1-1-2015

Cycles and Stability in Linguistic Signaling

Christopher Andrew Ahern
University of Pennsylvania, christopher.ahern@gmail.com

Cycles and Stability in Linguistic Signaling

Abstract

This dissertation advances our understanding of the roles played by pragmatic and grammatical competence in theories of language change by using mathematical and statistical methods to analyze the cross-linguistic change in the expression of negation known as Jespersen's cycle. In the history of Middle English this change is characterized by two transitions: from pre-verbal ne to an initially emphatic embracing ne...not; from embracing ne...not to post-verbal not. This description conflates two often related process: the formal cycle describes changes in the forms of negation available and consists of the transitions from pre-verbal to embracing to post-verbal negation; the functional cycle describes changes in how forms are used to signal meaning and consists of the transition from pre-verbal to embracing negation.

Using tools from evolutionary game theory, we show that the functional cycle can be explained by limits on our pragmatic competence. The incoming embracing form is initially restricted to negating propositions that are common information between interlocutors. But, experimental evidence shows that speakers have difficulty in distinguishing common and privileged information. Speakers use the initially restricted form in more and more contexts that are less and less closely tied to the discourse, and it undergoes a kind of informational bleaching. Applying statistical methods developed in population genetics, we show that grammatical competence, and the process of acquisition through which it is formed, predict stability rather than change in both transitions of the formal cycle unless the observed transitions are the result of the accumulation of small random changes akin to genetic drift in finite populations. We show that we can reject this possibility in the first transition of the formal cycle, but not in the second. The possibility of random change in the second transition of the formal cycle offers some insight into the varying amount of time it takes across languages.

The main contribution of this dissertation is demonstrating the need for articulated models of both pragmatic and grammatical competence in explanatory theories of language change, while also offering a set of tools and methods for analyzing different factors in historical corpora.

Subject Categories
Linguistics

CYCLES AND STABILITY IN LINGUISTIC SIGNALING

Christopher Andrew Ahern

A DISSERTATION

in

Linguistics

Presented to the Faculties of the University of Pennsylvania in Partial

Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

_____

Robin Clark, Professor of Linguistics

Graduate Group Chairperson

_____

Rolf Noyer, Associate Professor of Linguistics

Dissertation committee
Mark Liberman, Professor of Linguistics
Florian Schwarz, Assistant Professor of Linguistics

CYCLES AND STABILITY IN LINGUISTIC SIGNALING

COPYRIGHT

2016

Christopher Andrew Ahern

Everything changes and nothing remains still...you cannot step into the same stream twice.

–Heraclitus, as quoted in Plato's *Cratylus*

Why do you go away? So that you can come back. So that you can see the place you came from with new eyes and extra colors...Coming back to where you started is not the same as never leaving.

–Terry Pratchett

Acknowledgements

ABSTRACT

CYCLES AND STABILITY IN LINGUISTIC SIGNALING

Christopher Andrew Ahern

Robin Clark

This dissertation advances our understanding of the roles played by pragmatic and grammatical competence in theories of language change by using mathematical and statistical methods to analyze the cross-linguistic change in the expression of negation known as Jespersen's cycle. In the history of Middle English this change is characterized by two transitions: from pre-verbal *ne* to an initially emphatic embracing *ne...not*; from embracing *ne...not* to post-verbal *not*. This description conflates two often related process: the formal cycle describes changes in the forms of negation available and consists of the transitions from pre-verbal to embracing to post-verbal negation; the functional cycle describes changes in how forms are used to signal meaning and consists of the transition from pre-verbal to embracing negation.

Using tools from evolutionary game theory, we show that the functional cycle can be explained by limits on our pragmatic competence. The incoming embracing form is initially restricted to negating propositions that are common information between interlocutors. But, experimental evidence shows that speakers have difficulty in distinguishing common and privileged information. Speakers use the initially restricted form in more and more contexts that are less and less closely tied to the discourse, and it undergoes a kind of informational bleaching. Applying statistical methods developed in population genetics, we show that grammatical competence, and the process of acquisition through which it is formed, predict stability rather than change in both transitions of the formal cycle unless the observed transitions are the result of the accumulation of small random changes akin to genetic drift in finite populations. We show that we can reject this possibility in the first

transition of the formal cycle, but not in the second. The possibility of random change in the second transition of the formal cycle offers some insight into the varying amount of time it takes across languages.

The main contribution of this dissertation is demonstrating the need for articulated models of both pragmatic and grammatical competence in explanatory theories of language change, while also offering a set of tools and methods for analyzing different factors in historical corpora.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

> I am, however, enough of a rationalist to want to find a basis that underlies these facts,
> undeniable though they may be; I would like to be able to think of the standard type of
> conversational practice not merely as something that all or most do in fact follow but as
> something that it is reasonable for us to follow, that we should not abandon.
> –Paul Grice (1975:29)

Intuitively, everyone can agree that languages change. But, this intuition depends on exactly what we mean when we say *language*. On the one hand the term can be used to refer to the various nuanced ways that different linguistic forms are used in communication, and on the other hand it can also be used to refer to the unique human faculty that allows for the acquisition of that combinatorially rich set of linguistic forms. Broadly speaking then, change can refer to either a difference in the grammatical knowledge that learners internalize, or a difference in how that grammatical knowledge is externalized and put towards communicative ends. So, our intuitive agreement about change may persist, but we might seem to differ in what we take to be changing or, perhaps more importantly, how we can study and ultimately understand the causes of such change.
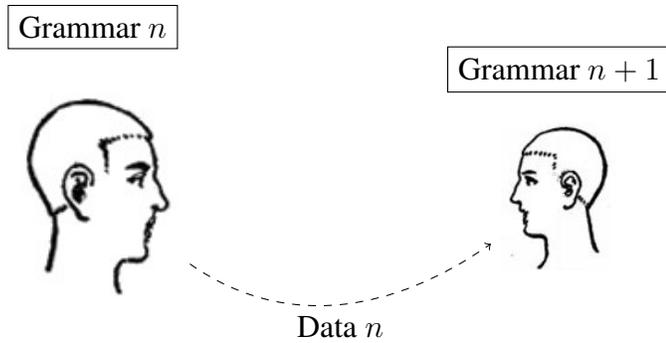
Figure 1.1: The process of language acquisition

Indeed, much of linguistics, the generative tradition in particular, has focussed almost exclusively on characterizing how the grammatical knowledge internalized through the process of acquisition might lead to change. In the terms of Chomsky (1986), language change proceeds as the process of language acquisition maps the externalized *E-Language* of one generation to the internalized *I-Language* of the next. We can visualize this schematically as in Figure 1.1 where the output from one generation serves as the input for acquisition in the next generation. The definition of change under this conception is expressed in terms of differences between subsequent grammars. In fact, as Chomsky and Halle (1968:249) put it, an observed change can only come about through a change in the underlying grammars in subsequent generations.

Yet, while this may be true, it does not necessarily reveal the underlying cause of the change. Crucially, the process of acquisition does not act in a vacuum. The output of one generation serves as the input to the next. And, while this input to acquisition arises from the interplay of many factors, it is not arbitrary. Rather, it is governed by a *pragmatic competence* that "underlies the ability to use [*grammatical competence*] along with the conceptual system to achieve certain ends or purposes" (Chomsky 1980:59). Or, in Gricean terms, the output from the previous generation arises from the rational use of an internalized grammar. We can visualize this schematically as in Figure 1.2 where the output of one

2
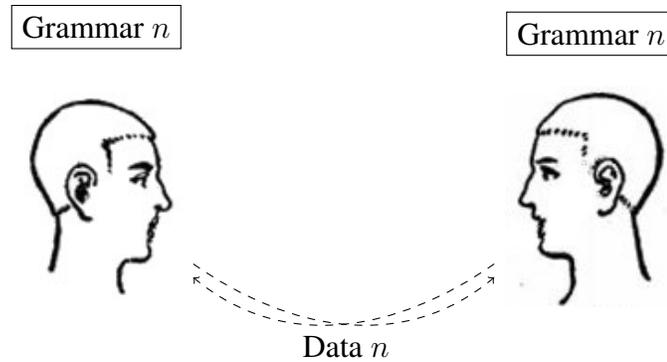
Figure 1.2: The process of language use

generation arises through the strategic use of the forms made available by an internalized grammar. Where grammatical competence is formed by a mapping from one generation's *E-Language* to *I-Language* of the next, pragmatic competence is what governs the mapping from each generation's *I-Language* to its *E-Language*.

