Conflict, cheap talk, and Jespersen's cycle [*]

Christopher Ahern
Department of Linguistics,
University of Pennsylvania

Robin Clark
Department of Linguistics,
University of Pennsylvania

Abstract Game-theory has found broad application in modeling pragmatic reasoning in both the classical Gricean case of common interests between interlocutors and, more recently, in cases of conflicting interests. This work brings these considerations of common and conflicting interests to diachronic patterns of language use. We use tools from evolutionary game theory to characterize the effect of conflicting interests on how meaning is signaled with costless signals in a population over time. We show that the dynamics of a particular class of inflationary processes in language, including Jespersen's cycle, can be modeled as a consequence of signaling under conflicting interests and fit the resulting model to historical corpus data.

1 Introduction

Conflicts of interest play markedly different roles in Linguistics and Biology. In Linguistics, Gricean pragmatics has aimed at understanding the inferences that a listener can draw from a speaker's contribution on the explicit assumption of a shared set of purposes for an exchange. By contrast, research in animal communication has aimed at understanding the existence and persistence of signaling systems in the face of inter- and intra-species conflict. Both endeavors hinge on the role of conflict, either in its presence or absence. In the abstract, though, both deal with the transmission and interpretation of signals by senders and receivers. This leads us to consider what happens when we extend our linguistic considerations beyond perfectly aligned interests.

At first glance, this step outside the idealized realm of common causes yields some forbidding results: conflicts of interest would seem to erode communication. The following reasoning from animal communication makes clear the root of this unraveling (Searcy & Nowicki 2005). Imagine two agents: a sender who sends

---

a signal and a receiver who receives the signal. Suppose that the sender has no incentive to be truthful, in fact, let us suppose that he has every incentive to deceive the receiver. If the sender has an incentive to deceive, then the receiver should not listen. If the receiver does not listen, then the sender has no motive to signal in the first place. Crucially, the actions of the sender depend on those of the receiver and vice versa. Given this interdependence, conflicting interests undermine the information conveyed by signals, rendering them, so to speak, meaningless.

The same reasoning holds in the case of an entire population of senders and receivers interacting over time. Senders will learn or evolve to dissimulate and receivers will learn or evolve to distrust. This process takes on the familiar form of the *tragedy of the commons* (Hardin 1968). Individuals will always be tempted to exploit the common resource of credulity. Collectively, this incentive to exploit exhausts the resource. Only a fool would tell the truth when there is something to be gained from deception, and only a fool would trust others to be truthful.

Silence, or at best meaningless babble, is the equilibrium state in the population: neither senders nor receivers have reason to unilaterally change their behavior. Thomas Schelling's grim pronouncement comes to mind (1978: 26):

> The body of a hanged man is in equilibrium when it finally stops swinging, but nobody is going to insist that the man is all right.

This sentiment holds in two ways. The inability to convey information would seem problematic and we do observe informative signaling both in the case of human language and animal communication more broadly. The existence of such signaling despite conflicting interests is a genuine puzzle.

This problem was not lost on Grice, insofar as he recognized the fundamental role of his *Maxim of quality* to the entire enterprise.

> It is obvious that the observance of some of these maxims is a matter of less urgency than is the observance of others; a man who has expressed himself with undue prolixity would, in general, be open to milder comment than would a man who has said something he believes to be false...[O]ther maxims come into operation only on the assumption that this maxim of Quality is satisfied (Grice 1975: 27)

Yet, while we have every incentive to abide by the maxim of quality when it serves our interests, we have every reason to do otherwise when it does not. So, what keeps human language from the downward spiral to silence? In this regard we can look to animal communication where much work has been devoted to explaining the evolutionary stability of communication. These solutions take the form of different mechanisms that mitigate conflicts of interest between senders and receivers.

For example, a sender might guarantee his commitment to the truth by incurring a sufficiently high cost to send a signal. This *handicap principle* (Zahavi 1975) allows for stable signaling despite conflicting interests.[1] To take the usual example, a peacock incurs a cost for his magnificent tail: significant metabolic resources are required to develop the tail, and once developed, his ability to fly is hampered. In all, the tail makes the peacock far more conspicuous and thus vulnerable to predators. But, successfully bearing the tail serves as a signal of his genetic worth. A weaker peacock would not have been able to support the tail and avoid becoming something else's lunch. Thus, potential mates can take the tail as a signal of a a truly fit peacock. However, when we turn our attention to language, this sort of mechanism need not be appropriate. In fact, the notion that truthfulness is enforced by cost is problematic: truth tellers expend as much effort learning and producing their language as liars, and no more. More importantly, no amount of verbiage can ever guarantee the veracity of a statement. As we say, talk is cheap.

There are, of course, various alternatives to handicaps that could be appropriate for the case of language (Scott-Phillips 2008). Regardless of the details of the actual mechanisms that mitigate conflict, we know two things. First, given that language exists, such mechanisms are clearly sufficient to stave off total collapse. Second, given that signals are not always used in perfect accordance with the maxim of quality, such mechanism are not sufficient to ensure the idealized case of Gricean commonality. These facts prompt two important questions.

First, if language is indeed subject to a host of competing pressures, what does this mean for the usual Gricean rationale for linguistic behavior as something that is "*reasonable* for us to follow, that we *should not* abandon" (Grice 1975: 29)? When the interests of speakers and hearers diverge can we flip this rationale on its head and find a reason for change rather than stability? Second, given the stability of language as a whole, if not a tragedy of the commons, might we find a smaller scale *tragedy of the conversation*, where particular linguistic signals, but not the system as a whole, are destabilized through repeated interactions? Do we find instantiations of the predicted patterns of language use? Can we explain why particular kinds of change happen?

To address these questions, we consider the class of *inflationary* processes noted by Dahl (2001) in which the conditions of use for a particular linguistic signal are extended over time. In particular, we focus on the development in the expression of negation over time known as *Jespersen's cycle* (1917) where an initially emphatic form of negation increases in frequency over time and undergoes a kind of *bleaching*. We argue that this increase in frequency and loss of information can be explained by the fact that language is used both to convey information and negotiate social status

---

1 See Maynard Smith & Harper (2004) and Searcy & Nowicki (2005) for thorough discussions of handicaps in animal signaling.

(Dessalles 2007, Franke et al. 2012). Simply put, hearers benefit when speakers share useful information, but speakers also benefit socially when hearers confer status by listening to what they have to say. Under these conditions, the interests of speakers and hearers are largely, but not perfectly aligned. Using *evolutionary game theory* (Maynard Smith 1982), we model how the interaction between speakers and hearers changes the information carried by signals over time.

The main contributions of this work are threefold. The first contribution is that we offer a formal game-theoretic model to characterize the pressures underlying inflationary processes and show how they can be used to explain the increase in frequency of the emphatic form of negation observed in Jespersen's cycle. While these pressures are often assumed, they are not explicitly formulated (cf. Detges & Waltereit 2002, Hopper & Traugott 2003, Kiparsky & Condoravdi 2006, *inter alia*). In doing so, we also make explicit the information-theoretic foundation of grammaticalization processes such as bleaching.

The second contribution of this work is that while we follow recent examinations of the role of common and conflicting interests in pragmatic reasoning (Benz et al. 2006, Franke et al. 2012, De Jaegher & van Rooij 2013), using evolutionary game theory allows us to relax the assumption of rationality in the dynamics. That is, here we assume that individuals are *boundedly rational* agents with limited cognitive and informational resources, rather than the logically omniscient and computationally omnipotent *homo economicus* (Simon 1955). This is important insofar as language change is not the product of conscious introspection, deliberation, or calculation. Rather, it is a kind of invisible-hand phenomenon in the sense of Keller (1994), which arises from the myopic interactions of an entire population of individuals. Taking this approach, the dynamics of signaling in a population can be modeled as a simple kind of learning which has deep connections with one of the most well-studied *evolutionary game dynamics*.

The third and final contribution of this work is that we fit the resulting dynamic model to corpus data. Recent work using both evolutionary game theory, along with other mathematical and computational tools, has focused on explaining qualitative patterns in historical change (Schaden 2012, Deo 2015, Yanovich 2015, 2016, Enke et al. 2016). To our knowledge, this is the first quantitative approach to modeling historical change using a dynamic evolutionary game-theoretic model. This approach yields several benefits. First, it offers more points of contact between abstract models and empirical evidence. The fitted parameters of a model can be compared to corpus or experimental data, or guide the collection of new data. Second, this approach also brings us closer to the use of statistical tools within quantitative historical linguistics (Kroch 1989, Altmann et al. 1983) and the ability to directly compare and select between models (Burnham & Anderson 2003). Being able to quantify and compare formal dynamic models is a crucial step towards understanding language change.

The rest of the paper is organized as follows. In Section 2, we present a general overview of Jespersen's cycle, noting that the term is used to refer to two often-related but distinct processes: the *formal cycle* describes the forms of negation over time; the *functional cycle* describes how those forms are used to signal information over time. We note the logical relationship between the two cycles and focus on functional cycle as an inflationary process. In Section 3 we model the interaction between speakers and hearers, as well as their respective interests, as a *signaling game*. In Section 4 we determine the *evolutionarily stable strategies* of the signaling game, which correspond to speaker and hearer behaviors in a population that are resistant to change. In Section 5 we consider signaling in a population over time under varying degrees of conflicting interests under a particular *evolutionary game dynamics*. We fit the resulting model to corpus data from the functional cycle in Middle English and discuss the fitted parameters. In Section 6 we conclude by discussing the implications of the model and future directions for research.

