many possible strict Nash equilibria available.\(^5\)

Linguistic interactions are sequential, not simultaneous as in the simple coordination game above. A more realistic picture of linguistic action is given by signaling games, which are widely employed in game-theoretic studies of communication.\(^6\) The players are a Sender \(S\) and a Receiver \(R\). In the customary interpretation, \(S\) observes a certain state of the world, and sends a message. \(R\) observes the message but not the

\(^{5}\)I follow Lewis in taking the unit of significance to be the utterance. Word meanings are then calculated compositionally from the meanings of utterances, as determined by a convention. This derivation is not without complications, but I leave a discussion of compositionality on the side. See Scott-Phillips and Kirby (2010) and Franke (2016) for discussion.

\(^{6}\)Signaling games are widely used in game-theoretic pragmatics: a generalization of Gricean pragmatics in which strategic inferences about what interlocutors must have meant are studied as games. On the most natural interpretation of games in GTP, meanings are fixed throughout the interaction, and function as ‘focal points’ to facilitate coordination (Schelling 1960). This stipulation is explicitly made by Franke (2009, 46), and followed by others. On GTP, see Parikh (1991, 2000, 2007), Pinker et al. (2008), Franke et al. (2012), Asher and Lascarides (2013), De Jaegher and van Rooij (2014), and Benz et al. (2010). GTP is conceptually very different from the framework I’m discussing here, in which meanings are not fixed, but instead “built up” through the interaction. In this, I follow Skyrms (1996, 2010); Huttegger (2007); Wagner (2015), and others.
initial worldly state. A round of play is in two parts: First, S produces an utterance; secondly, upon receiving S’s utterance R performs an action, which is typically the formation of an interpretation. If R interprets S’s message with the same meaning as S assigned to it then, in a sense, communication has been successful: R has learned what the state of the world is from S, merely by linguistic means.

I shall depart slightly from the customary interpretation, because games will be here interpreted not as systems to transfer information, but as meaning-building interactions. To revive an old Wittgensteinian theme: the meaning of an expression is determined by speakers coordinating on its use. The intuitive picture is as follows: there is an alternative between meanings i that a sentence ψ might have; then, S encodes one such meaning in a message (an utterance); finally, R decodes a certain meaning from the observed message. Formally, a signaling game is a tuple:

\[ \langle P, J, M, A_{S,R}, U_{S,R} \rangle \]

- \( P = \{ S, R \} \) is a set of (at least) two roles, sender S and receiver R;
- \( J = \{ i_1, i_2, \ldots \} \) is a set of meanings, or interpretations of the messages;
- \( M = \{ m_1, m_2, \ldots \} \) is a set of messages;
- \( A_{S,R} \) are sender and receiver strategies, where \( A_S = \{ j : J \rightarrow M \} \) and \( A_R = \{ k : M \rightarrow J \} \);
- \( U_{S,R} \) are utility functions for S and R, defined as \( A_S \times A_R \rightarrow \{ 0, 1 \} \), assigning a payoff (either 0 or 1) to a combination of sender/receiver strategies.

Instead of a set of messages \( M \), we could alternatively define a set \( E = \{ f', f'', \ldots \} \) of codings, i.e. one-to-many relations on the domain of meanings \( J \) (so that the same
meaning may be encoded in more than one way), and let $M = E(J)$. Intuitively, messages are just encoded meanings. We shall then say that communication succeeds in case the decoded meaning is the same as the meaning the sender had encoded. This is a formal description of the familiar “code model” of communication (Lewis 1969). Meanings must somehow be encoded in messages in order to travel to a medium (e.g. air).

It is only required that there be at least two players, at least one sentence, and at least two meanings, hence four messages; other than this, numbers can be multiplied. Consider for example the sentence Raphael is condescending, and call it $\psi$. Let there be two meanings $i_1$ and $i_2$ that $\psi$ might have. S plays first and may produce one of two messages by an utterance of $\psi$, one on which Raphael is gentle and affable, and one on which Raphael is patronizing and contemptuous. The first message carries the meaning determined by Milton’s use, the second by our own.

R has an alternatives upon receiving S’s message: to interpret it as $i_1$ or to interpret it as $i_2$. Sender and Receiver understand each other only if the meaning encoded by S in an utterance of Raphael is condescending is the meaning decoded by R, regardless of the initial meaning. We can represent the signaling game as follows:

\footnote{It is not accurate to say that condescension (and cognates) are ambiguous: the alternative meanings are meant to represent change in discrete units, not an ambiguity account. This is obviously only a simplification for modeling purposes.}
Both players are confronted with a dependent choice: the sender’s action depends on which belief state initially obtains, the receiver’s action depends on which message was sent.

Notice that $S$ sends a message that encodes a meaning of $\psi$, but intuitively $S$ does not “manipulate” meanings: the initial branching represents alternative meanings of $\psi$. This is regardless of the meanings $\psi$ may encode, hence regardless of the messages $S$ may send. Hence $S$ may utter $\psi$ when in state $i_2$, even though $\psi$ encodes $i_1$ (or vice versa). The game has a general structure that can be applied to other scenarios. It’s important to recognize that there is a central decision problem for $R$: how should I interpret the message? and a central decision problem for $S$: which message should I send? Both problems are determined by the same set of meanings, namely $J$. Formally, a set of meanings is a question, which is salient in different ways to the two players because of their different roles. I shall discuss the role of questions in the structure of the game in more detail in the conclusion, where the connection with the literature on the QUD (Roberts, 2012) will help highlight the assumptions about competence that have been made.
For now, let us turn to a more pressing issue: How do players choose between alternatives?

### 3.1 Strategy profiles

Agents may be divided into two different *profiles*, defined by sets of strategies. In the game above, there are two (relevant) profiles: map to $i_1$, and map to $i_2$. In other words, agents may play odd-numbered strategies, i.e. $j_1/j_3$ as senders and $k_1/k_3$ as receivers, or even-numbered strategies, i.e. $j_2/j_4$ as senders and $k_2/k_4$ as receivers. Agents that map to $i_1$ play odd-numbered strategies, agents that map to $i_2$ play even-numbered strategies. (“Mixed” behavior is, of course, possible: that is, some players may move between odd- and even-numbered strategies as senders and as receivers, but since we imagine them taking turns at playing $S$ or $R$ with equal probabilities, there is no complication here to take into account.)