Taken together, we can summarize the interaction of these two processes as in Figure 1.3, where both use and acquisition are entwined in the process of change. So, while we can certainly define change in terms of the internalized grammars of speakers at different points in time, the process of acquisition is not the only locus of change. That is, both the process of externalization, internalization, and the interaction between the two can lead to change in the grammars acquired over time. The central goal of this dissertation is to provide the mathematical tools for defining and analyzing models of change stemming from both use and acquisition. In doing so, we aim to demonstrate that the notion of pragmatic competence can be integrated into broader theories of change, and is not only incredibly useful but sometimes necessary to explain linguistic change.

At the center of this endeavor will be a diachronic process that implicates both use and acquisition, the development in the expression of sentential negation over time known as *Jespersen's cycle* (1917). The process is often described as the result of two transitions. The first transition occurs when a preverbal form of negation is replaced by an embracing form,

3
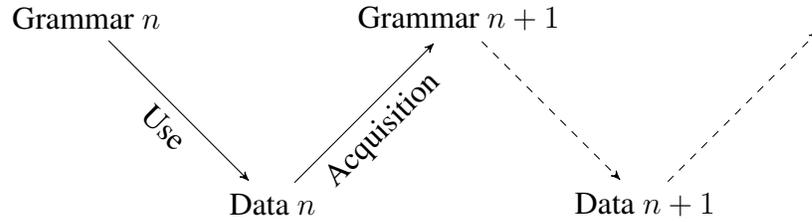
Figure 1.3: The iterated process of language change through acquisition and use

which is initially characterized as being more emphatic. The second transition occurs when this embracing form is subsequently replaced by a purely postverbal form. In the history of English, we observe both of these two transition in Middle English from *ne* to *ne...not* and from *ne...not* to *not*. Our goal will be to determine the role of use and acquisition in each of these transitions.

In Chapter 2 we begin by distinguishing between two phenomena that have often been conflated in investigations of Jespersen's cycle. In particular, we argue that Jespersen's cycle as it is often described consists of both a *formal* and a *functional cycle*. The formal cycle describes the change in the formal complexity of negation over time. It takes place as negation becomes more and then less formally complex, as can be seen in the transitions in the history of English from *ne* to *ne...not* to *not*. The functional cycle describes the way that different forms of negation are used to signal meaning. It takes place as one form of plain negation is replaced by another form. This can be seen in the history of English from *ne* to *ne...not* where the originally empathic *ne...not* displaces *ne* as it increases in frequency, loses its emphasis, and comes to signal plain negation. We note the logical and empirical relationship between the two cycles: the functional cycle can occur independently of the formal cycle. This result informs the structure of the rest of the dissertation; we start by addressing the functional cycle before turning to the formal cycle.

The first part of this dissertation addresses the functional cycle. In Chapter 3 we introduce the mathematical tools we will use to model the functional cycle. In particular, we

show how we can use evolutionary game theory to describe how meaning is signaled in a population over time. Importantly, these tools allow us to model a qualified kind of Gricean rationality. That is, individuals are *boundedly rational* insofar as they have limited cognitive and informational resources (Simon 1955, 1957). Yet, these tools allow us to show how the actions of individuals can give rise to change at the population level, even when those small decisions are not the product of conscious deliberation (Keller 1994). This is particularly important when we turn to the functional cycle in Chapter 4, where we show that the first transition from *ne* to *ne...not* can be explained as the result of speakers' limitations in keeping track of common versus private knowledge. So, just as Gricean rationality has been used to explain particular patterns of synchronic use, a kind of bounded rationality allows us to explain the functional cycle. So, how we use these two forms explains why they change over time, and the transition from *ne* to *ne...not*.

However, the same model does not apply to the transition from *ne...not* to *not*, so we turn to the formal cycle in the second part of this dissertation. In Chapter 5 we describe a model of syntactic acquisition and determine its predictions for both of the transitions of the formal cycle. In particular, we show that acquisition cannot explain either of the two transitions from *ne* to *ne...not* or from *ne...not* to *not*, other than as the result of a random change in the grammars acquired. In Chapter 6 we test this possibility using statistical methods developed in population genetics to test for random drift versus selection. We find that we can reject random drift in the case of the first transition from *ne* to *ne...not*, but we cannot reject drift in the case of the second drift from *ne...not* to *not*. This first result suggests that use is the driving force behind the first transition as part of the functional cycle. The second result shows that acquisition does not play a role in any of the observed transitions. So, insofar as we can offer an explanation of either of the observed changes, we need the notion of pragmatic competence to do so.

Chapter 2

Jespersen's Cycles

> 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. (Jespersen 1917:4)

Originally coined by Dahl (1979), the term *Jespersen's cycle* is often used in reference to the observation cited above. It is certainly the most quoted aspect of Jespersen's (1917) seminal work, which prefigures many of the current empirical and theoretical issues in the study of negation (cf. Horn 1989). Yet, this passage is also interpreted in two very distinct ways. This fundamental ambiguity stems from the fact that Jespersen noted both *formal* and *functional* patterns in the expression of sentential negation over time.[1] Both patterns can be conceived of as cycles in their own right. That is, we can find a series of transitions

---

[1] Sentential negation refers to the semantic property of negating an entire proposition, not just some sub-part. 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.

from and back to states that are in some sense formally or functionally equivalent. But, the term Jespersen's cycle is often used in one of these two senses or the other, without clear distinction.

Perhaps even worse, the canonical presentation of what is referred to as Jespersen's cycle conflates these two uses (Posner 1985, Schwegler 1988, 1990, Ladusaw 1993, Schwenter 2005, 2006). 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

This chapter is devoted to defining and distinguishing 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 meaning, respectively. Our goal is to clarify the relationship between the two cycles and what would count as an explanation of each.

First, we outline the formal aspects of how negation is expressed at the stages of the formal cycle. We provide a historical description of the structural forms that express negation at different points in the history of English, and compare this trajectory with that of French to emphasize particular aspects of the formal cycle. Second, we outline the function of those forms at different stages of the functional cycle. We provide a historical description of the functions of the different forms at points in the history of both English and French, noting the relationship between optionality and these different functions. We also note the logical relationship between the two cycles: the formal cycle entails the functional cycle, but not vice versa. Third, in light of our definitions and the logical relationship between

7

the two cycles, we discuss two ways of conceptualizing the causes of the cycles and the potential role of syntactic acquisition and pragmatic use in both.

While the main motivation of this chapter is terminological, distinguishing between the two kinds of cycles has an important consequence. Given the logical relationship between the two cycles, an explanation of the formal cycle requires an explanation of the functional cycle. This informs the structure of the dissertation. Since an explanation of the functional cycle must be our first priority, we begin in Part I by pursuing such an explanation. In Chapter 3 we present a mathematical framework for understanding how different functional meanings are signaled in populations over time. In Chapter 4 we apply this framework to model the functional cycle in the history of English. With this in place, in Part II we turn to the formal cycle. In Chapter 5 we evaluate a model of syntactic acquisition and the role it might play in an explanation of the formal cycle. In Chapter 6 we consider the possibility that stochastic processes may lead to the transitions in the formal cycle. So, our goal in this chapter is to set the foundation for the rest of the dissertation by clarifying the distinction between the two kinds of cycles.