2 Jespersen's cycle

Originally coined by Dahl (1979: 88), the term *Jespersen's cycle* is often used in reference to the following observation regarding sentential negation[2] in several languages made by Jespersen (1917: 4).

> The history of negative expressions in various languages makes us witness the following curious fluctuation: the original negative adverb is first weakened, then found insufficient and therefore strengthened, generally through some additional word, and this in its turn may be felt as the negative proper and may then in course of time be subject to the same developments as the original word.

However, this passage can and has been interpreted in two very distinct ways (van der Auwera 2009). The fundamental ambiguity stems from the fact that Jespersen noted both *formal* and *functional* patterns in the expression of sentential negation over time. Both patterns can be conceived of as cycles in their own right. That is, we can define a series of transitions from and back to states that are in some sense formally or functionally equivalent.

---

2 Sentential negation refers to the semantic property of negating an entire proposition, not just some subpart. It can be distinguished from morphological (e.g. *un-*, *non-*) and constituent negation (e.g. *John might have not understood*) using several diagnostics such as tag questions (Klima 1964) and performative paraphrases (Payne 1985). Sentential negation is also distinct from but related to the syntactic notion of standard negation (Miestamo 2005), which refers to constructions that can reverse the truth value of a proposition.

But, the term *Jespersen's cycle* is often used to refer to one or the other of these aspects, or both of them simultaneously. For example, the canonical presentation of what is referred to as Jespersen's cycle conflates these two uses (cf. Posner 1985, Schwegler 1988, Ladusaw 1993). Where parentheses at the second stage indicate an optional post-verbal element that is characterized as being emphatic, the following stages are posited.

1. NEG V

2. NEG V (NEG)

3. NEG V NEG

4. V NEG

At the start of the cycle, negation is expressed by a single pre-verbal element. At the second stage of the cycle an optional post-verbal element creates an emphatic bipartite form. At the third stage of the cycle the post-verbal element is obligatory and the bipartite form ceases to be emphatic. At the fourth stage we see the loss of the pre-verbal element resulting in a post-verbal form.

In this section we define and distinguish the two uses of the term intertwined in this representation, which we will call the formal and functional Jespersen cycles. In short, the distinction is between changes in the forms of negation available and changes in how those forms are used to signal information, respectively. First, we outline the formal aspects of how negation is expressed at the stages of the formal cycle. Second, we outline the function of those forms at different stages of the functional cycle. Finally, we note how the functional cycle falls into a broader class of inflationary processes observed in language change.

The formal cycle is defined in terms of the forms that are used to express negation over time, and consists of two transitions. We see the first stage of the formal cycle in the history of English with the pre-verbal *ne* in Old English, which expresses sentential negation alone.

(1)   Ic ne   secge
       I   NEG say
       (Old English)

This is followed by a transition to the bipartite form where another negative element is added. This is seen in Middle English, where *ne* is supplemented by *not*.

(2)   I ne   seye not
       I NEG say   NEG
       (Middle English)

6

The final stage in the formal cycle is a return to a single negative element. In fact, all three forms are present and overlap in Middle English. But, in Early Modern English the preverbal element in *ne...not* is lost, leaving the post-verbal form *not*.

(3) I say not
     I say NEG
     (Early Modern English)

So, the formal cycle consists of the increase and then decrease in the formal complexity of sentential negation. It is cyclic in the sense that the forms of negation at the start and end are of equal formal complexity. Before moving on, there are two important points to emphasize.

First, the emergence of *do*-support in Early Modern English yields a state parallel to Old English with a sole pre-verbal negator, but it is not a necessary component of the formal cycle.[3] That is, while negation in Present-day English may be structurally parallel to Old English, negation has been formally parallel since the loss of the bipartite form *ne...not*.

(4) I don't   say
     I do-NEG say
     (Present-day English)

As a point of comparison, we observe the same transition from pre-verbal *ne* to bipartite *ne...pas* to post-verbal *pas* forms of negation in French, but no subsequent transition back to pre-verbal negation. When it comes to the formal cycle it matters how much material is used to express negation, not necessarily where that material stands in relation to the verb.

Second, the formal cycle is not necessarily the transition from pre- to post-verbal negation. That is, the bipartite form is not necessarily formed through the addition of a post-verbal element. For example, in modern African American Vernacular English, negation can be supplemented by a pre-verbal element *eem*, which can also express negation in its own right (Jones 2015).

(5) You don't eem know.

(6) You eem know.

The formal cycle could just as well be from post- to pre-verbal negation. It is a contingent historical fact, arising from the syntax of English and the source of the

---

3 Jespersen (1917: 10) noted the uniqueness of these developments, which he attributed to a tendency to place negation at the beginning of the sentence to avoid confusion on the part of hearers. Yet, despite this purported tendency, the majority of languages that Jespersen noted, including his native Danish, persist in a supposedly perplexing state of purely post-verbal negation.

additional material, rather than some necessary property of the formal cycle. Again, it is the formal status rather than the structural position that is relevant.

While the formal cycle is defined by the forms of negation over time, the functional cycle is defined by how those forms are put to use, and is characterized by two transitions. The first transition occurs with the introduction of a stronger more emphatic negative form, often the result of adding a *negative polarity item* or other material to the original plain form (Horn 1989, Van Gelderen 2008, Givón 1978, Croft 1991). The initial effect of the bipartite form, in Jespersen's words (1917: 15):[4]

> ...[I]n most cases the addition serves to make the negative more impressive as being more vivid or picturesque, generally through an exaggeration, as when substantives meaning something very small are used as subjuncts.

The second transition of the functional cycle occurs as the new emphatic form increases in frequency, weakens in intensity, and replaces the original negative form. The functional cycle occurs when one form of plain negation is replaced by another form. It is cyclic in the sense that the number of functionally distinct forms of negation increases then decreases. For example, in the history of English and French an initially emphatic bipartite form weakens over time and comes to have the same force as the pre-verbal form.

There are two important points to be made with regard to the functional cycle and its relationship to the formal cycle. First, the conflation of the formal and functional cycles understandably stems from the fact that the functional cycle often coincides with the first transition of the formal cycle. Intuitively, *ne...not* is a more complex form than *ne*, and thus we would expect it have a more restricted and hence stronger meaning. Note that this does not apply to the second transition of the formal cycle given that the same relationship between *not* and *ne...not* does not hold. Second, while the functional cycle often takes place within the first transition of the formal cycle, it can occur entirely independently of the formal cycle.

For instance, one form can be replaced by another of equal formal complexity. Or, in Meillet's (1912: 134) estimation, the functional cycle is achieved when "one adds new *or* different words". Kiparsky & Condoravdi (2006) note that this is exactly what takes place in the history of Greek. Historical forms of negation in Greek are listed in Table 1, where emphatic negation is taken to be the semantically

---

4 Despite the evocative phrasing, Jespersen was certainly not the first to notice the trajectory of the functional cycle. van der Auwera (2009) suggests *Meillet's spiral* (1912: 394) as a potentially more appropriate term for the functional cycle, to which we might add *Gardiner's gyre* (1905: 134).

PLAIN	EMPHATIC	SOURCE
οὔ...τι	οὔ-δε...εν	Ancient Greek
(οὐ)δέν...τι	δέν...τίποτε	Early Medieval Greek
δέν...τίποτε	δέν... πρᾶμα	Greek Dialects
δέν...πρᾶμα	δεν...ἀπαντοχή	Modern Cretan

Table 1   Historical forms of plain and emphatic negation in Greek

stronger form in comparison to *plain* negation at any point in time.[5] The sources of the different forms are ordered chronologically by row.

There is a consistent transition of forms between the two functions: the emphatic negation of the last century or millennium becomes the plain negation of this century or millennium. Crucially, at least some of these functional cycles occur without any concomitant formal cycle. For example, if we compare the formal complexity from Early Medieval Greek onwards, they would all be equivalent. All of them consist of a shared pre-verbal element δέν along with a single post-verbal element. Thus, we see several bipartite forms come to express plain negation over time.

So, the formal and functional cycles are often closely intertwined, but they are distinct phenomena. This means that the facts to be explained for each are also distinct. Here we focus on the functional cycle rather than the formal cycle, which in the case of English, means that our goal will be to provide a model that explains the transition from *ne* to *ne...not*. In contrast, a model of the formal cycle would need to explain both this transition and the subsequent transition from *ne...not* to *not*. This point bears emphasis insofar as the two cycles are so often conflated. Addressing the functional cycle in isolation clarifies and lightens the explanatory burden. Here we only need to explain the single change from one form of plain negation to another.

In what follows we focus on the functional cycle and its relation to a broader class of inflationary processes under which the contexts of use for a linguistic form expand over time along some evaluative scale (Dahl 2001). For example, in the case of politeness, terms like *lady* and *gentleman* expand from the original meaning of landed gentry towards a generalized form of polite address. But, if everyone is addressed as a lady or a gentleman, then the forms cease to convey information about the speaker's actual perception of or attitude towards the hearer's status. This is analogous to the functional cycle. The bipartite form in English may initially be

---

5 We omit some of the emphatic forms for a concise presentation (cf. Kiparsky & Condoravdi 2006: 1). Whether there any form is ever *the* emphatic form depends on what we mean by *emphasis*, a point we return to in Section 3. Crucially, however, it is only ever a single emphatic form that displaces the old plain form.

emphatic, but as it becomes obligatory it ceases to be. For any form, if it is the only one in use, then it cannot carry any special meaning. There is nothing else to be special in comparison to. As Kiparsky & Condoravdi (2006: 5) rightly put it, "to emphasize everything is to emphasize nothing."