We can imagine that $i_1$ is Milton’s meaning of *Raphael is condescending*, and that $i_2$ is our own. In this case, it is plausible to assume that at some point around 1870, a large majority of agents play odd-numbered strategies. At the same time, there is also a minority of “liberal” players who play even-numbered strategies. Let $C_e$ and $I_e$ be functions from $e \in P = \{S, R\}$ to strategies, where $C_e$ maps a player to $i_1$, and $I_e$ maps a player to $i_2$. Players in the first group follow the majority, at least around 1870, and therefore are called Conformists. Players in the second group are the Innovators. Notice that at this point, the only asymmetry between profiles is that we imagine them to be distributed in different proportions in the population.
Importantly, payoffs are awarded to coordinators: communication is successful when conformist meets conformist or when innovator meets innovator, not otherwise. That is, two kinds of combinations of actions maximize payoffs: \( \langle C_S, C_R \rangle \in \{ \langle j_1, k_1 \rangle, \langle j_3, k_1 \rangle \} \) and \( \langle I_S, I_R \rangle \in \{ \langle j_2, k_4 \rangle, \langle j_4, k_4 \rangle \} \). Either way, \( R \) interprets condescension to have the meaning which \( S \) encoded. Depending on which of \( \langle C_S, C_R \rangle \) and \( \langle I_S, I_R \rangle \) obtains, if any, \( S \) and \( R \) coordinate on either \( i_1 \) or on \( i_2 \). On other combinations, \( \langle C_S, I_R \rangle \) and \( \langle I_S, C_R \rangle \), sender and receiver fail to understand each other.

Coordinators need not be inclined either to conformity or innovation: what makes me a coordinator is that I’d rather conform, but I may be indifferent as to what I conform to. There is thus a problem of symmetry-breaking: which meaning should we coordinate upon? A symmetry-breaker between alternatives may be constrained by the notion of risk, from the individual-level perspective of rational choice theory, or by the notion of evolutionary fitness, from the population-level perspective of evolutionary game theory. Both of these notions, risk and fitness, depend on the proportion of players’ profiles. For reasons mentioned above, the latter approach is more adequate for the phenomenon under discussion, but I shall begin by briefly describing the first.

From a rational choice perspective, a rational speakers’ answer to the question With which meaning should I use \( \psi \)?, is one that maximizes expected utility. I assume for simplicity that players’ epistemic expectations roughly match the real frequencies of strategies in the population: that is, players are not making (huge) mistakes in estimating the relative proportions of different profiles. (As a matter of fact this may of course happen: linguistic agents may be under all sorts of misconceptions about the society they live in.) If \( \rho_{C_R} \) is the probability that \( R \) conforms to majority (and
thus, $1 - \rho_{CR}$ is the probability that $R$ is an innovator), the expected utility of a $C$ strategy for $S$ given an initial meaning is:

$$EU(C_S) = \rho_{CR} \times U_S(C_S, C_R) + (1 - \rho_{CR}) \times U_S(C_S, I_R)$$

where utilities are defined as above, so the expected utility for a sender to play $C$ is the weighted sum of the utility of playing $C$ against a conformist, and the utility of playing $C$ against an innovator. Note that the payoffs are arbitrary and can be scaled uniformly without changing the structure of the game.

The same holds for $R$, and for innovators, so we can generalize. Let $e \in P = \{S, R\}$ and $x \in \{C_e, I_e\}$. Let the “toggle” function $\bar{\cdot}$ be defined over elements of a pairs (of distinct players $S, R$ and of distinct profiles $C, I$) such that for $u$ an element of $\langle \alpha, \beta \rangle$, $\bar{u}$ is the other element of $\langle \alpha, \beta \rangle$. Finally, since I assume that there are only two profiles, the likelihood that your opponent is an innovator is just 1 less the probability that your opponent is a conformist. We may then calculate Generalized Expected Utility (GEU) in a signaling game:

$$EU(x_e) = \sum_{i \in I} \rho_{ye} \times U_e(x_e, ye)$$

Then, so long as $\rho_{xe} > \rho_{ye}$, it follows that $EU(x_e) > EU(y_e)$. In other words, so long as it is more likely that someone plays $x$ rather than $y$, it follows that $\langle x_e, x_e \rangle$ is a strict Nash equilibrium. Because expected utility is maximised by minimizing the risk of communication failure, conforming to the majority pays off among coordinators.

Given initial proportions, it’s more rational to conform to majority use. In terms of the notion of risk: innovative behavior jeopardizes the chances to get rewards, in the form of mutual understanding, hence it is always risky to violate a convention. In the rational choice interpretation of a coordination signaling game, risk-aversion is the
force driving coordination. Conformists are the risk-sensitive players who minimize the possibility of communication failure. As a result, if most use *queer* pejoratively, even though we might (in principle) all start using it positively tomorrow, in fact we won’t because almost nobody expects that anybody would. If most use *condescension* as a positive term, we won’t start using it negatively because almost nobody expects that anybody would. Conventions among coordinators are robust: if they exist, they tend to persist.

### 3.2 Evolution and Drift

An understanding of linguistic conventions within the rational choice paradigm is problematic, in particular because the normative pull to conform to a convention depends, from this perspective, on the hypothesis all interlocutors assume that they all are rational. The shared rationality assumption has indeed been made explicitly by Lewis (1969) and Schiffer (1972), and often assimilated to Grice’s (1989) assumption of shared cooperativity and Stalnaker’s Common Ground (Stalnaker, 1978, 2002). Choice in a decision problem that leads to coordination then follows by reasoning strategically from the assumption that all are rational, together with a preference to maximize utility. The shared rationality assumption, however, has been criticized by several authors, from Burge (1975) to Lepore and Stone (2014). I will have some remarks on this point in the conclusion.

Linguistic conventions may be understood, less problematically, from an evolutionary perspective. Symmetry-breaking among alternatives is then no longer a matter of risk-avoidance, but of resistance to invasion. Evolutionary game theory has been studied extensively in Biology to model how the fitness of strategies (which, in
that setting, represent phenotypes) correlates with their distribution over time (May-
nard Smith 1982). Evolutionary game theory is a study of equilibria that obtain in
nature among all sorts of living beings, including monkeys, bacteria, and bees, for
which it would be (1) implausible to assume that they behave according to principles
of rationality, and (2) such assumption would be explanatorily negligible as a “ratio-
nal reconstruction” of their actions. In the study of language, we can model how the
likelihood of communicative success (a sort of “linguistic fitness”) correlates with the
frequency of profiles over time.

We may introduce adaptive adjustments to linguistic behavior: that is, linguistic
actions may be determined not only by an agent’s profile, which is “assigned at birth”
so to speak, but also by imitation. Intuitively, frequencies of strategies, over time,
are proportional to communicative success: players observe successful communicative
behavior in their peers, and tend to imitate it at the next round of play. Unlike
biological evolution, which is a matter of reproductive success, linguistic (and more
generally, cultural) evolution is a matter of learning from experience. If a strategy is
not conducive to successful communication, it will die out; if a strategy is commu-
nicatively successful, it might be contagious.