2.1   The formal cycle

The formal cycle is defined in terms of the forms that are used to express negation over time, and consists of two transitions. At the start of the formal cycle, a single element is used to express negation. The first transition occurs when this single element is supplemented by additional lexical material. The second transition occurs when the original negative element is subsequently lost, leaving the added lexical material as the only expression of negation. In the most general sense, the formal cycle occurs when the form of negation becomes more and then less complex. It is cyclic in the sense that the forms of negation at the start and end of the cycle are of equal formal complexity.
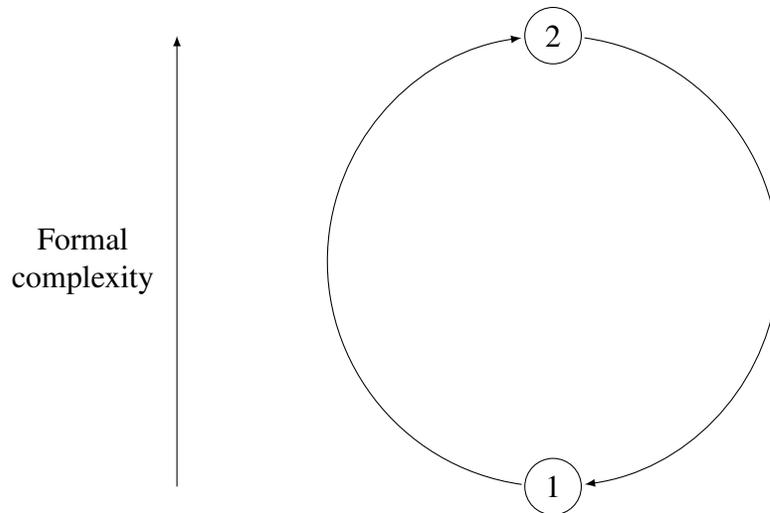
Figure 2.1: The formal Jespersen cycle

This can be shown schematically as in Figure 2.1, where the vertical axis represents the degree of formal complexity. The first transition takes place from ① to ② with the addition of material. The second transition takes place from ② back to ① with the loss of material. The addition of material leads to an increase in formal complexity of negation and the loss of material leads to a decrease. The formal cycle is then just a cycle from and back to a less complex form of negation: negation is expressed by some stuff, then more stuff, then less stuff.

For example, Jespersen noted that in the history of several European languages, including English and French, a pre-verbal negative element is supplemented by a post-verbal element, and the pre-verbal element is subsequently lost. This can be shown schematically as in Figure 2.2 where the different forms of negation are arranged according to formal complexity on the vertical axis and time along the horizontal axis. At both the start and the end of the formal cycle a single element expresses negation. Intuitively, if we abstract away from the structural realization of the forms as pre- or post-verbal, then Figure 2.2 maps onto Figure 2.1. The curious fluctuation in form that Jespersen noted simply becomes a closed orbit through the space of formal complexity.
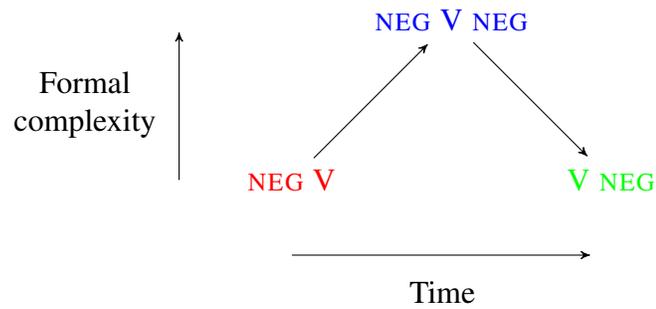
Figure 2.2: The realization of the formal cycle in English and French

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 stage of embracing or bipartite negation where a post-verbal 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)

There are several sources that these additional elements are drawn from, most notably from *negative polarity items*, overwhelmingly the set of *minimizers* (e.g. "not a drop", "not a hair") and *generalizers* (e.g. "not ever", "not at all", Horn 1989). For example, in the case of English, *not* comes from Old English *nawiht* (*literally* "no thing, creature, being"). Other sources include, indefinite pronouns (e.g. "no one"), negative adverbs (e.g. "never", cf. Horn 1989, Van Gelderen 2008), negative verbs (e.g. "refuse", "deny", "reject", "avoid",

"fail", and "lack", Givón 1978), and the concatenation of negative and existential verbs (Croft 1991).

In light of this broad range of sources, it is important to note that the first transition need not consist of 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 express negation in its own right (Jones 2015).

(3)   You don't eem know.

(4)   You eem know.

As Jones (2015) notes, this form comes from but is arguably distinct from *even*.

(5)   #eem numbers

Importantly, it is the formal status rather than the structural position that is relevant. Given the range of sources for additional negative elements, this is not surprising. But, it bears emphasis that the transition from pre- to post-verbal negation is not the only path through the formal cycle. The transition could just as well have been from post- to pre-verbal negation or from and back to pre-verbal negation. It is a contingent historical fact, arising from the syntax of English and the source of the additional material, rather than some necessary property of the formal cycle.

The final stage in the formal cycle is a return to a single negative element. This is seen in Early Modern English, where the preverbal element in *ne...not* is lost, leaving the post-verbal *not* as the sole negative element.

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

11

The emergence of *do*-support in Early Modern English yields a state parallel to Old English with a sole pre-verbal negator:

(7)   I do not   say
        I do NEG say
        (Present-day English)

The contraction of negation offers an even closer parallel to Old English where *ne* contracted with several verbs (e.g. *ne be → nis*).

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

But, however suggestive these further developments are, they are not necessary components of the formal cycle. Jespersen (1917:10) himself noted the uniqueness of these developments, which he attributed to a tendency to place negation at the beginning of the sentence to avoid confusion. As he put it, the effect of post-verbal negation in German results in a kind of semantic garden path.[2]

> [T]he hearer or reader is sometimes bewildered at first and thinks that the sentence is to be understood in a positive sense, till suddenly he comes upon the *nicht*, which changes everything; see, for instance "Das leben ist der güter höchstes nicht."

Yet, despite this purported tendency, the majority of languages that Jespersen noted, including his native Danish, persist in a perplexing state of post-verbal negation. This can only be taken as further evidence that we ought to interpret the implications of Jespersen's observation in formal rather than structural terms. It matters how much stuff is used to express negation, not where that stuff is.