Now, while we can conceive of the functional cycle as an inflationary process, this still leaves us with the question of *why* the inflation happens. In this regard, the case of politeness is again a useful point of comparison. The increased use of the restricted forms of address can be understood in social terms. There are no mechanisms that prevent using *lady* or *gentleman* in conversation, and indeed, using it to flatter an addressee may come with distinct benefits. The increase in frequency of these forms can be understood in terms of the social impact of their use. We argue that the functional cycle can be understood in similar terms. In the next section we make clear what might be meant by *emphasis*, what using emphatic forms signals about speakers, and what social consequences that information might have.

3   Signaling

The notion of emphasis is central to understanding the functional cycle as an inflationary process and has generally been defined by two functions. The first is that emphatic negation widens and strengthens negation to preclude exceptions (Kadmon & Landman 1993). That is, emphatic negation signals a stricter *standard of precision* for the assertion of a proposition (Krifka 1995). The second is that emphatic negation serves to deny or counter an expectation (Detges & Waltereit 2002, Kiparsky & Condoravdi 2006). In this section we argue that these two functions can be modeled and understood as two interrelated aspects of a *signaling game* (Lewis 1969). We define the components of the game in the abstract, then turn to how we can use these to model emphatic negation in the functional cycle.

A signaling game is played between a sender and a receiver, which in this case correspond to a speaker and a hearer. It has the following basic structure. First, the sender observes some private information, $t \in T$, drawn according to a prior distribution, p. This information is referred to as the sender's state and can be thought of as some information about the world. Next, the sender chooses a message, $m \in M$, to send to the receiver according to strategy that maps states to messages, $s : T \to M$. Finally, the receiver takes some action, $a \in A$, in response to the message according to a strategy that maps messages to actions, $r : M \to A$. Both senders and receivers have preferences over combinations of states and actions, which are represented by the utility functions that map the combination of states and actions to real numbers, $U_S : T \times A \to \mathbb{R}$ and $U_R : T \times A \to \mathbb{R}$, respectively. Senders and receivers prefer outcomes that yield higher utilities according to these functions.

Signaling games offer a natural model of communication, but their generality requires us to be clear about what each component represents in the case of emphatic negation and the functional cycle. First, it is crucial to specify what kind of private information a speaker might have. When it comes to negation, even if we assume that speakers abide by the maxim of quality and avoid saying what they know to be false, this still only partially captures how we actually assert and reason about propositions. As Austin (1962: 142) put it:

> Suppose that we confront 'France is Hexagonal' with the facts, in this case, I suppose, with France, is it true or false? Well, if you like, up to a point; of course I can see what you mean by saying that it is true for certain intents and purposes. It is good enough for a top-ranking general, perhaps, but not for a geographer.

Indeed, the felicitous use of certain forms requires more or less strict *standard of precision* on the part of speakers (Lewis 1979, Landman 1991, Krifka 1995). That is, if speakers use an emphatic form of negation, they are expected to have strong evidence to support such a precise assertion.

For example, imagine a cook for a large group of people being asked and responding to the following questions (cf. Kadmon & Landman 1993: 360).

(7)   a.   Will there be french fries tonight?
        b.   No, I don't have any potatoes.

(8)   a.   Maybe you have just a couple of potatoes that I could fry in my room.
        b.   No, I don't have any potatoes at all.

The cook's response in (7b) is compatible with the cook knowing that there are a few potatoes, but not enough for the whole group. The response in (8b) is only compatible with the cook knowing that there are absolutely no potatoes available. The standards of precision assumed by the plain and emphatic responses in (7b) and (8b) are naturally ordered along a scale. The standard of precision for asserting (8b) entails that of (7b), but not vice versa.

In what follows we take the speaker's private information to be a standard precision drawn from a scale, $T : [0,1]$. Here we take the state t_0 to be the minimum standard of precision required to truthfully assert a proposition and the state t_1 to be the strictest standard of precision possible. That is, we assume that speakers overwhelmingly abide by the maxim of quality, but vary within certain bounds as to how strictly they observe it. So, we might think of t_1 as corresponding to the case where the speaker knows that there are absolutely no potatoes, and $t' < t_1$ being a case where the speaker knows that there are perhaps a few potatoes. While (7b) would be felicitous given t_1 or t', (8b) would only be felicitous given t_1.

Regarding the functional cycle in English, then, a speaker's strategy consists of the mapping from different standards of precision to the two forms m_{ne} and $m_{ne...not}$. If we take emphasis to mean that a form is used with the highest standards of precision, then the notion that emphasis also serves to counter some prior expectation follows naturally. In fact, the notion of emphasis can be expressed in information-theoretic terms by comparing the *information gain*, or *Kullback-Leibler divergence* (1951), of plain and emphatic forms. Where $p(t \mid m)$ is the conditional probability of states given a message, the information gained by receiving a message is the following.

$$(1) \qquad KL(m) = \int_0^1 log\left(\frac{p(t \mid m)}{p(t)}\right) p(t \mid m) dt$$

If the form m_{ne} is used with all standards of precision, then the conditional probability of states given the message, $p(t \mid m_{ne})$, is equivalent to the prior probability over standards of precision, $p(t)$. It follows directly that $KL(m_{ne})$ approaches 0 since $log(1) = 0$. In other words, upon receiving m_{ne} hearers will know that negation was used, but will not have any additional information about how precisely it was used. That is, nothing will have changed from their prior expectations about how precise the speaker was being. In contrast, if the form $m_{ne...not}$ is restricted to use with high standards of precision, then it carries information about how precisely negation is being used. The conditional probability of states given the message is distinct from the prior, $p(t \mid m_{ne...not}) \neq p(t)$, and thus the message carries information, $KL(m_{ne...not}) > 0$. We can conceive of emphasis as a description of highly informative signals.

In other words, receiving $m_{ne...not}$ substantially shifts the hearer's expectations about how precise the speaker is being. So, both functions of emphasis that have previously been noted are straightforwardly related when we define the sender's state as a scale of standards of precision. Indeed, this definition of emphasis as information also offers a natural interpretation of the intuitive relation between the frequency of the forms and bleaching. As $m_{ne...not}$ increases in frequency, the conditional probability of states given the message necessarily approaches the prior. When $m_{ne...not}$ is the only form in use it ceases to carry any information because it simply cannot shift the prior expectation of hearers.

Now that we have defined the components of the signaling game that correspond to the speaker's states and strategies, we turn to the hearer. A hearer's strategy consists of the mapping from m_{ne} and $m_{ne...not}$ to a set of actions. The crucial point to clarify is what these actions correspond to. In this regard it is useful to note two ways of conceptualizing the purpose of communication. The first is that communication consists of the transmission of information from speaker to hearer. However, as

Franke et al. (2012) note, if the interests of speakers and hearers diverge, then sharing information presents a paradox. Various mechanisms might serve to bolster this conception of language, but they only succeed insofar as we actually observe the features of communication that they predict.

For example, sharing information might be warranted between kin, whose interests are overwhelmingly aligned (Hamilton 1963, 1964). But, language is certainly not restricted to communication between family members. Indeed, communication might be more likely to take place between unrelated individuals in relationships of reciprocal information sharing (Trivers 1971). However, such an arrangement would require that speakers monitor their interlocutors to prevent *free-riding*. That is, speakers would be expected to chastise interlocutors for not making substantively informative contributions to a conversation. This is the exact opposite of what we actually observe. Thoughtful listeners are commended for being attentive and polite, whereas long-winded speakers are chastised unless they make substantive contributions. In other words, hearers monitor speakers for informative contributions rather than the other way around.

Dessalles (2007) argues that we can make sense of this fact if we conceptualize the purpose of communication not just as the transmission of information per se, but as the transmission of information in exchange for social status. When hearers listen to what a given speaker has to say they increase the speaker's status insofar as they choose to listen. That is, given that hearers have a choice this choice acts as a kind of advertisement for speaker's ability to gather and reason about important information (cf. Gintis et al. 2001, Dessalles 2014). The longer and more keenly others listen, the better an indication that a speaker is saying something relevant, and the more social benefits accrue to the speaker. In what follows we take the set of actions available to hearers to constitute a scale, $A : [0, 1]$, which correspond to the amount of time and attention paid to the speaker by the hearer in a given exchange.

If we conceive of communication as the exchange of information for social status, the final part of the signaling game to specify is the preferences of speakers and hearers over the exchange rate. Intuitively, hearers want the best return on their investment of time and attention in the form of information about the world. Hearers want to spend as much attention as necessary to gain information from a given standard of precision, no more and no less. In contrast, speakers are biased towards accruing social status, regardless of the actual standard of precision. In short, the preferences of speakers and hearers diverge depending on the strength of this speaker bias. We can model the difference in the preferences of speakers and hearers using

the following utility functions, modified slightly from Crawford & Sobel's (1982) seminal work on information transmission.[6]

(2)
$$U_S(t,a) = 1 - (a - t - (1-t)b)^2$$

(3)
$$U_R(t,a) = 1 - (a - t)^2$$

Hearers prefer that the standard of precision and the amount of attention invested match in some sense, which is exactly the case where $a = t$. In contrast, speakers prefer that the amount of attention invested exceeds the level warranted by the standard of precision, $a = t + (1-t)b$. The degree of speaker bias is captured by the parameter b. For the special case where $b = 0$, the interests of speakers and hearers are perfectly aligned and both want the action taken by the hearer to match as closely with state of the speaker. Speakers want a fair exchange of information for status. As b grows, however, the speaker prefers higher actions on the part of hearers. That is, speakers want hearers to grant them more status for a given standard of precision than hearers would prefer.