Let’s suppose that we have a large population of agents who take turns at being $S$
or $R$, and are randomly matched to one another at each interaction. Large, possibly
infinite, populations are commonly assumed in order to normalize statistical fluctu-
ations in the long run. Random matching is also a mathematical simplification, but
one that we will need to revise later.

The replicator dynamics is the most common model of evolution—whether by
reproduction or by imitation (Huttegger, 2007). Given that GEU above defines the
value of individual strategies in one interaction, the replicator dynamics defines the increase over a time step in the frequency of strategies in proportion to the difference between its communicative success and the population’s average payoff. The average payoff for the population relative to a role \( e \), \( W_e \), is calculated as follows:

\[
W_e = \rho_{xe} \times EU(x_e) + \rho_{\bar{x}e} \times EU(\bar{x}_e)
\]

The replicator equation RD defines the frequency of an \( e \) strategy at the next time step in terms of its communicative advantage over \( W_e \) (Deo, 2015b):

\[
\rho'_{xe} = \rho_{xe} \frac{EU(x_e)}{W_e}
\]

The likelihood \( \rho'_{xe} \) of a strategy \( x_e \) being played at the next time step is the product of the likelihood \( \rho_{xe} \) of \( x_e \) at the previous time step by the proportion between \( x_e \)'s expected payoff and the average payoff. Under the assumptions made above, RD models how quickly a population is driven to conformity to the most frequently observed behavior. If players are coordinators, conformists succeed if matched with conformists and innovators succeed if matched with innovators, but since conformists are in the majority, in the long run they do better than the population average (having more chances of profitable encounters), and so conformist strategies take over. Non-conforming behavior tends to disappear.

Some strategies are stable with respect to RD. In a biological setting, evolutionary stability is taken to model the property of some phenotypes of being resistant to the invasion of a small number of “mutants” (i.e., alternative phenotypes). Resistance to mutant invasion is a mark of robustness. Likewise, some linguistic strategies are such that, if they are followed by a large enough number of agents, then even if a small number of innovators should appear, the population is driven back to conformity.
Conventions are *evolutionarily stable* in this sense. Following Maynard-Smith (1982, 14), and assuming the frequency of an alternative strategy \( y \neq z \) is small enough, \( z \) is evolutionarily stable if and only if:

either \( U_e(z, z) \geq U_e(y, z) \), or

\[
U_e(z, z) = U_e(y, z) \text{ only if } U_e(z, y) > U_e(y, y)
\]

Since linguistic agents are coordinators, \( U_e(x_e, x_e) > U_e(x_e, x_e) \), i.e. it’s better to be a conformist against a conformist than otherwise. A large number of conforming players resist invasion by a small number of innovators. Hence, it may be impossible for a linguistic innovation shared initially by a small number, to spread to the entire population. The dominant equilibrium is thus decided by the initial distribution of probabilities and by the imitative behavior of coordinators. In this sense, conventions are evolutionarily robust.

The robustness of conventions, whether from a rational choice or an evolutionary perspective, explains how a group of speakers, whose linguistic behavior may be initially quite erratic, could eventually crystallize in stable regularities. But it appears now that there is too much stability. Innovators, for instance about the use of *saleema*, *fish*, *queer*, or *condescension*, cannot hope to see their alternative convention replace the existing one, in a population of coordinators. Any innovative uses will be driven to extinction (or rationally dispreferred on the ground of risk-aversion). In the next section, some assumptions made by Lewis will be revised in order to provide an explanatorily adequate model of semantic social change.

There is one general mechanism of change that might explain linguistic evolution within Lewis’s coordination model, but as we shall see in the rest of this section, this is
not the right account of semantic social change. The mechanism is *drift*. Conventional
equililibria may be disrupted by random stochastic processes\(^8\) In particular, just
like biological reproduction is not cloning, cultural reproduction, or more specifically
linguistic imitation, need not be a matter of reproducing perfect copies. A certain
kind of randomness is built into evolutionary processes. Typically, this randomness
is interpreted as *learning mistakes*, of which examples may be found aplenty:

A case in point is the well-known development of the German words for
‘head’, *Kopf* and *Haupt*. ... In medieval German we have the word *Kopf*,
drinking vessel, ‘cup’, which is etymologically related to Vulgar Latin
*cuppa*, the source of English *cup*. In the 12th century soldiers started to
use the word *Kopf* as a sarcastic metaphor to refer to the heads of enemies.
In the course of the 15th century *Kopf* became widely used and the use
loses its sarcastic note. During this period the use of *Kopf* competes
with the traditional Germanic word *Haupt*, head, and the use of *Kopf*
for referring to drinking vessels becomes obsolete. From the 16th century
onward, *Kopf* is the everyday word for referring to the head, whereas
*Haupt* continues to be used only in formal language. (Fritz, 2012)

Speakers’ error may be a factor in the history of German *Kopf*: people just lost track
of what the word meant. A similar case is the story of English *bead*, derived from the
Old English word for ‘prayer’, plausibly because it had become unclear whether the
word was for rosary beads or for the prayers recited with them (Hock, 1991 296)\(^9\)

\(^8\)It is worth noting that the philosophical literature on reference has acknowledge
the role of drift in semantic change at least since Gareth Evans’s *Madagascar*
example, in Kripke (1980).

\(^9\)The literature on semantic change in the cognitivist tradition classifies examples of this kind
as “metaphorical” or “metonymical change” (Fritz, 2012, Geeraerts, 2002). In this tradition, which
includes Pustejovsky (1995) and goes back to Katz and Fodor (1963), meanings are understood as sets
of features (e.g., \(+\)Animated, +Physical...) and so on), and consequently semantic change is a matter
of losing or gaining features. While the exploitation of metaphors plays a role in linguistic evolution,
classifying a pattern of change by figures of speech is of dubious explanatory value. More generally,
the limits of the feature-based model of the lexicon are well known: successful feature-based accounts
of lexical items are few and far between, and always exposed to quite unpredictable idiosyncrasies
of use (Hanks, 2013). Grondelaers and Geeraerts (2003) appear to share some misgivings too, and
gesture toward a more pragmatic approach to semantic variation.
Uncertainty is already built into the model I described above, as the receiver’s lack of information about the sender’s intended use, and the sender’s lack of precision about the initial meaning. What’s not yet part of the model is genuine random variation. Plausibly, we may account for examples like Kopf and bead, from an evolutionary perspective, by introducing alternative strategies at random. Variation is thus a natural phenomenon, due in part to a combination of noise in the signal, and random drift. Randomness can be formalized by the mutator-replicator dynamics, introduced by Hadeler (1981) (on which see also Skyrms (2010, 61-62), and Newberry et al. (2017) for an overview of studies of linguistic variation based on drift).