---

[2]Living is not the highest good (lit. The living is the good highest not)
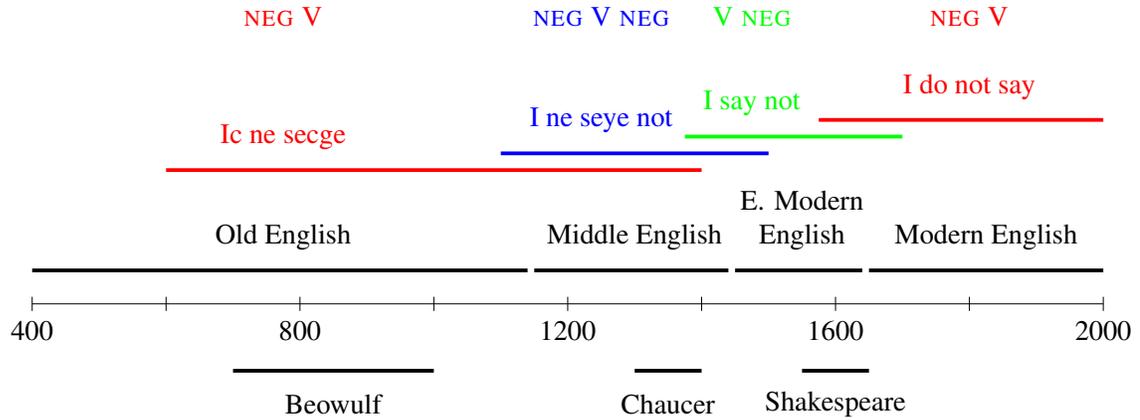
Figure 2.3: Timeline of negation in the history of English

Given the stages of the formal cycle in the history of English, it is useful to give them some historical context. The stages are summarized in Figure 2.3. The horizontal axis represents years in the common era. Immediately above the horizontal axis are commonly-used terms for historical periods: Old English (*ca.* 400-1175 CE), Middle English (*ca.*1175-1450 CE), Early Modern English (*ca.* 1450-1650 CE), and Modern English (*ca.* 1650-Present CE). Above these general time periods are the spans of the different forms of negation. From Old to Early Modern English we make a full formal cycle from and back to a single negative element. We also see the transition at the end of the formal cycle from post- back to pre-verbal negation in Modern English.

Below the axis are rough historical anchors that offer a general sense of the historical periods: Beouwulf, the Old English epic poem written some time between 700-1050 CE; Chaucer, the Late Middle English author who is widely-credited as the first to write in the vernacular language of his day, from around 1350-1400 CE; and Shakespeare, the noted dramatist and playwright, from around 1550-1650 CE.

The history of negation in French offers a parallel to the formal cycle in English, although the exact details of the timeline are not without dispute (cf. Martineau and Mougeon 2003). The Old French pre-verbal *ne* becomes Middle French *ne...pas*, and subsequently

Modern Colloquial French post-verbal *pas*.

(9)   Jeo ne   dis
        I   NEG say
        (Old French)

(10)   Je ne   dis pas
        I NEG say NEG
        (Middle French)

(11)   Je dis pas
        I   say NEG
        (Modern Colloquial French)

Again, it is useful to provide some historical context. The stages of the formal cycle in the history of French are summarized in Figure 2.4. Immediately above the horizontal axis are commonly-used terms for historical periods: Old French (*ca.* 900-1350 CE), Middle French (*ca.*1350-1600 CE), Classical French (*ca.* 1600-1700 CE) , and Modern French (*ca.* 1700-Present CE). Above these general time periods are the spans of the different forms of negation. From Old to Modern French we make a full formal cycle from and back to a single negative element.

Below the axis are rough historical anchors that offer a general sense of the historical periods: Charlemagne, the first Holy Roman Emperor from 750-800 CE; the Hundred Years' War from roughly 1350-1450 CE; Voltaire, the Enlightenment author and satirist from roughly 1700-1800 CE.

Hansen and Visconti (2009, 2012) note that negation in Louisiana Creole French negation is purely pre-verbal. Of course, this observation comes with the caveat that no conclusions about the future of French can be drawn from this change.

(12)   Mo pa   di
        I   NEG say
        (Louisiana Creole French)

14
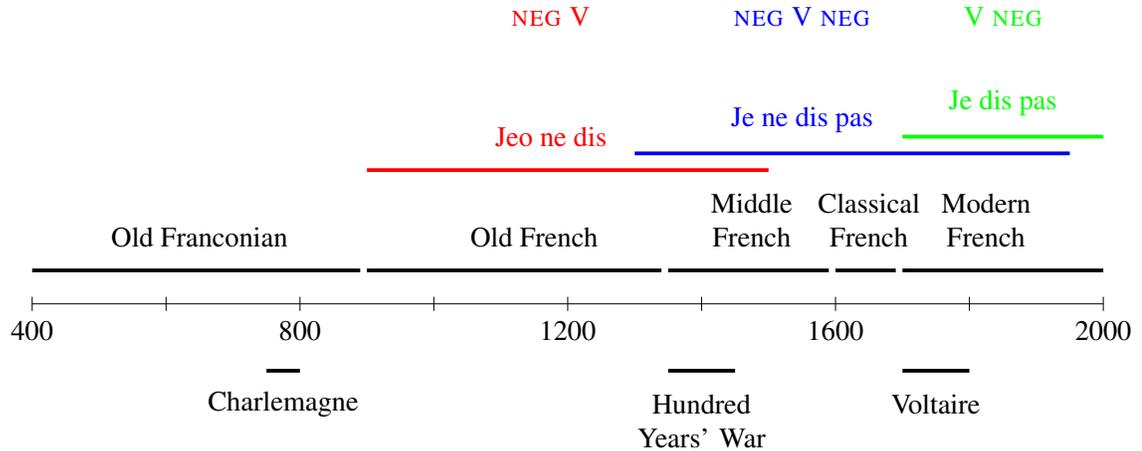
Figure 2.4: Timeline of negation in the history of French

But again, even if such developments were readily apparent, they would not constitute a necessary component of the formal cycle.

Thus, the crucial components of the formal cycle are a transition from one negative element to two, and eventually back to one. For the specific case of English, as well as French, the formal cycle is realized as a transition from pre-verbal to embracing to post-verbal negation. We can represent these transitions schematically as follows.

1. NEG V

2. NEG V NEG

3. V NEG

Note that these stages are not mutually exclusive. That is, there are transitions between these stages (van der Auwera 2009). This captures the theoretical intuition that diachronic change is rarely a dramatic shift. Where parentheses are taken to indicate optionality, the stages can be represented as the following.

1. NEG V

2. NEG V (NEG)

3. NEG V NEG

15

4. (NEG) V NEG

5. V NEG

This more articulated model also comes closer to the diachronic facts. For example, in Late Middle English, all the forms of negation are used contemporaneously: purely pre-verbal, embracing, and post-verbal negation co-occur for a brief period in time. For the moment then, we can take this as the abstract trajectory of the formal cycle over time.

Importantly, an explanation of the formal cycle in English must consist of two components: it must provide the conditions for the transition from pre-verbal to embracing negation as well as the conditions for the transition from embracing to post-verbal negation. It should be noted that criterion for such explanations can be both qualitative and quantitative. That is, an explanation of the formal cycle can predict both *that* a transition will occur as well as *how* it will occur. In what follows, we will use both kinds of criteria in evaluating models as explanations of change. With this definition of the formal cycle in place, along with what would constitute an explanation for its instantiation in the history of English, we turn to the functional cycle.

2.2 The functional cycle

If the formal cycle is defined by the forms of negation, then the functional cycle is defined by the function that those forms are put to. It is about how and how many forms are used to mean different things. At the first stage of the functional cycle a single form is generally used to express negation, often characterized as plain negation. The functional cycle begins with the introduction of another form that is used to express negation in a semantically stronger way, often characterized as being more emphatic. The functional cycle progresses as this new stronger form increases in frequency, weakens, and replaces the original negative form. In the most general sense, the functional cycle occurs when the
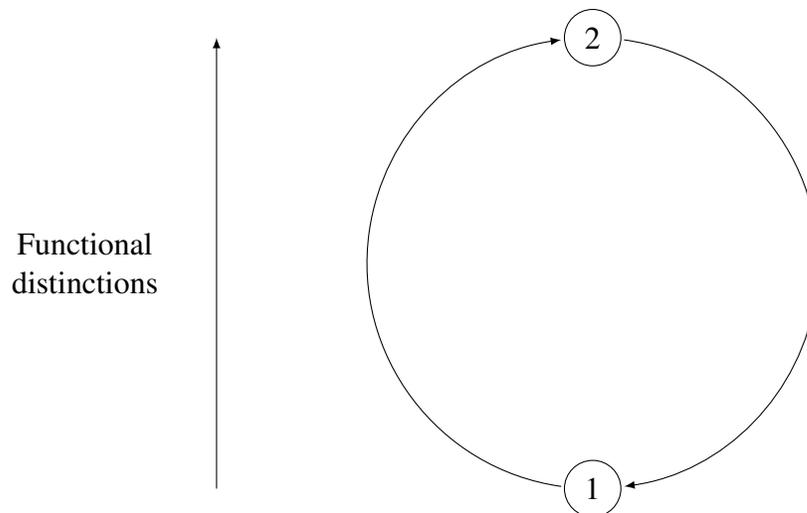
Figure 2.5: The functional Jespersen cycle

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.