So far we have defined all of the components of the signaling game for the case of emphatic negation in the functional cycle. The speaker observes some information about the world which corresponds to a standard of precision and chooses a message from the forms m_{ne} and $m_{ne...not}$ according to a strategy. In response to this message, the hearer takes an action which corresponds to investing some amount of time and attention with the speaker. Hearers prefer an equal exchange of information for the status granted by this attention, but speakers are biased towards a higher rate of attention for a given standard of precision. Note that this description of the signaling game does not in and of itself predict the behavior of speakers and hearers. To do so, we need to supply a *solution concept* that we can use to determine the equilibria of the game.

4   Equilibrium

With the components of the game defined, we now turn to analyzing its properties in order to understand the conditions for the functional cycle. In particular, we

---

6 This formulation simply constrains the preferences of speakers and hearers to fall within the unit interval. Also, note that the bias parameter can be interpreted as the expected value of a bias conditioned probabilistically on the state, $\mathbf{b} = E[b \mid t] = \int b \cdot p(b \mid t)db$. So, speakers might vary in the amount of bias exhibited in any given interaction, but exhibit a certain amount of bias on average. Here we simply assume that there is a linear relationship between the state and the expected degree of speaker bias.

want to know what speaker and hearer strategies are *evolutionarily stable strategies* (Maynard Smith 1982), behaviors that are resistant to change, and which strategies are susceptible to change. We begin by defining speaker and hearer strategies, the expected utility of different strategies given the prior probability over states, and then determine the evolutionarily stable strategies of the game based on the expected utilities. Finally, we note two important points regarding the relationship between speaker bias and the functional cycle. First, if speaker bias is sufficiently large, then only a single message can be used in equilibrium. Second, this single message equilibrium is not evolutionarily stable since it can always be disrupted by the introduction of an appropriately conditioned signal. We discuss the results in the context of the functional cycle.

To begin we define the set of speaker and hearer strategies. The set of speaker strategies is all potential mappings from the unit interval to a discrete set $S : [0,1] \rightarrow M$. This is problematic given that the domain is uncountable. To simplify things we use the following condensed representation. Let $\mathscr{P}_n(T)$ be a partition of the state space into n equal length subintervals $t_0 = 0 < t_1 < ... < t_{n-1} < t_n = 1$. For each properly defined subinterval, (t_{i-1}, t_i) the speaker uses the message m_i. A speaker's strategy is then a function from this partition to messages $S : \mathscr{P}_n(T) \rightarrow M$. Intuitively, this is simply a way of carving up the state space into discrete contiguous regions and using those regions to determine which signal to send. For example, we will deal with the case of two messages $\mathscr{P}_2(T)$, where the speaker uses m_{ne} for $t \in (0, t_1)$ and $m_{ne...not}$ for $t \in (t_1, 1)$. The set of hearer strategies is all potential mappings from the set of messages to the unit interval $R : M \rightarrow [0,1]$. Since the domain is finite, this is more straightforward than the set of speaker strategies. For each message m_i the hearer takes an action a_i. So, for example, a_{ne} would be the hearer's response to message m_{ne}, and $a_{ne...not}$ would be the hearer's response to message $m_{ne...not}$.

So, a speaker with strategy s observes a state t, and assuming that $t < t_1$, sends m_{ne}. In response to this message a hearer responds according to her strategy r and takes an action a_{ne}. In less abstract terms, a speaker observes some fact about the world corresponding to a standard of precision, uses either a plain or emphatic form of negation, and the hearer responds by paying a certain amount of attention to the speaker. While we can describe a given outcome in these terms, we are interested in how well speaker and hearer strategies do on average, we want the expected utilities of speaker and hearer strategies. With slight abuse of notation, for a pair of speaker and hearer strategies $\langle s, r \rangle$ these are given by the following.

(4) $$E[U_S(s,r)] = \int_0^1 \left(1 - (r(s(t)) - t - (1-t)b)^2\right) p(t)dt$$

$$(5) \qquad E[U_R(s,r)] = \int_0^1 \left(1 - (r(s(t)) - t)^2\right) p(t) dt$$

We obtain these expected utilities by doing three things. First, for a given state, we determine the message sent by the speaker according to his strategy, $s(t)$, and the action taken by the hearer in response to this message by the hearer according to her strategy, $r(s(t))$. Second, we substitute these actions into the utility functions we defined in Equations (2) and (3) to determine the outcome for a given state. Finally, we weight these outcomes according to the prior probability over states. These expected utilities quantify how well speakers and hearers using particular strategies do in bringing about their preferred outcomes on average.

Now, in order to calculate the expected utilities we need to define the prior distribution that determines how likely speakers are to have a given standard of precision. We model the prior distribution over standards of precision as a *beta distribution*, often noted as $\mathscr{B}(\alpha, \beta)$ where $\alpha, \beta > 0$ are parameters that control the shape of the distribution.[7] In what follows we assume the prior is of the general form $\mathscr{B}(1, \beta_p)$, where $\beta_p > 1$, which is meant to capture two facts. The first fact is that as boundedly rational agents, we are often uncertain about the actual state of the world. That is, we are less often, if ever, certain of all of the facts. Thus, we are rarely in the position to support a high standard of precision with evidence. A larger β_p simply corresponds to the prior probability being skewed towards low standards of precision, reflecting our uncertainty.

We can visualize the prior distribution over standards of precision for various values of β_p as in Figure 1. For $\beta_p = 1$, we have a uniform distribution over standards of precision. For $\beta_p = 2$ the distribution is skewed towards lower standards of precision, and as we increase β_p further it becomes even more skewed towards these lower standards of precision. The crucial intuition is that the two shape parameters allow us to model a wide range of distributions. In what follows we use beta distributions extensively, so it is useful to note that we could achieve symmetric results in the opposite direction by holding $\beta = 1$ constant and increasing α.

The second fact that this asymmetric prior allows us to continue defining emphasis as information while maintaining the distinction between emphatic and attenuating forms. (Israel 2011). That is, there are emphatic negative polarity items (e.g. *a wink*, *an inch*, *at all*) as well as emphatic positive polarity items (e.g. *tons of*,

---

7 As a special case, the uniform distribution over the interval corresponds to $\mathscr{B}(1,1)$. The expected value of a beta distribution is given by $\frac{\alpha}{\alpha+\beta}$, so the expected value of a random variable uniformly distributed over the unit interval is as we would expect $\frac{1}{2}$. The variance of the distribution decreases as α and β grow. So, $\mathscr{B}(10,10)$ would have the same expected value as $\mathscr{B}(1,1)$, but would be more bunched up around the expected value of $\frac{1}{2}$.
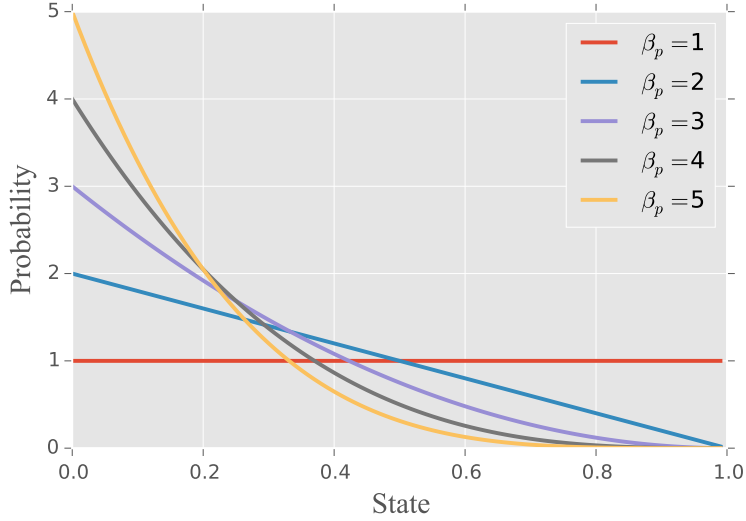
Figure 1   Prior distribution over standards of precision as a beta distribution $\mathscr{B}(1, \beta_p)$ for various values of β_p.

utterly, *totally*) that are both used with the highest standards of precision. Note that these emphatic forms pick out the same standards of precision because NPIs occur in scale-reversing contexts (Fauconnier 1975). In contrast, there are also attenuating negative polarity items (e.g. *much*, *long*, *all that*) and attenuating positive polarity items (e.g. *a little bit*, *sorta*, *somewhat*) that are used with the lowest standards of precision. Under a uniform distribution, if these sets of standards of precision are roughly equal in size, then both emphatic and attenuating forms are equally informative insofar as they shift the prior in equal but opposite directions. However, if the low standards of precision associated with the attenuating forms are more likely to begin with, then the attenuating forms are less informative. This is not to say that attenuating forms are not informative at all, but rather that emphatic forms are *highly* informative. So, allowing for an asymmetric prior allows us to capture both the informational reality of boundedly rational agents and the definition of emphatic forms as highly informative.