Let $V$ be an ordered set of $n$ strategies such that $A_S \cup A_R \subseteq V$, allowing for the possibility that $V$ contains more strategies than in the initial setup, defined over suitable supersets of $A_S$ and $A_R$. We assign a mutation probability $q_{xy}$ to any pairs $\langle x, y \rangle \in V \times V$. For example, $q_{k_1j_2}$ is the probability that a conformist receiver mutates randomly into an innovative sender at the next round. A matrix $Q_{xy}$ is the set of all mutation probabilities. We can now define the mutator-replicator equation (Deo, 2015b, 35-37):

\[
\rho'_x = \sum_{y \in V} Q_{xy} \frac{\rho_y \times EU(y)}{W_y} \tag{MR}
\]

The frequency of a strategy $x_e$ at the next round of play is the sum for all strategies $y_g \in V$ of their replication rate $EU(y_g)/W_g$, multiplied by the frequency of $y_g$, weighted by the probability that $y_g$ mutates into $x_e$. The possibility of random mutation may destabilize a conventional equilibrium, and possibly lead to new conventions. There are learning mistakes.

Therefore, the coordination model is compatible with some instances of semantic change, e.g. in which change is effected by chance. As I pointed out above, it is
the speakers’ shared interest in successful communication that drives coordinators to conformity. If you are a coordinator, your only interest in the outcome of a linguistic interaction is to get your message across, and the most efficient way of doing so, is by using language in the way others do—setting errors aside, there is no spontaneous variation in this picture.

Learning mistakes may well account for semantic variation as observed in German Kopf and English beads. However, sometimes speakers innovate not by mistake but in response to some goal they have. Indeed, this is the case of social semantic change. Sometimes speakers innovate, for example, out of a concern for how relevant parts of the world are represented in their speech. This is the case with the evolution of saleema, fish, queer, condescension, none of which is plausibly the result of a stochastic process. So some other mechanism for change must be introduced in a model of linguistic interaction that be diachronically plausible.

The discussion of Lewisian conventions highlighted that it is the speakers’ shared interest for mutual understanding that drives conformity. However, it is not the case that speakers are only interested in understanding each other. Mutual understanding is the basis of communication, but a lot more is achieved in a linguistic interaction than that. Dropping the assumption that interlocutors are merely coordinators will get us to an account of semantic social change, and to a more plausible model of communication.

4 Dynamic conventions

On grounds of rational choice, existing conventions tend to persist, because innovating against a majority has low expected utility. On evolutionary grounds, existing
conventions tend to persist, because innovators are driven to extinction under the replicator dynamics. While the possibility of learning mistakes under the mutator-replicator dynamics may introduce some instability in the long run, chance is not a good explanation of semantic social change. The problem is to describe how semantic change might reflect the attitude of a community of speakers, as suggested by the examples of saleema, queer, fish, and condescension.

Consider saleema. Intuitively, while an interest for mutual understanding pulls speakers towards conformity to the existing convention, in which saleema is used to talk about “whole and healthy” things, speakers might have practical goals, in particular extralinguistic goals, that could shape their linguistic policies. An initially small group of Sudanese Arabic speakers, for example, might innovate on the use of saleema moved by a concern about certain traditional practices. The new use might then spread to the rest of the population, should the innovators be able to exercise enough peer pressure on other speakers to force them to comply.

The two elements to be added to signaling games to study diachrony are: (1) a multiplicity of interests for each player, and (2) normative pressure for compliance to the innovators’ convention, even though it might not what the majority conforms to.

A generalization of linguistic games to allow for multiple interests finds precedent in recent literature. Jespersen’s cycle is a cross-linguistically well attested regularity in the history of the meaning of negation: roughly, emphatic negation becomes plain out of excessive use, and is replaced by new material to express emphatic meaning. Ahern and Clark (2017) depart from the basic coordination model to account for Jespersen’s cycle. The historical evolution of the meaning of negation, according to their proposal, is driven by two different interests interlocutors have, but don’t
share equally: informativity and status. Listeners have an interest in the speaker being informative, whereas speakers have an interest in listeners paying attention to them (this is what they call status). These interests need not pull in the same direction: emphatic negation gets overused in the speakers’ attempt to gain status, and thus loses its informative value. So the emphatic negation of yesterday becomes the plain negation of today. At that point, new material gets recruited for emphatic use. Ahern and Clark’s account can be generalized straightforwardly, potentially providing a recipe for a variety of inflationary cycles in language evolution.[10]

A multiplicity of interests gives linguistic interactions the structure of a social dilemma, having multiple interests implies that they need not pull all in the same direction. For the reasons noted above, Ahern and Clark’s account does not apply to semantic social change, which is neither cyclical, nor regular, nor driven by the grammar, as instances of change like Jespersen’s cycle are. Instead, semantic social change is determined by societal pressure: it’s a kind of social change that becomes visible in the history of the lexicon. Importantly however, the formal study of social dilemmas leads to an account of how conventions can be made genuinely dynamic.

For this purpose, we shall no longer assume that the behavior of linguistic agents is wholly determined by a single interest. People may have multiple goals. (This is, of course, a welcome step away from too much idealization.) It follows that the players’ interests may be misaligned, and games in which the interests of players may diverge

10The notion of status has also been employed by Franke et al. (2012), and by others. Status is a useful tool in game-theoretic modeling, because it allows to manipulate the structure of games and study what happens when speakers are not pure coordinators. Franke et al. (2012) employ games of conflicting interests to study strategic pragmatic inferences when speakers don’t follow Grice’s Cooperative Principle. I am not sure that the discussion below could be rephrased in terms of status, rather than practical goals, preserving its most important elements. In part this is because these authors aren’t very explicit about what ‘status’ is.
are mixed-motive games. Typical examples are the Prisoner’s Dilemma and the Stag Hunt. In mixed-motive games, there are asymmetries between possible solutions. I will focus on SH, because the structure of payoffs in this case brings out very clearly the difference between alternative strategies (however, the proposal below works for PD games just as well, as shown by Bicchieri (2006)). We may start by considering a normal form SH. There are two players, You and Myself, and two strategies for both:

\[
\begin{array}{ccc}
\text{You} & y & x \\
\text{Myself} & y & (2,2) & (0,1) \\
& x & (1,0) & (1,1)
\end{array}
\]

As in the simple coordination game above, there are two strict Nash equilibria, namely \(\langle x, x \rangle\) and \(\langle y, y \rangle\). Notice however that the outcomes are different in the two states.