This can be shown schematically as in Figure 2.5 where the vertical axis represents the number of functionally distinct forms of negation. The addition of a new form increases the number of distinctions from ① to ②, and the loss of the original form decreases the number of distinctions from ② back to ①. However, we should note that Figure 2.5 does not convey all of the necessary information about the functional cycle.

Namely, the functional cycle does not hold for just any pair of semantic distinctions. Rather, the incoming form is semantically stronger than the incumbent form. We can more accurately represent the details of the functional cycle as in Figure 2.6, where the vertical axis represents semantic strength and the horizontal axis represents time. The original form is supplemented with an additional form that is semantically stronger and this new form weakens over time as it replaces the original from.

The stronger negative form is often a result of elements being added to the original plain form. For example, in the history of English and French a pre-verbal element is supplemented by a post-verbal element, which strengthens negation. The meaning of the
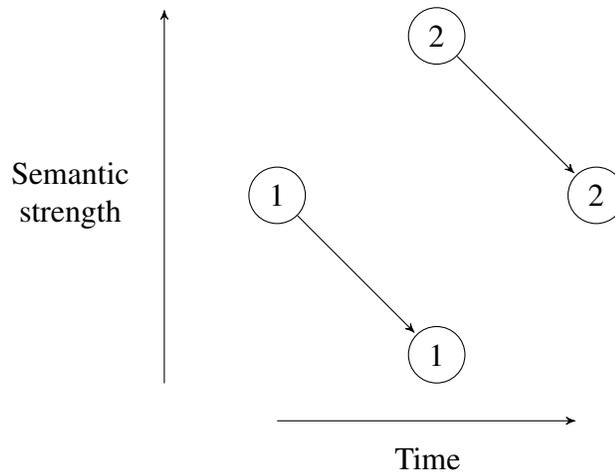
Figure 2.6: The functional Jespersen cycle in more detail

combined pre- and post-verbal elements weakens over time, and the two elements come to have the same force as the original pre-verbal element in isolation. This can be shown schematically as in Figure 2.7 where the different forms are arranged according to semantic strength along the vertical axis and time along the horizontal axis. If we abstract away from the realization of the particular forms then Figure 2.7 maps onto Figure 2.6, which in turn maps onto Figure 2.5. The curious fluctuation in function that Jespersen noted becomes a particular kind of closed orbit through the space of functional distinctions, which stems from the weakening of negative forms along a semantic dimension.

We see the first stage of the functional cycle in the history of English and French, with pre-verbal *ne* expressing plain negation.

(13)   Ic ne   secge
       I NEG say

(14)   Jeo ne   dis
       I   NEG say

The optional addition of a post-verbal element in the embracing forms *ne...not* and *ne...pas* is used to express a stronger negation.
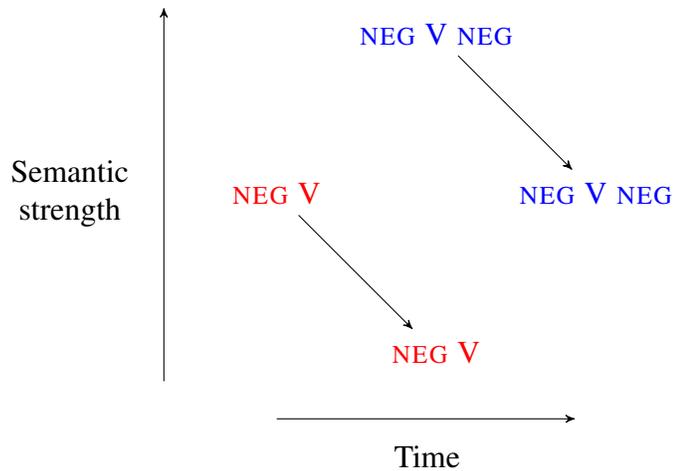
Figure 2.7: The realization of the functional cycle in English and French

(15)   I ne   seye not
        I NEG say   NEG

(16)   Je ne   dis pas
        I   NEG say NEG

The initial effect of the embracing form, in Jespersen's words (1917:15):

> ...[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.

Despite the evocative phrasing, Jespersen was certainly not the first to notice the trajectory
of the functional cycle. It was noted in great detail by both Meillet (1912:393) and Gardiner
(1905:134) in French and several other languages.

> [French *pas* and *point*], from the Latin *passum* and *punctum*, were orig-
> inally adverbial accusatives placed at the end of negative sentences for the
> purpose of emphasis; just like the English "not a jot", "not a straw"....*Pas* and
> *point*, and like them the Demotic B, Coptic AN, next lose their emphasizing

19

force, and become mere adjuncts of the negative words (French *ne*, Coptic 'N).

Last of all, they come themselves to be looked upon as negative words.

Indeed, van der Auwera (2009) suggests that *Meillet's spiral* (1912:394) may be the more appropriate term for the functional cycle.[3]

> Thus, languages follow a sort of spiral development: they add extra words to intensify expression; these words fade; decay and fall to the level of simple grammatical tools; one adds new or different words on account of expressiveness; the fading begins again, and so on endlessly.

If the functional cycle in English and French begins with the introduction of the optional post-verbal element to create an embracing form that intensifies expression, then it ends when the post-verbal element ceases to be optional. That is, when the post-vebal element becomes obligatory only the embracing form is used, and its intensity fades. This is because the embracing form ceases to be able to signal anything about the distinction between plain and emphatic negation. As Kiparsky and Condoravdi (2006) rightly put it, to emphasize everything is to emphasize nothing. This inverse relation between frequency and informativeness has been argued to underly multiple linguistic phenomena. As forms increase in frequency, they undergo a kind of *rhetorical devaluation* (Dahl 2001). Simply put, 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.

Now that we have defined the formal and functional cycles, there are several important points to be made regarding the relationship between them. First, the effect of one cycle often has implications for the other. For example, the addition of formal material almost always comes with a more restricted and hence stronger meaning.[4] This is a natural con-

---

[3]Translation McMahon (1994:165)

[4]The rare exception being truly empty obligatory pleonastic or expletive elements: "It's raining." Even periphrastic *do*, which is redundant outside of emphatic affirmatives: "I DO want pizza", originally carried some information on its way to becoming obligatory (Ecay 2015)

sequence of how semantic composition proceeds in a generally intersective fashion, to put it set-theoretically. For example, "a black bear" is certainly more specific, and hence semantically stronger than "a bear". In the same way, we would expect *ne...not* to be more specific in comparison to *ne*. This means that the first transition in the formal cycle is virtually guaranteed to coincide with the entirety of a functional cycle, which is indeed what we see in both English and French.

Second, the first transition of the formal cycle has a functional cycle tucked inside of it. But, the second transition of the formal cycle does not correspond to another functional cycle. That is, in the case of English and French the transition from embracing to post-verbal negation does not correspond to the same functional trajectory as the preceding transition from pre-verbal to embracing negation. This follows intuitively, again, from the compositional nature of meaning. The lexical content of the post-verbal form *not* is a proper subset of the embracing form *ne...not*, and thus we would not expect the post-verbal form to have a stronger or more restricted meaning than the embracing form.