   With the expected utilities defined, we might ask whether particular strategies constitute *evolutionarily stable strategies* (Maynard Smith 1982). These evolutionarily stable strategies correspond to speaker and hearer behaviors that are resistant to change, and meet the Gricean criterion of being reasonable to follow rather than abandon. For asymmetric games, such as signaling games where players have specific roles like speaker and hearer, the evolutionarily stable strategies correspond to

the *strict Nash equilibria* of the game (Selten 1980). As a point of reference, a pair of speaker and hearer strategies is a Nash equilibrium if neither speaker nor hearer would do better by unilaterally changing behavior. Such an equilibrium is *strict* if both speaker and hearer would do worse by unilaterally changing behavior. If strict Nash equilibria are behaviors that jointly strict maximize the utility functions, then they can be determined by solving a system of partial differential equations of the utility functions. In Figure 2 we plot the Nash equilibria strategies of speakers and hearers as a function of speaker bias, where the prior probability is taken be a beta distribution over standards of precision $p(t) \sim \mathscr{B}(1,2)$.[8] In what follows we evaluate whether or not given Nash equilibria of the game are strict and thus evolutionarily stable strategies.

The horizontal axis of Figure 2 represents the degree of speaker bias, where $b = 0$ indicates the case where speakers are interested in a fair exchange of information for attention and $b > 0$ indicates an increasing bias towards higher rates of exchange. The vertical axis represents the actions and states taken by speakers and hearers in equilibrium for a given amount of bias. The solid line indicates the point at which speakers partition states, below which speakers send m_{ne} and above which they send $m_{ne...not}$. The dashed lines indicate the hearer response to a_{ne} and $a_{ne...not}$. We can interpret this figure by fixing a value of b and examining how speakers partition the state space and how hearers respond to the messages. For example, when $b = 0$, speakers partitions states at $t_1 = .3819$, sending m_{ne} for $t \in (0, .3819)$ and $m_{ne...not}$ for $t \in (.3819, 1)$. In response, hearers take actions $a_{ne} = .1759$ and $a_{ne...not} = .5879$.

There are two important things to note about these results. First, if speaker bias is sufficiently high, then only a single message is used in equilibrium. That is, for all $b > \frac{1}{6}$ only $m_{ne...not}$ will be used in equilibrium. When speaker bias is this large the form carries no information, and the best response for speakers is the action that corresponds to the expected value of the prior. In this case, we see that for sufficiently high speaker bias the best response is $a_{ne...not} = \frac{1}{3}$. Second, if speaker bias is sufficiently high, this single message equilibrium is not a strict Nash equilibrium and thus is not evolutionarily stable. That is, it can be invaded by strategies corresponding to other behaviors. To see why this is the case note that if only $m_{ne...not}$ is used by speakers, then hearers' response to m_{ne} is free to vary without affecting the expected utility of speakers or hearers, and thus the single message equilibrium is not evolutionarily stable. In fact, a single message equilibrium is not *neologism-proof* (Farrell 1993) in the sense that it can always be disturbed by the introduction of a new message used with standards of precision higher than the expected value of the prior.

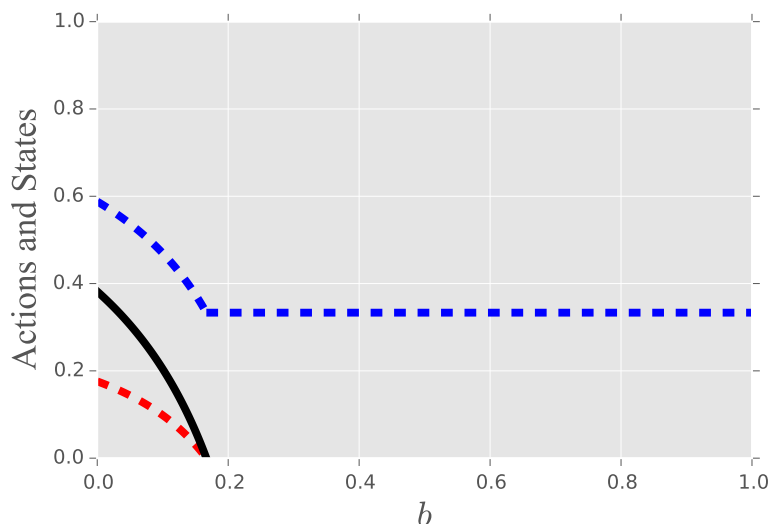8 See the supplementary material for the details of calculating these equilibria.

Figure 2   Nash equilibrium solutions for two messages for values of bias with prior distribution $p(t) \sim \mathscr{B}(1,2)$

So, if speakers are sufficiently biased when it comes to the exchange of information for attention, then using two forms is never a component of a strict Nash equilibrium and thus never evolutionarily stable. Yet, a single form can always be displaced by an appropriately conditioned form. This means that the functional cycle can always be set in motion with the introduction of the right kind of message. However, if this is indeed the case, we might ask ourselves why the functional cycle is not constantly occurring. Given the vast number of negative polarity items available in any given language it seems that there would always be new emphatic forms of negation available to speakers (cf. Horn 1989). If these messages were indeed available, then we would expect functional cycles one right after another rather than separated by centuries.

One potential explanation for this discrepancy is that there are limits on what counts as an available emphatic form for the purposes of the functional cycle. A reasonable restriction is that new potential emphatic forms must be free from selectional constraints. For example, the contrast between *step* with verbs of motion and others demonstrates this distinction.

(9)   a.   #I didn't sleep a step.
        b.   #I didn't eat a step.

The reason that new emphatic forms are not always available to set the functional cycle in motion is precisely because they are semantically anomalous. At some point a new emphatic form is freed from these restrictions. For example, in the history of French, *pas* 'step' lost its restrictions and became part of the bipartite form, but we leave further investigation of this process to future work. Indeed, one means of exploring the genesis of emphatic forms might be to examine the mutual information of classes of verbs and negative polarity items over time (cf. Danescu-Niculescu-Mizil & Lee 2010)

We have determined the evolutionarily stable strategies of the population, but this notion is an essentially static concept. That is, we can reason about what would happen if we started at a particular state, but not whether we will ever reach that state in the first place. More importantly, it does not allow us to understand how a population evolves in a particular historical change. We must posit a process that underlies how speakers and hearers interact and respond to each other. Doing so will allow us to examine how different degrees of speaker bias impact the trajectory of meaning.

5   Dynamics

In the previous section we examined the equilibrium properties of speaker and hearer strategies in the signaling game as we varied how biased speakers are in the exchange of information for attention and status. If speakers are sufficiently biased, then the functional cycle can be set in motion with the introduction of a new emphatic signal. In this section we turn to the details of the motion of the functional cycle by providing *evolutionary game dynamics* that define how a population of speakers and hearers changes over time (Hofbauer & Sigmund 1998). We begin by discussing the *replicator dynamics* (Taylor & Jonker 1978) as an appropriate evolutionary game dynamics for studying changes in meaning. We then turn to data from the functional cycle in a parsed corpus of Middle English. Finally, we fit a dynamic model to this data and discuss the resulting parameters.

The replicator dynamics were originally introduced as an explicitly dynamic model of biological replication. The fundamental intuition underlying them is that strategies that have a higher than average expected utility should increase in prevalence in a population, and that strategies with lower than average expected utility should decrease. However, this simple idea has since been shown to have deep connections with some of the most widely-studied models of learning. In particular, Börgers & Sarin (1997) prove that if agents interact frequently and change their behavior slowly then the asymmetric replicator dynamics are equivalent to the expected behavior of agents playing an asymmetric game while learning according

to a simple form of learning (cf. Bush & Mosteller 1955, Sutton & Barto 1998). That is, if speakers and hearers tend to do things more if they yield higher utility, then their expected behaviors can be modeled by the replicator dynamics. There are, of course, a few conceptual points and clarifications to be made to justify the use of the replicator dynamics in modeling the functional cycle.

First, given that negation is one of the most frequently aspects of language it is safe to assume that it is used frequently and speakers and hearers do not dramatically alter their probability of use or interpretation of forms from one point in time to the next. Second, we assume that each individual acts as a speaker and hearer, but cannot introspectively reason about the impact of one strategy on the other. That is, individuals cannot use their behavior as speakers to change their own behavior as hearers, nor vice versa. Third, and most importantly, since the replicator dynamics correspond to the expected behavior of individual learners, they can also be interpreted as the expected or *mean dynamics* of an entire population. That is, In this case of the functional cycle, the replicator dynamics are not a model of individual behavior, but of the aggregate behavior in a population.

Now, choosing an evolutionary game dynamics allows us to investigate the effect of speaker bias on the functional cycle in the abstract, but we are really interested in how the resulting model can be applied to the actual historical trajectory of negation. In particular, we are interested in what happens when we fit the model to data from the history of negation in English. Towards this end, we fit the resulting model to data drawn from the second edition of the Penn Parsed Corpus of Middle English (Kroch & Taylor 2000).[9] The data are shown in Figure 3. The horizontal axis represents years, spanning Middle English from 1100 to 1500 CE. Each circle represents tokens in a given document. The size of the circle represents the number of tokens. The height of the circle represents the proportion of those instances that are a particular form. Locally-weighted regression lines are fit to these proportions. We see the transition from *ne* to *ne...not* starting around the 12th century, followed closely by the transition from *ne...not* to *not* in the 14th century.