The most famous interpretation, and earliest description, of SH may be found in Jean-Jacques Rousseau, in the *Discourse on the Origin of Inequality*. Rousseau describes a contrast between an activity with higher rewards but lower chances of success (hunting deer), and an activity with lesser rewards but easier to accomplish (hunting hare). Deer hunting has lower chances of success because it requires the full cooperation of other hunters. So it has a higher risk of failure. Rousseau describes the conflict between higher-risk high-payoff behavior, and lower-risk low-payoff behavior:

If a deer was to be taken, every one saw that, in order to succeed, he must abide faithfully by his post: but if a hare happened to come within the reach of any one of them, it is not to be doubted that he pursued it without scruple, and, having seized his prey, cared very little, if by so doing he caused his companions to miss theirs.
Just like coordination games, SH has two stable equilibria, \langle y, y \rangle and \langle x, x \rangle: everybody hunts deer and everybody hunts hare. However, outcomes are now asymmetric. The \langle x, x \rangle equilibrium is lower-risk, since hare hunters do better \textit{ceteris paribus} against deer hunters than conversely (given that hare hunters are more likely to walk away with a non-zero payoff anyhow). However, \langle y, y \rangle is (Pareto) optimal, for no alternative combination would make someone better off without making anyone worse off. Conversely, \langle y, y \rangle is riskier, for it obtains only if no one defects, but \langle x, x \rangle is not optimal. SH has the structure of social dilemmas.

In a linguistic environment undergoing semantic social change, speakers are confronted with a dilemmas of roughly the same sort as Rousseau’s hunters. They are caught in the alternative between a linguistic policy that minimizes the risk of communication failure (conformity to the existing convention), and an innovative policy that awards better payoffs (since it better represents our own practices and goals). The two come apart because the more optimal alternative comes with a higher risk of communication failure, given that it is initially only a small minority of innovators that follow it.

Conformity to an existing convention despite the availability of a more optimal alternative is thus akin to the selfish behavior discussed by Rousseau: players prefer a higher likelihood of suboptimal success. Dynamic conventions explain why anybody would take risks and deviate from an existing Lewisian convention, choosing instead a more optimal solution. After all, if we can relatively easily catch a hare, why coordinate instead on a riskier deer hunting? Because a normative expectation might be operative, given broader (often non-linguistic) goals that are shared through an agents’ social network. So even though we understand each other perfectly well using
ψ to mean \(i_1\), normative pressure might force us to innovate and use \(ψ\) to mean \(i_2\). Compliance with a new convention may thus be established. Consider a SH signaling game in extended form:

As above, we can put the distinction between conformists and innovators in terms of risk proclivities, as well as evolutionarily. As above, communication succeeds if players coordinate on the same meaning, hence for states \(\langle C_S, C_R \rangle \in \{\langle j_1, k_1 \rangle, \langle j_3, k_1 \rangle\}\) and \(\langle I_S, I_R \rangle \in \{\langle j_2, k_4 \rangle, \langle j_4, k_4 \rangle\}\). The first option is the safe bet, even though, intuitively, by following it we keep on talking in ways that are now judged to be offensive, misleading, or otherwise misguided. In other words, players are never penalized for conforming to the existing equilibrium, but they are better rewarded for innovating. Hence, conformists behave just like in coordination games: they tend to play strategies that map to \(i_1\) and minimize risks. The \(\langle C_S, C_R \rangle\) equilibria are just the risk-minimizing solutions to a coordination game. Things have changed for innovators, who are less sensitive to risks, but who, unlike in coordination games (where they had no higher rewards than conformists), tend to maximize payoffs: communication may succeed if players coordinate on \(\langle I_S, I_R \rangle\) equilibria, but by doing so, they get
better rewards than conformists. This type of equilibrium is specific to mixed-motive games: it’s the risk-dominated but optimal solution to SH.

If a population is divided in the two types, conformists and innovators, conformists speak to each other by coordinating on \( \langle C_S, C_R \rangle \), innovators speak to each other by coordinating on \( \langle I_S, I_R \rangle \), and what remains to be explained is what happens when the two types interact. As I pointed out above, what must be explained is how the population does not split into subcommunities with different sublanguages. Following Bicchieri (2006, 11) and (2017, 35), I shall introduce a mechanism whereby, despite the higher risks of failure, there is an incentive to innovate (or more precisely, a cost for failing to do so). This way, innovative behavior initially only present in a small minority, may take on the rest of the population, getting (nearly) everyone to conform to a new equilibrium.

Intuitively, players in either group are affected by a normative pressure to follow their peers. It thus becomes costly to fail to innovate when the pressure on an agent coming from innovators becomes preponderant. This cost makes it more advantageous to join the innovative equilibrium. The combination of strategies that leads to the optimal solution to SH is a \textit{dynamic convention}\footnote{Bicchieri (2006), from which I draw, calls this arrangement a \textit{social norm}. I reserve this term instead for the norm issuing from an agent’s social network that influences the agent’s behavior. See below for discussion.}

A behavioral regularity \( R \) in a situation of type \( G \), where \( G \) is a mixed-motive game, is a \textit{dynamic convention} in a population \( P \) if and only if there is a large enough \( A \subseteq P \) such that for any \( a \in A \),

1. \( a \) believes that \( R \) applies to \( G \) (\textit{Contingency condition}), and
2. \( a \) has an interest in conformity to \( R \) provided that:
   (a) \( a \) believes that enough individuals in \( P \) conform (\textit{Epistemic expectation}), and
A Lewisian convention is the special case of a dynamic convention, in which the normative component (b) is empty. Violation of a convention is not sanctioned, for an agents’ conditional interest suffices to explain conformity to the convention. In order for me to be at the cafeteria for our meetings, it suffices that I want to meet you: there is no need for penalties to force me to conform. On the contrary, dynamic conventions are supported, in addition to a conditional interest for conformity, by the belief that compliance is enforced. This is a normative expectation: in a dynamic convention you (believe that you) are expected to conform, by enough individuals, who may sanction your failure to comply. Violations of a dynamic convention activates a social norm: normative pressure from a group of peers that force compliance by sanctioning transgressors. Sanctions may take very different forms, ranging from manifest discomfort, to reproach, to alienation: “If others believe one ought to conform, the reaction to nonconformity may range from slight displeasure to active or even extreme punishment” (Bicchieri, 2017, 35). Importantly, the normative pressure exercised by a dynamic conventions explains the possibility of shifting to an innovative equilibrium.