Third, 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. In Meillet's estimation, the functional cycle is achieved when "one adds new *or* different words". Kiparsky and Condoravdi (2006) show that this is exactly what takes place in the history of Greek. Historical forms of negation in Greek are listed in Table 2.1, where emphatic negation is taken to be the semantically stronger form in comparison to *plain* negation at any point in time.[5] The sources of the different forms are ordered chronologically. Importantly, there is a consistent transition

---

[5]We omit some of the forms for a concise presentation, but see (Kiparsky and Condoravdi 2006:1) for a full list. Also, we should note that it is a bit of a misnomer to call any particular form *the* emphatic form. There are always means for augmenting plain negation to make it emphatic (e.g. 'at all', 'ever', cf. Horn (1989:452) and Israel (2011:258)). We might think of the forms in Table 2.1 as *plain* and *frequently-used-but-stronger-than-plain* negation. We return to a particular interpretation of what is meant by *emphasis* in Chapter 4.

of forms between the two functions: the emphatic negation of the last millennium becomes the plain negation of this millennium.

PLAIN	EMPHATIC	SOURCE
ου...τι	ου-δε...εν	Ancient Greek
(ου)δεν...τι	δεν...τιποτε	Early Medieval Greek
δεν...τιποτε	δεν... πραμα	Greek Dialects
δεν...πραμα	δεν...απαντοξη	Modern Cretan

Table 2.1: Historical forms of plain and emphatic negation in Greek

Crucially, at least some of these functional cycles occur without any concomitant formal cycle. For example, if we were to compare the formal complexity of forms after Early Medieval Greek, they would all be equivalent. All of them consist of a shared pre-verbal element δεν along with a single post-verbal element. Thus, we see several embracing forms come to express plain negation over time.

Taken together, these points indicate a particular logical relationship between the two cycles. Namely, the formal cycle entails the functional cycle, but not vice versa. This relationship is important because it sets a clear limit on how much an explanation for one kind of cycle can extend to the other. That is, the conditions for the formal cycle can be, at most, sufficient for the functional cycle. In the other direction, the conditions for the functional cycle can be, at most, necessary conditions for the formal cycle. This means that a full understanding of both cycles crucially rests on understanding the functional cycle.

For now we will largely be concerned with the realization of the functional cycle in English. The crucial component of the functional cycle is the transition from and back to a single plain form of negation. For the case of English and French, the functional cycle is realized schematically as the transition from pre-verbal to embracing negation.

1. NEG V

2. NEG V (NEG)

3. NEG V NEG

As we noted above, this leaves out the important detail of semantic strength, but also shows the frequent relation between the functional and formal cycles. An explanation of the functional cycle in English must consist of one component: it must provide the conditions for the transition from pre-verbal to embracing negation. Note that an explanation of the formal cycle requires an explanation of the functional cycle, but not vice versa. With the definition of both cycles, along with the relationship between them and their explanations, we now turn to two ways of conceptualizing their causes.

2.3   Causes of the cycles

We have distinguished between the two kinds of cycles that Jespersen noted. The formal cycle is constituted by a transition from and back to equally complex forms of negation. The functional cycle is constituted by a particular transition from and back to a single form being used to express plain negation. We now consider the two major kinds of scenarios that have been used to conceptualize the causes of the cycles. Drawing on the terminology of sound change, the two approaches can be though of as *pull-chains* and *push-chains* involving the different forms of negation.

The pull-chain scenario finds its most natural application in the case of the formal cycle, where a new form is *pulled* into expressing negation due to formal weakening. The old form *pulls* in the new form. The push-chain scenario finds its most natural application in the case of the functional cycle, where the old form is pushed out of expressing negation due to functional weakening. The new form *pushes* out the old form. We assess the plausibility of both scenarios before turning to how different process such as syntactic acquisition and pragmatic use might cause the dynamics of each transition.

2.3.1 A pull-chain scenario

While Jespersen did not distinguish between the cycles as we have, he did conceive of change as the product of both formal and functional weakening and strengthening. In particular, he took the role of formal phonetic weakening of the pre-verbal element as a potential cause of change. The clearest interpretation of this cause is in terms of the first transition of the formal cycle. But, it also has potential explanatory power with regards to the second transition.

Regarding the first transition of the formal cycle, Jespersen noted that the pre-verbal elements in English and French were prone to not receiving stress. This creates a problem, insofar as this lack of stress arguably also made negation hard to perceive (Jespersen 1917:5):

> The incongruity between the notional importance and the formal insignificance of the negative (often, perhaps, even the fear of the hearer failing to perceive it) may then cause the speaker to add something to make the sense perfectly clear to the hearer.

That is, given the importance of the distinction between affirmation and negation, he reasoned that some additional word is used "to increase the phonetic bulk" of the negative signal to bolster its perception (Jespersen 1917:14). That is, the new embracing form is *pulled* into expressing negation because of the weakness of the purely pre-verbal form.

There are several reasons to be skeptical of this kind of pull-chain scenario. First, phonetic weakening is quite common (Bybee 2003). This prevalence suggests that we should be cautious in attributing to it a role in any particular morphosyntactic change. On balance, phonetic weakening is neither necessary nor sufficient for morphosyntactic change. Second, Labov (1994:547-599) offers a thorough critical evaluation of the functional preservation of meaning in the face of sound change. By and large, sound change proceeds in a

24

mechanical fashion without the conscious adjustment to avoid communicative pitfalls. This is true even when such change leads to the loss of the distinction between negation and affirmation. Labov (2010:320) notes that in his own native north New Jersey dialect, the pronunciation of the affirmative *can* and the negative *can't* are at times indistinguishable:

> A very common utterance among residents of this Northern New Jersey area was "Did you say C–A–N or C–A–N–T?," since the vowel is tense in both words and the /t/ is often neutralized before a following apical obstruent (as in "I can't tell you").

Despite the importance of the functional distinction, no additional material has been added to differentiate the two senses. Third, (Posner 1985:177) argues from Italian dialect data that the strength of the pre-verbal form is not correlated with whether the first stage of the formal cycle takes place or not. Taken together, this suggests that a pull-chain scenario is unlikely. Or, at the very least, formal weakening cannot be considered as a necessary or sufficient cause of the first transition of the formal cycle.

Regarding the second transition of the formal cycle, it is useful to note that no such transition takes place in the history of Greek. That is, in Table 2.1 we see that from Early Medieval Greek onwards there is no formal weakening of either the pre- or post-verbal elements. The crucial difference between the pre-verbal form in those cases and English and French is that the first constitutes a full closed syllable, whereas the second two do not. Thus, while a pull-chain scenario may not offer a causal explanation of the first transition, the role of formal weakness may be important at different stages of the formal cycle. That is, the loss of the pre-verbal element in the transition from the embracing form to the post-verbal form may indeed be related to its formal weight.

2.3.2 A push-chain scenario

Unsurprisingly, Jespersen prefigured the other major conception of the cycle insofar as he took weakening and strengthening to be both formal and functional. Following Meillet, more recent approaches have focused on the role of functional strengthening and weakening (Detges and Waltereit 2002, Hopper and Traugott 2003, Eckardt 2006, Kiparsky and Condoravdi 2006, *inter alia*). The clearest interpretation of this cause is in terms of the functional cycle as a kind of *push-chain*.

Crucially, these accounts assume that a stronger form is introduced, increases in frequency due to overuse, and is thus weakened (Dahl 2001). That is, once the new more emphatic form is introduced, its frequency increases due to pragmatic pressures. The subsequent weakening of the new form follows from the information-theoretic properties of signals: to emphasize everything is to emphasize nothing. As the incoming form becomes obligatory, it takes over the expression of plain negation in its own right, *pushing* the original form out. In the case of English and French, the obligatorification of the post-verbal element pushes the purely pre-verbal form out of the picture. It should be noted that this push-chain is all that is required to account for the functional cycle. That is, so long as one form of plain negation is replaced by a formerly emphatic form, the functional cycle has occurred. The push-chain ends with the end of the functional cycle.