Since we are interested in the functional cycle, we only care about the transition from *ne* to *ne...not*. While the subsequent rise of *not* represents the second transition in the formal cycle, it is not a part of the functional cycle. Here we treat tokens of *not* as if they were tokens of *ne...not*. This avoids the problem of excluding these tokens and attributing meaning to small fluctuations between *ne* and *ne...not* after the middle of the fourteenth century. More importantly, it captures the contingency of the second transition of the formal cycle to purely post-verbal negation. That

---

9 Following Wallage (2008) and Ecay & Tamminga (2015), we use negative declarative tokens, but exclude contracted forms, negative concord, and cases that appear to be constituent negation, among other cases that pattern in a substantially different manner. The code for generating and processing the queries can be found at: https://github.com/christopherahern/digs15-negative-priming
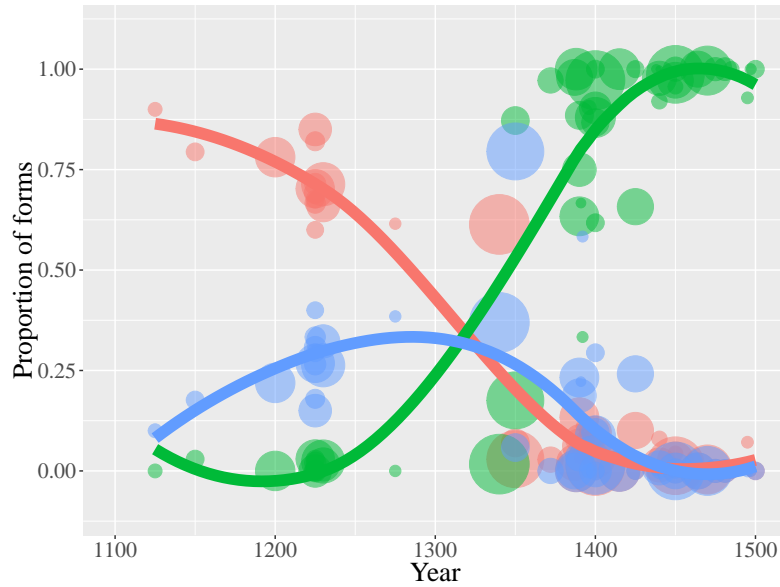
Figure 3 Proportion of *ne*, *ne...not*, and *not* in negative declaratives over time in the PPCME

is, the rise of *not* is not a part of the functional cycle, nor is it a necessary and immediate consequence of the functional cycle. We only need to compare the history of negation in French where the embracing form goes to completion before being eventually replaced by the post-verbal form. This will allow us apply the same model across languages without regard to subsequent contingent developments. The results of doing so are shown in Figure 4, where the horizontal axis represents the same time span, but the vertical axis represents the proportion of bipartite and post-verbal *ne...not* and *not*. Again, each circle represents an individual document whose size corresponds to the number of tokens.

We cannot evaluate the standard of precision for any given token of negation, the trajectory of forms in Figure 4 constitutes the data available to fit our model, and to fit the model, we need to specify its parameters. In particular, we need to define the initial state of how speakers use the different forms and how hearers respond to them. From there we can simulate the dynamics and adjust these parameters to find the most likely parameters given the data. In fact, we have quite a bit of information regarding what the initial state of the functional cycle actually is. That is, we know that $m_{ne...not}$ is fairly infrequent and largely restricted to high standards of precision. Likewise, we know that hearers' response to $m_{ne...not}$ is also largely restricted to actions corresponding to high standards of precision. We can
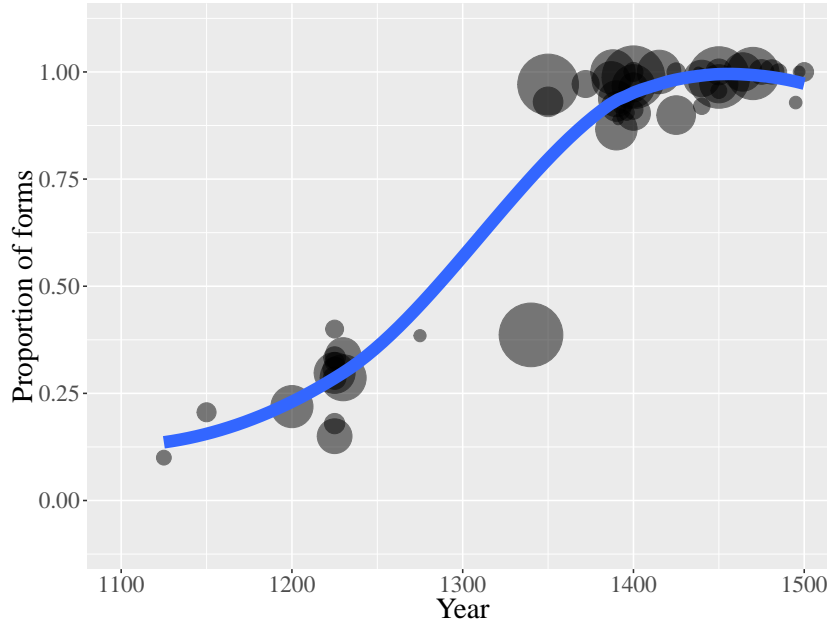
Figure 4   Proportion of emphatic *ne...not* and *not* versus *ne* in negative declaratives over time in the PPCME

translate this information into conditions on the initial states of the speaker and hearer populations.

Regarding speakers we assume that both forms have a particular meaning, which is captured by conditional probability of states given a form. First, we assume that $m_{ne...not}$ is largely used in states with high degrees of activation. In particular, we assume that it satisfies the following conditional distribution $p(m_{ne...not} \mid t) \sim \mathscr{B}(\alpha_s, 1)$. This guarantees that speakers use $m_{ne...not}$ for all standards of precision, but more with higher standards of precision. As we noted above regarding Figure 1, keeping $\beta = 1$ and increasing α_s pushes this distribution towards using higher and higher standards of precision. From this distribution and the prior probability over states, we can calculate the conditional probability of standards of precision given the message according to Bayes' rule, $p(t \mid m_{ne...not}) = \frac{p(m_{ne...not}|t)p(t)}{\int_t p(m_{ne...not}|t)p(t)dt}$. The larger that α_s is, the more likely that $m_{ne...not}$ is used with high standards of precision. In contrast, m_{ne} is the default form and does not carry any information above and beyond the prior. Since it is used almost evenly across all standards of precision it roughly satisfies the following conditional distribution $p(t \mid m_{ne}) \sim \mathscr{B}(1, \beta_p)$. Here β_p is the shape parameter of the prior, which determines its skew towards lower standards of precision as we showed in Figure 1. The last parameter to be fit for speakers is the degree of speaker bias b.

23

In all, then, we have three parameters, α_s, β_p, and b, to fit to model the initial state of and subsequent changes to the speaker population. To put these in perspective, the parameter α_s allows us to capture the fact that $m_{ne...not}$ is initially restricted to the highest standards of precision, and is thus emphatic. The parameter β_p allows us to capture the fact that we are boundedly rational agents with limited informational resources and thus often find ourselves in low standards of precision. The parameter b allows us to capture the fact that despite often having very little information as speakers, we prefer that our hearers pay attention to us.

In fact, these three parameters are sufficient since we can reasonably define the initial state of the hearer population with them. Intuitively, we want hearers to have reasonably accurate responses to the two forms. This can be achieved by ensuring that expected value of the response to each form corresponds to the expected value of the conditional probability of states given the form. For m_{ne} this is satisfied by infinitely many distributions, including the prior, so for simplicity we take the probability of an action given m_{ne} to be equivalent to the prior $p(a \mid m_{ne}) \sim \mathscr{B}(1, \beta_p)$. For $m_{ne...not}$, let the expected value of the standard of precision given the message, $\gamma = \int_t t \cdot p(t \mid m_{ne...not}) dt$, which is determined by the parameters α_s and β_p. Hearers have a reasonable response to $m_{ne...not}$ for any conditional probability of actions given the message $\mathscr{B}(\alpha, \beta)$ such that $\alpha = \left(\frac{\gamma}{1-\gamma} \right) \beta$. Again, for simplicity, we take the conditional probability of actions to be $p(a \mid m_{ne...not}) \sim \mathscr{B}\left(\left(\frac{\gamma}{1-\gamma} \right), 1 \right)$. In all then, we have no additional parameters beyond those needed for speakers to fit in order to capture the initial state of and subsequent changes to the hearer population.

To simulate the dynamics we discretize the set of states and actions for speakers and hearers.[10] That is, for some n, we treat the set of states $T : \{t_0, ..., t_n\}$ and actions $A : \{a_0, ..., a_n\}$, where $t_i = a_i = \frac{i}{n}$. In this case, we use one hundred states and actions, $n = 100$, and use *beta-binomial* distributions over these states and actions, which are the discrete analogue of the beta distribution. We use the discrete-time behavioral representation of the replicator dynamics suggested by Hofbauer & Huttegger (2015), which treats each state as its own population in which messages compete with each other, and each message as its own population in which actions compete with each other. This allows us to reduce the dimensionality of the system and simplifies things conceptually and computationally.

We fit the dynamic model to the data from the functional cycle by doing the following. First, we specify the parameters of the initial condition, α_s, β_p, b. Then we simulate the replicator dynamics for the four hundred years from 1100 to

---

10 For our purposes, this discretization allows us a tractable means of simulating and fitting the dynamics to data. Some analytical results can be derived regarding the details of the replicator dynamics in continuous strategy spaces, but are more complicated and ultimately rely on numerical simulations for assessing stability. See Oechssler & Riedel (2001, 2002), Jäger et al. (2011) for relevant discussion.