The set of individuals in $P$ that define a’s normative expectation is a’s social network. Elements of a social network are those among an agent’s peers whose proximity to the agent and whose role in their social life may contribute to determine linguistic action. The influence of peer pressure in guiding our actions and strengthening our opinions has long been recognized (Asch, 1951). A social network may be as small as containing only one element, who acts like an opinion leader, whose influence may
force a to comply to a convention. Of special significance is the case of trendsetters: individuals who, in part because of the particular position they occupy in their own social network, may initiate change—the prototypical example in matters of social change is Rosa Parks (see the discussion in ch. 5 of Bicchieri (2017)), but another example familiar to linguists might be Celeste Sullivan, described by Labov (2001) as a ‘leader’ of linguistic change in Philadelphia. It’s only to be expected that a similar situation is to be found in semantic change. The relevant properties of social networks have been studied in the sociolinguistic literature and in the social sciences, isolating measurable parameters to describe their structure. Likewise, the psychological features of trendsetters have been extensively discussed elsewhere (see Bicchieri’s works listed above and references therein).

Formally, dynamic conventions solve mixed-motive games in the following sense: if a normative expectation is active, a social dilemma like SH is transformed into a coordination game (Bicchieri, 2006, 3). Thus normative expectations directly intervene on the structure of payoffs of an SH game, by introducing costs incurred by transgressors. Let a constant $c_e$ for $e \in P = \{S, R\}$ measure a player’s sensitivity to a dynamic convention: the payoff deduction inflicted to the convention violator will be proportional to $c_e$. The operator $\max(\forall g)$ is defined over all players $g \in \{e, \bar{e}\}$ other than the norm transgressor, so that costs are calculated differently depending on which group one is complying with. Given two players $e$ and $\bar{e}$ (recall that $\bar{\cdot}$ is the “toggle” function) and two strategies $x$ and $y$ respectively, fix $z$ so that $z = x$ if $g = e$, and $z = y$ if $g = \bar{e}$. A norm-based utility function $N$ is then defined as follows:

$$N_e(x_e, y_{\bar{e}}) = U_e(x_e, y_{\bar{e}}) - c_e \times \max(\forall g)\{U_g(z_e, z_{\bar{e}}) - U_g(x_e, y_{\bar{e}}), 0\}$$
The norm-based utility for $x_e$ against $y_e$ (e.g., a conformist sender against an innovative receiver), is the utility for $e$ of $x_e$ against $y_e$ minus some costs. These costs are proportional to a measure of $e$’s normative expectation, $c_ε$, and to the difference between the payoff one gets by complying and the current payoff (where this value just goes to 0 in case such difference is negative). The more one is sensitive to social pressure to comply, or the more it would be advantageous to comply, the more failure to comply is costly.

For example, in a two-players SH signaling game, utilities for a sender $S$ are calculated as follows, where we assume that there is normative pressure to comply with the innovative behavior (i.e., the norm transgressor is the conformist player, who fails to comply with the innovators)\footnote{The utility calculation in case normative pressure is exercised by conformists is analogous. If that’s the case, more countervailing pressure against innovation than a mere preference for conformity is cast upon the agent. This does not rule out change, though it makes it harder.}

\begin{align*}
N_S(C_S, C_R) &= U_S(C_S, C_R) \\
N_S(I_S, I_R) &= U_S(I_S, I_R) - c_S \times \max \left\{ U_S(I_S, I_R) - U_S(I_S, I_R),
U_R(I_S, I_R) - U_R(I_S, I_R),
0 \right\} \\
N_S(C_S, I_R) &= U_S(C_S, I_R) - c_S \times \max \left\{ U_R(I_S, I_R) - U_R(C_S, I_R), 0 \right\} \\
N_S(I_S, C_R) &= U_S(I_S, C_R) - c_S \times \max \left\{ U_R(I_S, I_R) - U_S(I_S, C_R), 0 \right\}
\end{align*}

In the first term, $N_S(C_S, C_R)$, the subtrahend goes to 0, because there both $S$ and $R$, if conformists, fail to comply with the innovator’s behavior, hence the argument of $\max$ is 0. In the second term, $N_S(I_S, I_R)$, $\max$ keeps all players as argument, but given the utilities assigned to $\langle I_S, I_R \rangle$, the subtrahend goes to 0 as well. Hence
\[ N_S(C_S, C_R) = U_S(C_S, C_R) \] and \[ N_S(I_S, I_R) = U_S(I_S, I_R) \]. Otherwise put, utility functions in a dynamic convention are calculated just like in Lewisian conventions if players coordinate (though of course their values may be different).

Matters are different for combinations of strategies where coordination fails, i.e. \( \langle C_S, I_R \rangle \) and \( \langle I_S, C_R \rangle \). Given the payoffs above, these reduce to \( N_S(C_S, I_R) = 1 - 2c_S \) and \( N_S(I_S, C_R) = -2c_S \). Similar calculations hold for the receiver.

A SH becomes a coordination game in which equilibrium is found among innovators for values of \( c_S \) and \( c_R \) such that

\[ N_S(C_S, I_R) < N_S(I_S, I_R) \] and \[ N_R(I_S, C_R) < N_R(I_R, I_R) \]

i.e. in which it is more convenient to comply with the innovative behavior for both sender and receiver. The intuitive interpretation is that if a conformist sender meets an innovative receiver, the activation of a social norm may make conformity less appealing, and eventually tip the balance in favor of innovation. Likewise, if an innovative sender meets a conformist receiver, the sender’s payoff is lower than it would have been, had the receiver been an innovator.

For example, suppose you and I participate in an efficient communication system in which fish is used to mean ‘animal for fishing’ (i.e., including whales). The conformist behavior is to use fish to mean ‘animal for fishing’, and this being the existing Lewisian convention, there is no incentive for us to deviate from it. Some elements in our social network, however, may start using fish in a more biologically accurate manner, in which fish is distinct from mammals not by reason of the human practices in which the animal gets involved but of its phylogenetic tree. Perhaps we have no particular commitment to the existing convention, and perhaps we are particularly exposed to the pressure of those amongst our peers who have innovated in their speech with
regards to *fish*. If this is case, it may be that it is less convenient for us to stick to the existing convention, for we may come to be regarded as speaking in an obtuse and unpractical way. Thus, we comply with the innovative convention.

### 4.1 Social networks

Social dilemmas have been used to model the emergence of altruistic behavior in a population of self-interested agents. In an evolutionary setting, we have seen that the replicator dynamics drives a population to conform to the behavior that is most frequently observed in the population. So the task is to explain how this trend can be overturned.

The assumption that gets dropped is random matching: that all types of players (Conformists and Innovators) meet all other types randomly.