This point is important insofar as some analyses have emphasized the completion of the functional cycle as setting the stage for the second transition of the formal cycle. However, these analyses treat the push chain as one between negative elements rather than negative forms. For example, Detges and Waltereit (2002:187) argue that the loss of the pre-verbal element follows from a kind of *constructional iconicity* where the simple meaning of plain negation is expressed by a simple form. That is, the sentence just is not big enough for two negative elements. Similarly, Frisch (1997) argues that the loss of the pre-verbal element results from the unstable functional doublet created by the use of both *ne* and *not* in *ne...not*.

26

From this viewpoint, Burridge (1993:201) flips the reasoning of the pull-chain, noting that the loss of the pre-verbal element can be seen as the effect, rather than the cause of the addition of the post-verbal element. In a certain sense, this is a kind of functionally-mediated push-chain for the formal cycle. That is, the introduction of the post-verbal element pushes the pre-verbal element out, due to a functional constraint on the number of negative elements in a sentence.

While the notion of simplicity in form to match simplicity in function is a compelling one, it is not a necessary component in understanding the functional cycle as a push-chain. However, this notion may again be helpful in understanding the second transition of the formal cycle. That is, in addition to the formal weight of the pre-verbal element, some preference for a correspondence between form and function may lead to its loss.

2.4   Explaining the cycles

With the mechanics and plausibility of the two kinds of scenarios in mind, it is useful to pause and reconsider what there is to be explained. Given our focus on the history of negation in English, the empirical facts to explain are the transitions from *ne* to *ne...not* and from *ne...not* to *not*. Taken together, these constitute both a functional and formal cycle. Importantly, we want to understand the role syntactic acquisition and pragmatic use might play in such explanations.

Before addressing each transition in turn, we note two things. First, both use and acquisition necessarily play a role in any change. This follows from the simple fact that learners have to acquire a language to use it, and other speakers have to use a language for learners to acquire it. So, in a certain trivial sense, both must play some role in language change. In what follows, we will be interested more specifically in the way that use modifies the evidence available to acquisition, and the way that acquisition acts on the evidence from
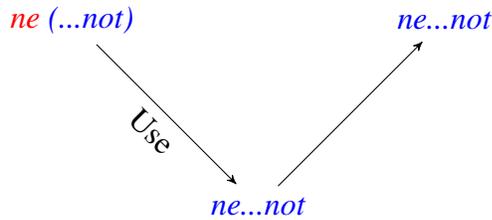
Figure 2.8: Pragmatic use as the cause of the first transition

use. Second, our goal is not to explain why forms come about in the first place, but rather how their introduction leads to change. That is, we are not aiming to explain why *ne...not* is introduced into English. We take variation in the forms of negation as a consequence of the broader fact of language variation.

Regarding the first transition, we noted that we can think of it as a kind of push-chain where the pre-verbal form is being pushed out by the embracing form as it increases in frequency. We are interested in what is causing the increase in the frequency of the embracing form and thus the pushing. There are at least two possibilities. First, it could be the case that pragmatic use leads to an increase in the embracing form. This compounds over time to the point where the embracing form is used exclusively. This means that it is the only form present in the linguistic evidence for learners. Thus, learners will only acquire the embracing form. We can represent this schematically as in Figure 2.8, where the top row indicates the grammatical knowledge of speakers and the lower row indicates their use of the forms provide by their grammatical knowledge.

So, the upper left *ne (...not)* indicates grammatical knowledge that includes both the pre-verbal and embracing form. We indicate the increase in the embracing form due to use as the downwards arrow in Figure 2.8. The result of use is that embracing form is used categorically in the evidence available to subsequent learners. While we only show a single step, this process could just as well be the cumulative effect of use over several iterations. The important things is that use is the force acting to steadily increase the frequency of the
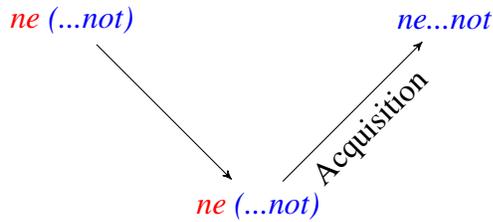
Figure 2.9: Syntactic acquisition as the cause of the first transition

embracing form over time, as opposed to acquisition.

While pragmatic pressures are often taken to be the cause of the pushing in the push-chain, there are certainly other options. A second possibility is that syntactic acquisition leads to an increase in the embracing form. This compounds over time to the point where the embracing form is the only one learned. Thus, it is the only one available to speakers to use. We can represent this schematically as in Figure 2.9, where the top row again indicates grammatical knowledge and the lower indicates the use of forms.

So, the upper left *ne (...not)* indicates grammatical knowledge that includes both the pre-verbal and embracing form. But, there is no increase in the embracing form due to use. We indicate the increase in the embracing form due to acquisition as the upwards arrow in Figure 2.9. The result of acquisition is that only the embracing form is acquired. Again, while we only show a single step, this could just as well be the result of several iterations. Crucially, it is acquisition rather than use that is driving the increase in the embracing form.

While we can present these two causes in isolation, another possibility is that both use and acquisition interact over time, giving rise to the embracing form. However, the important thing in any case is the form that an explanation must take. That is, for pragmatic use or syntactic acquisition to serve as explanations of the first transition we must demonstrate how they cause the increase in the frequency of the embracing form. We need a model of use or acquisition that makes both qualitative and quantitative predictions. That is, we need a model that predicts *that* the first transition will happen, as well as *how* it will happen.
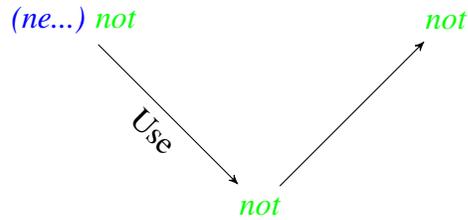
Figure 2.10: Pragmatic use as the cause of the second transition
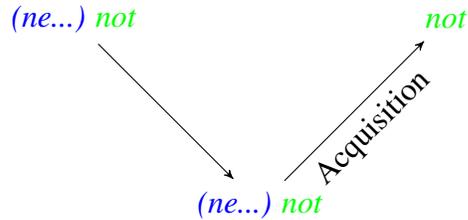
Figure 2.11: Syntactic acquisition as the cause of the second transition

The same requirements holds for the second transition. We can represent what a pragmatic or syntactic explanation would like in Figures 2.10 and 2.11. For either to serve as an explanation, we would have to demonstrate how they cause the increase in *not* over time. Note that these requirements are independent of whether it makes sense to conceive of the second transition as a push-chain or whether we think that use or acquisition play any role whatsoever. In fact, it serves as an important check on any models we posit for the first transition. For example, if we have no reason to think that pragmatic use is what drives the second transition, then a model of the first transition based on use should not predict the second transition.

Summary

In this chapter we made the important terminological distinction between the formal and functional Jespersen cycles and noted the logical relationship between the two phenomena. An explanation of the formal cycle requires an explanation of the functional cycle, but

not vice versa. importantly, given that the functional cycle can occur independently of the formal cycle, an explanation of the first may be fundamentally different from an explanation of the second. This guides our approach to explaining the two processes in the history of English, observed as the transitions from *ne* to *ne...not* to *not*.