1500, calculating $p(m_{ne...not})$ over time.[11] Third, we calculate the likelihood of the parameters given the data. Finally, we maximize the likelihood of the parameters using the bounded *L-BFGS* algorithm (Byrd et al. 1995). This allows us to specify facts about the model such as the speak bias being constrained, $b \in [0, 1]$.

The maximum likelihood parameters for the model are the following. Regarding the initial use of the incoming form, $\hat{\alpha}_s = 4.728$, $\hat{\beta}_p = 13.793$. As a point of comparison, we can think about these parameters in terms of Figure 1. We noted above, that if we keep $\beta = 1$ and increase α_s then the distribution is more and more skewed to the right. The fact that $\hat{\alpha}_s = 4.728$ simply means that at the beginning of the functional cycle speakers are more likely to use $m_{ne...not}$ with high standards of precision. The fact that $\hat{\beta}_p = 13.793$ simply means that the prior distribution over standards of precision is highly skewed towards low standards of precision. That is, we very rarely find ourselves in a position to make strong claims, we are often uncertain about the world. The fact that $\hat{b} = 0.270$ means that our results agree with the static equilibrium model presented in Section 4. That is, the degree of speaker shown in Figure 2 that would lead to the functional cycle, $b > \frac{1}{6}$, is consistent with the fitted parameter. At the very least, the parameters of the dynamic model justify our intuitions regarding the equilibrium predictions, and demonstrate the importance of both perspectives when considering historical instances of change.

We can visualize these results by simulating the replicator dynamics using these parameters, and showing the use of $m_{ne...not}$ over time as in Figure 5. The horizontal axis represents years and the vertical axis represents the probability of use for $m_{ne...not}$. As a point of comparison, the results of the dynamic model largely resemble those presented in Figure 4, which shows a locally-weighted fit to the data.

Perhaps more importantly, we can actually inspect the inner workings of the model as they relate to the functional cycle. That is, we can examine the information carried by the emphatic form over time as in Figure 6. The horizontal axis represents standards of precision and the vertical axis represents the conditional probability of standards of precision given that $m_{ne...not}$ was used. We show this conditional probability at the turn of each century from 1100 to 1500 as the functional cycle proceeds, where the dashed line indicates the prior probability over standards of precision. The initial meaning of the incoming emphatic form at 1100 is represented by the rightmost curve. This conditional probability of states given $m_{ne...not}$ is distinct from the prior, with a much higher expected value. At this point in time the incoming emphatic form carries the most information. At 1200, the conditional probability of states given $m_{ne...not}$ is still distinct from the prior, but its expected value is not as far away. That is, it carries less information. As time goes on, $m_{ne...not}$ spreads to

---

11 The replicator dynamics evolve in abstract time units. We compared the fit of several ratios of abstract time units per year. The best fit was obtained where each abstract time unit of the replicator dynamics corresponded to a year. See the supplementary materials for the code and details.
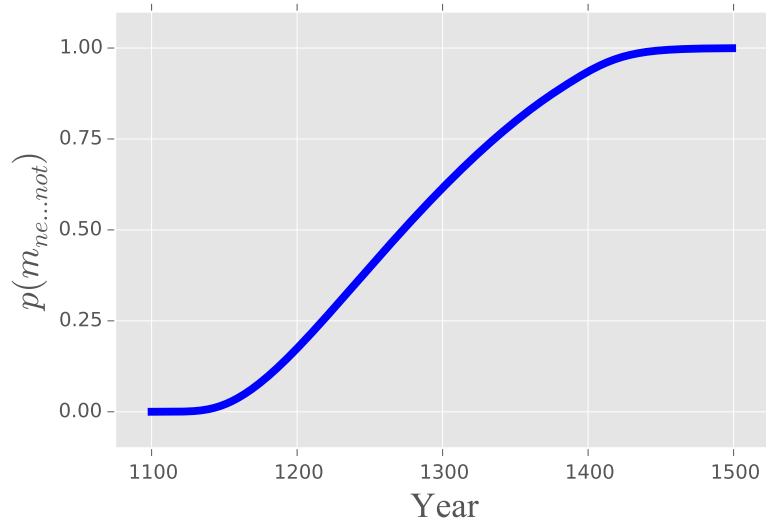
Figure 5   Predicted probability of *ne...not* over time for fitted model of functional cycle.

more and more standards of precision as the form increases in frequency. By 1300 $m_{ne...not}$ still carries information, but not much and by 1400 the probability of states given $m_{ne...not}$ is almost identical to the prior. We represent this loss of information by the thickness of the line. By 1400, $m_{ne...not}$ is hardly distinguishable from the prior and thus carries little information. Visually speaking, at this point its emphasis has almost entirely faded.

At this point it is useful to pause and summarize our results. We started off by distinguishing between the formal and functional Jespersen cycles, and focused on the functional cycle as a kind of inflationary process. At the start of the cycle, a new, initially emphatic form is introduced, but loses emphasis as it increases in frequency. We modeled these facts by defining a signaling game between speakers and hearers. Speakers have some information about the world, which they signal using different forms of negation, and hearers respond by paying varying amounts of attention and thus granting social status to the speaker. While hearers want a fair exchange of information for status, speakers are biased towards gaining more status for less information. Emphatic forms, which we defined as highly informative, are exploited by speakers due to this conflict of interests. As emphatic forms increase in frequency, however, they cease to be highly informative and hearers discount the amount attention paid to them. We fit a dynamic model of this interaction to corpus data from the functional cycle in Middle English.
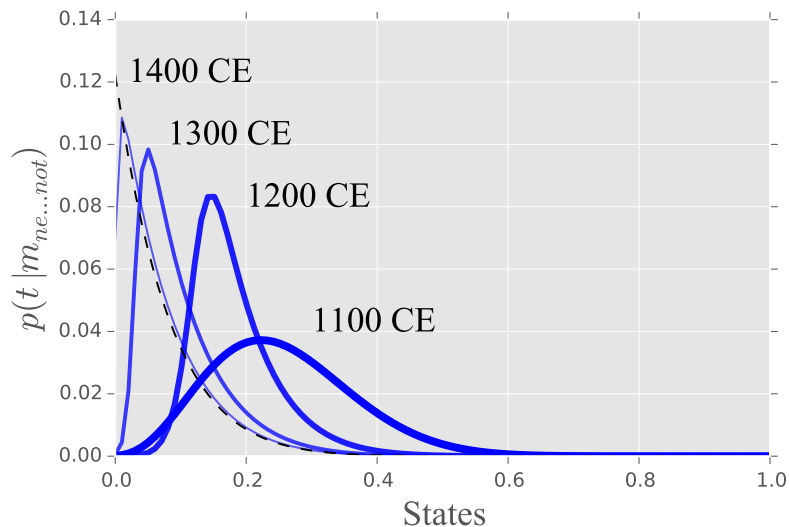
Figure 6   The meaning of the emphatic form over time as given by the conditional probability of states given $m_{ne...not}$.

6   Conclusion

Allowing for slight differences in the preferences of speakers and hearers can serve as an explanation for why populations abandon the use of particular forms in particular ways. That is, the fact that speakers prefer an uneven exchange of information for social status offers a kind of inverted Gricean explanation for the observed change in the functional Jespersen cycle. Our main contributions consist of the transformation of a general notion that the change is driven by inflationary pressures to a dynamic evolutionary game-theoretic model of the change that can be fit to corpus data.

There is, of course, more work to be done to further test the model. In particular, we might wonder how we can assess the reasonableness of $\hat{b}$. For example, Blume et al. (2001) had participants play a signaling game with abstract messages under varying degrees of diverging interest. That is, they had different cohorts play games with varied b and found that senders exploit and receivers discount messages over time. So, it would seem that if speakers have a bias, then speakers and hearers will behave in the expected manner. However, translating these results to an ecologically valid experimental paradigm for language change remains for future work. We still need to know when and how much speakers exhibit these kinds of biases.

References

Altmann, Gabriel, Haro von Buttlar, Walter Rott & Udo Strauss. 1983. A law of change in language. In Brian Brainard (ed.), *Historical linguistics*, 104–115.

Austin, John. 1962. *How to do things with words*. Clarendon Press.

Benz, Anton, Gerhard Jaeger & Robert van Rooij (eds.). 2006. *Game Theory and Pragmatics*. Palgrave Macmillan.

Blume, Andreas, Douglas V DeJong, Yong-Gwan Kim & Geoffrey B Sprinkle. 2001. Evolution of communication with partial common interest. *Games and Economic Behavior* 37(1). 79–120.

Börgers, Tilman & Rajiv Sarin. 1997. Learning through reinforcement and replicator dynamics. *Journal of Economic Theory* 77(1). 1–14.

Burnham, Kenneth P & David R Anderson. 2003. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media.

Bush, Robert & Frederick Mosteller. 1955. *Stochastic models for learning*. Wiley.

Byrd, Richard H, Peihuang Lu, Jorge Nocedal & Ciyou Zhu. 1995. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing* 16(5). 1190–1208.

Crawford, Vincent P & Joel Sobel. 1982. Strategic information transmission. *Econometrica* 1431–1451.

Croft, William. 1991. The evolution of negation. *Journal of Linguistics* 27(01). 1–27.

Dahl, Osten. 1979. Typology of Sentence Negation. *Linguistics* 17(1-2). 79–106.