Assortment of encounters—that is, positive correlation of types in encounters—plays the major role in explanations of the evolution of altruism. Altruism, modeled as cooperation in the Prisoner’s Dilemma, cannot evolve with random pairing. But it can when there is sufficient positive correlation of types, so that cooperators tend to meet cooperators and defectors tend to meet defectors. Mechanisms exist in nature to promote an assortment of encounters. There is no reason to believe that they should operate only in Prisoner’s Dilemma situations. They can make a difference in evolution of signaling. ([Skyrms, 2010](#skyrms2010) 70)

Altruism corresponds with the deer hunting solution to the SH: where payoffs are maximized at higher risks, provided that no one selfishly defects and hunts hare. If outcomes are asymmetric, players confronted with the dilemma between optimization of payoffs and risk-minimization, ideally, ought to coordinate on the optimal solution. But this won’t happen under random matching.
Random matching is mathematically useful for its simplicity, but it’s an extreme idealization anyway. Lewisian conventions correspond to evolutionary stable strategies in a replicator dynamics under the assumption of random matching (Skyrms, 1996). We have seen that dynamic conventions are equilibria of mixed-motive games. In an evolutionary setting, the optimal but riskier strategy, such as cooperation in the Prisoner’s Dilemma or an innovative equilibrium in the SH game above, cannot evolve under random matching. The population is always driven back to the less optimal equilibrium.

Things change if we take into account the fact that agents tend to meet like-minded agents often enough. This seems to be a quite realistic assumption, and allows us to quantify the role of social networks in an evolutionary setting. Speakers with similar concerns, that share relevant aspects of their social identities, and similar attitudes and practical goals, tend to speak to each other more often than they speak to others who have altogether different beliefs and desires. Innovators tend to speak to innovators, and conformists tend to speak to conformists. There is no random matching in the real world. Innovative conventions may thus develop.

Earlier I indicated with $\rho_{CR}$ the probability to encounter a conformist receiver, and noted that, at least initially, $\rho_{CR}$ is just a measure of the distribution of players’ profiles. Reading ‘$x_e \bullet \bar{x}_e$’ as ‘$x_e$ meets $\bar{x}_e$’, the following equations define the likelihood of targeted matching (Skyrms, 2010, 71):

\[
\rho(x_e \bullet x_e) = \rho_{xx} + \varepsilon(1 - \rho_{xx})
\]
\[
\rho(x_e \bullet \bar{x}_e) = \rho_{x\bar{x}} - \varepsilon \rho_{x\bar{x}}
\]
We can now replace these matching probabilities in the mutator-replicator equation MR for a strategy $x_e$:

$$\rho'_{x_e} = \sum_{i=1}^{n} Q_{y_i,x_e} \frac{\rho(x_e \bullet y_i) \times EU(y_i)}{W_h}$$

The frequency of a strategy $x_e$ at the next time step is the weighted sum, given mutation probabilities, of any strategy’s replication rate multiplied by the frequency of the strategy under a regime of targeted matching. Notice that for $\varepsilon = 0$ we have random matching, and the equation above reduces to MR. For $\varepsilon = 1$ we have perfectly targeted matching: innovators and conformists never interact. For all remaining real values of $\varepsilon \in [0, 1]$, we may observe the emergence of dynamic equilibria under the mutator-replicator dynamic. As Skyrms remarks, “there is always some degree of correlation that will do the trick” (Skyrms [2010] 71), i.e. on which a payoff-maximizing solution spreads contagiously, and an innovative convention is established.

The parameter $\varepsilon$ is a value associated with the influence of a social network in an agents’ linguistic action. One extreme, $\varepsilon = 0$, indicates a complete absence of a network: there is thus no normative pressure exercised on an agent, who is free to minimize risks. In this scenario, conformists strategies define linguistic behavior. Social networks may be more or less tight, and thus more or less determinant forces. Close-knit groups activate more powerful social norms, and individuals in such groups share goals, values, and expectations. It may be mostly a vague matter how to define the impact of one’s social network at any specific time or circumstances. However, it is important to recognize the role played by one’s membership in a social network to account for semantic social change. The resulting model of communication is explanatory, since it isolates the significant variables that track change. These variables can be manipulated and measured, e.g. by means of questionnaires (for one’s sensitivity
to a social network), or quantitatively by collocation data about the use of words in corpora (giving a statistical measure of semantic change over time).

5 Concluding remarks

The aim of this paper was to lay out a generalization of Lewisian conventions, to account for observable facts about semantic change. Lewisian conventions are robust because coordinators share the same interest. From an evolutionary perspective, the assumption that agents may imitate successful behavior is formalized by the replicator dynamic, which may lead an initially chaotic distribution of actions to a stable equilibrium. A better approximation of linguistic history is given by the mutator-replication dynamic, which includes randomized learning mistakes in the transmission of conventions. Some instances of semantic variation, however, do not seem to be random, but rather oriented by the agents’ practical goals. As speakers’ attitudes and goal may change over time, semantic regularities may change as well.

Some assumptions about Lewisian conventions have been modified to account for semantic social change, adapting Bicchieri’s work to sociolinguistics: first, following Ahern and Clark (2017), mixed motive games allow for the possibility of asymmetric outcomes, given by a multiplicity of interests for agents who are no longer mere coordinators; secondly, social norms imposed by a social network introduce the possibility of sanctioning transgressions to a type of innovation that, in the network, commands compliance, despite being followed only by a minority. The dynamics thus introduced

13As a more general feature of scientific generalizations in the social domain (Epstein 2008; Hausman 2018), the model may not be predictive, but this is a more general issue that deserves to be discussed elsewhere. Another issue that deserves further discussion is the structure of the lexicon. The question is how to integrate the relationship between lexical items in a semantic field with the present approach.
explains how a strategy played initially only by a small set of agents may nevertheless propagate to all speakers of a language for which semantic conventions already exist. By way of concluding remarks, I wish to return to a point briefly sketched earlier on what the model described here requires of speakers by way of their competence.

A signaling game between sender and receiver, whether amongst coordination or competitors, is defined over a set of meanings \( J = \{i_1, i_2, \ldots \} \). As I remarked above, a set of meanings of the same semantic type (e.g., a set of propositions) is a question. Such question articulates the different moves available to the players: Which meaning should I encode in the message? and Which interpretation should I decode from the message?

These questions are central to the abstract structure of the signaling game. In acting in response, an understanding of the interlocutor’s communicative intentions seems necessary. Understanding of intentions is, in part, an awareness of relevance relations, including what is salient to the interlocutors, and relevance is, of course, a well accepted element of conversational competence (Sperber and Wilson 1995; Roberts 2012).

There is more to say though. In particular, competence in a sender/receiver game includes two main elements: (1) an evaluation of the likelihood that the interlocutor belongs to a certain profile (innovator vs. conformist), and (2) an understanding of the interlocutor’s goals. These two components are not independent: understanding your interlocutors’ goals requires, among other things, a correct evaluation of which type of player they are. For example, perhaps your interlocutor is a Victorian liberal for whom condescension implies haughtiness, and perhaps she is using condescension (and its cognates) in a way that manifests her dissent for the social and economical inequality
of traditional English society. These are modest assumptions about competence, for they only appear to require a capacity to interpret hints from your interlocutor’s behavior about their profile.  

A description of the game in terms of questions, and of the goals of the participants, highlights the connection with the literature on the Question Under Discussion (Roberts, 2012), which generalizes the traditional approach of Grice (1989). Like the games described here, a discourse in the QUD model is structured by questions, which define relevance relations in a conversation, and goals, which determine the speakers’ strategic contributions (Thomason, 1990). Like QUDs, the metalinguistic questions that structure the game of sender and receiver are at best implicit in most cases (not every discourse is an orderly sequence of questions and answers). Unlike QUDs, the metalinguistic alternatives that constrain the sender/receiver interaction are neither directly nor indirectly determined by focus structure (Rooth, 1992; Roberts, 2012). This suggests that, both grammatically and cognitively, the role of alternatives in semantic social change is different from what is known about the dynamics of a conversation.

There is independent evidence that semantic social change and classical pragmatics are not on the same level, for reasons discussed above. Semantic social change does not take place at the level of a conversation between a small number of speakers, in a relatively short time, but is a large-scale dynamic in a population in evolutionary time.

As Burnett (2017, 254) remarks, “speakers strategically exploit listeners’ interpretation processes to communicate properties about themselves and, in doing so, construct their identities”. By their linguistic policies, speakers identify themselves as members of a particular social group, perhaps defined by accepted social practices (saleema), or beliefs about society (queer, condescension). The process of identity construction in language use is central to much discussion in sociolinguistic pragmatics, a development of ‘third wave’ studies of variation (Eckert, 2013), and is parallel to the process of semantic change described here. I hope to discuss this connection in more detail elsewhere.
Hence a description in terms of questions and of the properties of individual sender and receiver is at best only a rational reconstruction of the linguistic interaction. Hence it does not purport to be an empirically accurate description of interlocutors’ cognitive life. Questions that define the central decision problems for sender and receiver are not semantic objects, as QUDs are.

It is certainly too much of an idealization to suppose that speakers perform expected utility calculations under the guidance of implicit metalinguistic questions to decide what is the most rational course of action in a sender/receiver game. As an idealization, it might nevertheless be useful insofar as it provides constraints on performance that can be tested empirically. For an analogy, Chomsky-style idealizations about grammatical competence, that attribute to speakers the capacity to produce sentences they will never, as a matter of fact, produce, is nevertheless empirically significant because it enables us to trace to e.g. working memory limitations whatever deviations from competence we may observe in performance. The value of the rational choice paradigm in a reconstruction of semantic social change, depends on the possibility of delivering such constraints—but I leave for another time a thorough discussion of this point.

A distinct but related point of contact with Pragmatics in the Gricean tradition is given by Stalnaker’s very influential notion of Common Ground (Stalnaker, 1978): the common ground is just common or mutual belief, and what a speaker presupposes is what she believes to be common or mutual belief. The common beliefs of the parties to a conversation are the beliefs they share, and that they recognize that they share: a proposition $\psi$ is common belief of a group of believers if and only if all in the group believe that $\psi$, all believe that all believe it, all believe that all believe that all believe it, etc. (Stalnaker, 2002, 704)
Stalnaker’s reconstruction of Gricean pragmatics based on the CG has been very successful both conceptually and empirically. If conventions are supported by a shared belief that we are playing a coordination game, and by believing about each other that we are rational, I believe that you prefer to conform, and you believe that I so believe, and so on, and vice versa. That is a quite expensive stack of iterated attitudes to carry around—what Michael Thompson calls a ‘doxastic pancake’\textsuperscript{15}

On Lewis’s definition of conventions, conventions are the result of conscious Gricean inferences drawn by speakers about each others’ rationality. Lewis assumes that all players know that they’re playing a coordination game: that’s why no one starts wreaking havoc once coordination is reached. His original definition thus includes a shared rationality condition, besides the contingency condition and epistemic expectation I have kept, that resembles Stalnaker’s common ground.

Lewis’s shared rationality assumption has attracted much criticism, as I have anticipated. It is in tension with the observation that all sorts of individuals use language competently, despite holding all sorts of beliefs, and even without a sophisticated theory of mind (Laurence \textit{1996}). Cognitive abilities that are required to use language competently are much less demanding than those required for higher order Gricean reasoning, which implies a recognition of each others’ action plans (Thomason \textit{1990}). Sperber and Wilson \textit{1995} go so far as rejecting the whole code model of communication on the ground that it must rely on shared iterated attitudes, but it is implausible to assume that assumption for elementary communicative interactions.

\textsuperscript{15}You and I’, talk delivered at the Aristotelian Society on May 31st, 2012. Available at \url{https://www.aristoteliensociety.org.uk/the-proceedings/proceedings-podcasts/}
Sperber and Wilson’s argument is unsound. It is not the case that all communication must rely on shared iterated attitudes. An evolutionary perspective on conventionality helps to isolate a level of linguistic action that does not depend on rational calculations that would allegedly follow from such “common ground” assumption. Formally, semantic regularities emerge as a result of an adaptive learning algorithm: this is how coordination in the behavior of all sorts of animals is often accounted for in biology (Garrod and Doherty, 1994; Barr, 2004). The kind of cognitive abilities involved in this type of behavior (e.g. epistemic expectations) should be understood as operating subconsciously or implicitly, as Bicchieri (2006, 97) emphasizes. So a basic level of linguistic communication plausibly takes place as a result of self-organizing processes driven by low-level subconscious attitudes. But no doxastic pancakes.

These considerations motivate an evolutionary interpretation of the framework over one based on rational choice. Linguistic actions are taken without much deliberation or inferences, in most cases. Even though semantic change is a matter of collective action, not all collective actions are like participating in a democratic election, or serving in a jury. Some are like not colliding with strangers on a crowded sidewalk. Semantic social change is more like the latter. Although the CG is needed to explain various aspects of communication, it is not required to account for collective action in general, including unsophisticated linguistic action. An explanatory account of semantic change may well take place at this level.