Dahl, Osten. 2001. *Inflationary effects in language and elsewhere* 471–480. Philadelphia: John Benjamins Publishing Company.

Danescu-Niculescu-Mizil, Cristian & Lillian Lee. 2010. Don't 'have a clue'? Unsupervised co-learning of downward-entailing operators. In *Proceedings of acl*, 247–252.

De Jaegher, Kris & Robert van Rooij. 2013. Game-theoretic pragmatics under conflicting and common interests. *Erkenntnis* 1–52.

Deo, Ashwini. 2015. The semantic and pragmatic underpinnings of grammaticalization paths: The progressive to imperfective shift. *Semantics and Pragmatics* 8. 1–52.

Dessalles, Jean-Louis. 2014. Optimal investment in social signals. *Evolution* 68(6). 1640–1650.

Dessalles, JL. 2007. *Why we talk*. Oxford University Press, Oxford.

Detges, Ulrich & Richard Waltereit. 2002. Grammaticalization vs. reanalysis: A semantic-pragmatic account of functional change in grammar. *Zeitschrift f"ur Sprachwissenschaft* 21(2). 151–195.

Ecay, Aaron & Meredith Tamminga. 2015. Persistence as a diagnostic of grammatical status: The case of middle english negation. In Rob Truswell & Mathieu F (eds.), *From micro-change to macro-change*, 11. Oxford University Press.

Enke, Dankmar, Roland Muhlenbernd & Igor Yanovich. 2016. The emergence of the progressive to imperfective diachronic cycle in reinforcement-learning agents. In Sean Roberts, Christine Cuskley, Luke McCrohon, Lluis Barcelo-Coblijn, Olga Feher & Tessa Verhoef (eds.), *The evolution of language: Proceedings of the eleventh international conference on the evolution of language*, .

Farrell, Joseph. 1993. Meaning and Credibility in Cheap-Talk Games. *Games and Economic Behavior* 5. 514–531.

Fauconnier, Gilles. 1975. Pragmatic scales and logical structure. *Linguistic inquiry* 353–375.

Franke, Michael, Tikitu de Jager & Robert van Rooij. 2012. Relevance in Cooperation and Conflict. *Journal of Logic and Computation* 22(1). 23–54.

Gardiner, Alan H. 1905. The word. *Zeitschrift für Ägyptische Sprache* 41(JG). 130–135.

Gintis, Herbert, Eric Alden Smith & Samuel Bowles. 2001. Costly signaling and cooperation. *Journal of theoretical biology* 213(1). 103–119.

Givón, Talmy. 1978. Negation in language. In Peter Cole (ed.), *Syntax and Semantics, Volume 9: Pragmatics*, 69–112. Mouton de Gruyter.

Grice, H.P. 1975. Logic and conversation. In *Studies in the Way of Words*, 22–40. Harvard University Press.

Hamilton, William D. 1963. The evolution of altruistic behavior. *American naturalist* 354–356.

Hamilton, William D. 1964. The genetical evolution of social behaviour. *Journal of theoretical biology* 7(1). 1–16.

Hardin, G. 1968. The Tragedy of the Commons. *Science* 162(3859). 1292–1297.

Hofbauer, Josef & Simon M Huttegger. 2015. Selection-mutation dynamics of signaling games. *Games* 6(1). 2–31.

Hofbauer, Josef & Karl Sigmund. 1998. *Evolutionary games and population dynamics*. Cambridge University Press.

Hopper, Paul J & Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge University Press.

Horn, Laurence R. 1989. *A natural history of negation*. University of Chicago Press.

Israel, Michael. 2011. *The grammar of polarity: Pragmatics, sensitivity, and the logic of scales*. Cambridge University Press.

Jäger, Gerhard, Lars P Metzger & Frank Riedel. 2011. Voronoi languages: Equilibria in cheap-talk games with high-dimensional types and few signals. *Games and economic behavior* 73(2). 517–537.

Jespersen, Otto. 1917. *Negation in English and other Languages*. Host.

Jones, Taylor. 2015. Eem neagation in AAVE: A next step in Jespersen's Cycle, .

Kadmon, Nirit & Fred Landman. 1993. Any. *Linguistics and philosophy* 16(4). 353–422.

Keller, R. 1994. *On language change: The invisible hand in language*. Routledge.

Kiparsky, Paul & Cleo Condoravdi. 2006. Tracking Jespersen's Cycle. In Mark Janse (ed.), *Proceedings of the 2nd International Conference of Modern Greek Dialects and Linguistic Theory*, University of Patras.

Klima, Edward S. 1964. Negation in english. In Jerry Fodor & Jerrold Katz (eds.), *The structure of language*, 246–323. na.

Krifka, Manfred. 1995. The semantics and pragmatics of polarity items. *Linguistic analysis* 25(3-4). 209–257.

Kroch, Anthony & Ann Taylor (eds.). 2000. *Penn-Helsinki Parsed Corpus of Middle English Second Edition (PPCME2)*.

Kroch, Anthony S. 1989. Reflexes of grammar in patterns of language change. *Language variation and change* 1(03). 199–244.

Kullback, Solomon & Richard A Leibler. 1951. On information and sufficiency. *The Annals of Mathematical Statistics* 79–86.

Ladusaw, William A. 1993. Negation, indefinites, and the jespersen cycle. In *Annual meeting of the berkeley linguistics society*, vol. 19, 437–46.

Landman, Fred. 1991. *Structures for semantics*. Springer.

Lewis, David. 1969. *Convention*. Cambridge: Harvard University Press.

Lewis, David. 1979. Scorekeeping in a language game. *Journal of philosophical logic* 8(1). 339–359.

Maynard Smith, John. 1982. *Evolution and the Theory of Games*. Cambridge university press.

Maynard Smith, John & David Harper. 2004. *Animal Signals*. USA: Oxford University Press.

Meillet, Antoine. 1912. *L'évolution des formes grammaticales*. Nicola Zanichelli.

Miestamo, Matti. 2005. *Standard negation: The negation of declarative verbal main clauses in a typological perspective*, vol. 31. Walter de Gruyter.

Oechssler, Jörg & Frank Riedel. 2001. Evolutionary dynamics on infinite strategy spaces. *Economic Theory* 17(1). 141–162.

Oechssler, Jörg & Frank Riedel. 2002. On the dynamic foundation of evolutionary stability in continuous models. *Journal of Economic Theory* 107(2). 223–252.

Payne, John R. 1985. Negation. In Timothy Shopen (ed.), *Language typology and syntactic description*, 197–242. Cambridge University Press.

Posner, Rebecca. 1985. Post-verbal Negation in Non-Standard French, a Historical and Comparative View. *Romance Philology* 39(2). 170–197.

Schaden, Gerhard. 2012. Modelling the "aoristic drift of the present perfect" as inflation an essay in historical pragmatics. *International Review of Pragmatics*

4(2). 261–292.

Schelling, T.C. 1978. *Micromotives and Macrobehavior*. WW Norton.

Schwegler, Armin. 1988. Word-order changes in predicate negation strategies in romance languages. *Diachronica* 5(1-2). 21–58.

Scott-Phillips, T.C. 2008. On the correct application of animal signalling theory to human communication. In A. D. M. Smith, K. Smith & R. Ferrer i Cancho (eds.), *The evolution of language: Proceedings of the 7th International Conference on the Evolution of Language*, 275–282. Singapore: World Scientific Press.

Searcy, William A. & Stephen Nowicki. 2005. *The Evolution of Animal Communication: Reliability and Deception in Signaling Systems*. Princeton: Princeton University Press.

Selten, Reinhard. 1980. A note on evolutionarily stable strategies in asymmetric animal conflicts. *Journal of Theoretical Biology* 84(1). 93–101.

Simon, Herbert A. 1955. A behavioral model of rational choice. *The Quarterly Journal of Economics* 99–118.

Sutton, Richard S & Andrew G Barto. 1998. *Reinforcement learning: An introduction*. MIT press.

Taylor, Peter D. & Leo B. Jonker. 1978. Evolutionary stable strategies and game dynamics. *Mathematical biosciences* 40(1). 145–156.

Trivers, Robert L. 1971. The evolution of reciprocal altruism. *Quarterly review of biology* 35–57.

van der Auwera, Johan. 2009. The Jespersen Cycles. In Elly van Gelderen (ed.), *Cyclical Change*, John Benjamins Publishing Company.

Van Gelderen, Elly. 2008. Negative cycles. *Linguistic Typology* 12(2). 195–243.

Wallage, Phillip. 2008. Jespersen's cycle in Middle English: parametric variation and grammatical competition. *Lingua* 118(5). 643–674.

Yanovich, Igor. 2015. Analyzing imperfective games.

Yanovich, Igor. 2016. Genetic drift explains sapir's drift in semantic change. In Sean Roberts, Christine Cuskley, Luke McCrohon, Lluis Barcelo-Coblijn, Olga Feher & Tessa Verhoef (eds.), *The evolution of language: Proceedings of the eleventh international conference on the evolution of language*, .

Zahavi, A. 1975. Mate selection–a selection for a handicap. *Journal of theoretical Biology* 53(1). 205–214.

Christopher Ahern
University of Pennsylvania
Department of Linguistics
619 Williams Hall
Philadelphia, PA
cahern@ling.upenn.edu

Robin Clark
University of Pennsylvania
Department of Linguistics
619 Williams Hall
Philadelphia, PA
rclark@ling.upenn.edu