In Part I we pursue an explanation of the functional cycle based on pragmatics. We present a mathematical framework for modeling how meaning is signaled in a population over time and apply it to modeling how pragmatic use leads to the transition from *ne* to *ne...not*, but not the transition from *ne...not* to *not*. In Part II we turn to the formal cycle as a whole and evaluate whether a model of syntactic acquisition can explain either of the transitions, from *ne* to *ne...not* and from *ne...not* to *not*.

Part I

The Functional Cycle

Chapter 3

Evolutionary Game Theory

> We repeat most emphatically that our theory is thoroughly static. A dynamic theory
> would unquestionably be more complete and preferable.
> –Von Neumann and Morgenstern (1944:44-45)

> We shall now take up the "mass-action" interpretation of equilibrium points...It is
> unnecessary to assume that the participants have full knowledge of the total structure of
> the game, or the ability and inclination to go through any complex reasoning process.
> –Nash (1950b:21)

Game theory is a branch of applied mathematics originally developed by Von Neumann and Morgenstern (1944) that models human decision-making when the actions, outcomes, and preferences of multiple agents are intertwined. This differs from decision theory, where a single agent is faced with a choice to bring about his or her most preferred outcome. Rather, a game arises when the choices made by each player crucially depend on those made by others. It is this interdependence that distinguishes game theory and makes it such a useful tool in understanding social interactions, from the routes we choose to drive to

33

work to the way we use words to mean things.

The rest of this chapter provides an outline of the crucial concepts for our application of evolutionary game theory to language change. First, we offer a general overview of the dimensions along which games vary. We focus on two simple games, and develop intuitions about expected behavior. Second, we define a set of solution concepts that can be applied to predict behavior in the two simple games. We discuss the limitations of solutions concepts for predicting behavior. Third, we supply game dynamics and examine the trajectory of the population under those game dynamics for the two simple games. We then we define signaling games as a natural model of communication and present the simplest case. Finally, we present a generalized means of describing the dynamics of arbitrarily complex signaling games.

3.1   The role of rationality

Taking up one aspect of this first example, we will assume that the vast majority of people would prefer to avoid car accidents. When driving a car, this is straightforwardly achievable by always being on the opposite side of the road of the next oncoming vehicle. We can think about this from the perspective of two drivers approaching each other. Each driver has the choice of driving on the left or the right. Both parties would prefer to coordinate on driving on whichever side. The side of the road that best achieves the outcome of not crashing is irrelevant, but it depends on how both parties drive. If both parties choose to drive on the right, from their own perspective, then they will avoid each other. Likewise if they choose to drive on the left.

The solutions to this interaction are readily apparent, everyone should choose one side of the road or the other. Provided that both drivers choose the same side, then neither has an incentive to switch. This fundamental insight into behavior in games is what defines a

Nash equilibrium: the action of all agents is optimal given the choices of all other agents. Yet, grounding this solution in rationalistic terms requires a little more effort. First, it requires that all agents are *rational*, acting to maximize their own benefit. This is implicit in our example above, most humans prefer to avoid being injured. Second, it requires that players have *common knowledge* of the game structure, including the actions available to all players, the outcomes of different combinations, and know that they all know these things, and so forth. At least for this game it follows naturally from the number of sides of the road there are, and the ability to reason about collisions, but it is not always so straightforward. Finally, it requires that agents have *equilibrium knowledge*, that they are able to accurately anticipate the actions of other players. This last requirement may seem readily apparent, at least for the example of driving, but it deserves further comment.

Depending on the country, we know that everyone will drive on one side of the road. In fact, there are laws and legal repercussion for not doing so. Yet, if we abstract away from our everyday experience in this particular case, the problem becomes clearer. For example, imagine that you are asked to choose between two options, that another person has been asked to choose between the same two options, that both of you have received the same instructions, and that you will both succeed if you choose the same option. These starker terms may seem to border on the absurd, but they bring the problem into relief. Their absurdity lies in the fact that it does not do to reason about *this* option or *that* option, because there is nothing to distinguish the two. There is no *thisness* or *thatness* to motivate one choice or the other. That is, there is no rule of the road that simply states "Drive on the right". Instead, we are faced with the problem of deciding the undecidable, as Aristotle, a bit sarcastically put it in his *On the Heavens*:

> [A] man, being just as hungry as thirsty, and placed in between food and drink, must necessarily remain where he is and starve to death.

We are particularly good at looking for reasons why one option is preferable, or more likely to be chosen than another. For example, if we were to enrich our description above to a choice between an option *A* or an option *B*, an intuitive rationale would be to choose *A*. The reason for doing so being that *A* comes first in the alphabet, and that both people might expect each other to know this. There is some intuitive *thisness* about option *A* that we can leverage on the assumption that it will be readily apparent to others as well. Schelling (1960:57) called these partial escapes from the dilemma of equilibrium knowledge *focal points* "for each person's expectation of what the other expects him to expect to be expected to do." In the pre-cellular age, Schelling posed the following scenario: "Tomorrow you have to meet a stranger in NYC. Where and when do you meet them?" The intuitive meeting place of Schelling's day was Grand Central station at noon. The rationale for both choices is that they are prominent in their respective geographical and temporal landscapes. Experimental results demonstrate that people are generally quite good at grounding their coordination via reasoning about these kinds of focal points (Mehta et al. 1994a,b, Bardsley et al. 2010).

But, even with our ability to reason about what we might expect others to expect us to expect them to expect us to do, we still cannot perfectly anticipate what others will do. That is, focal points are useful heuristics, but they are never foolproof: not everyone would choose option *A* or grand central station at noon. In fact, it would not be foolish for example to suggest meeting at Times Square at noon. The problem of justifying equilibrium in general still remains. Nash himself was well aware of this problem, which was in part what prompted his consideration of a "mass-action" interpretation of equilibria.[1] In short, this change in perspective moves from focusing on a single agent to a population of agents that

---

[1]The main motivation was understanding *mixed Nash equilibria*, which involve some probability distribution over actions. Nash (1950b,a, 1951) proved that such an equilibrium is guaranteed to exist. What it means for an individual to adopt the corresponding *mixed strategy* is not unproblematic. For a discussion of these problems see Aumann (1985), and for the population approach see Rosenthal (1979).

need not have either common knowledge or equilibrium knowledge. This perspective was naturally taken up in applications of game theory to biology, where loosening some of the equilibrium requirements was both reasonable and necessary (Maynard Smith 1982). This move also came with the reinterpretation of the game-theoretic machinery: actions were not choices to be made by the individual, but genetically pre-determined responses; the utility derived from a particular outcome was not a numerical representation of preferences, but a measure of Darwinian *reproductive fitness*.

This turn towards the biological allowed for further developments towards the explicitly dynamic theories that Von Neumann and Morgenstern (1944) envisioned. Drawing on the mechanics of population genetics, Taylor and Jonker (1978) offered the first *evolutionary game dynamics*, christened the *replicator dynamics* by Eigen and Schuster (1979), which explicitly defined how a population changes over time due to biological reproduction. This new approach allowed for insights into not just whether a population would remain at some equilibrium state if it started there, but also if the population would reach that equilibrium state in the first place. While successively detailed refinements of Nash equilibria were proposed with more and more unrealistic assumptions, evolutionary game theory offered a means of understanding equilibrium behavior under much simpler assumptions.

Despite its biological roots, however, work in evolutionary game theory has discovered profound connections with the rationalistic approach to equilibrium behavior. That is, the equilibrium predictions of the rationalistic approach often correspond to the effect of natural selection captured by evolutionary game dynamics. Perhaps more surprisingly, some of these game dynamics can be derived not just from the mechanics of biological reproduction, but from particular kinds of decision making or behavior. For example, the replicator dynamics defined by Taylor and Jonker (1978) can also be derived from particular forms of imitation (Schlag 1998, Björnerstedt and Weibull 1996) and learning (Börgers and Sarin 1997). The fact that such diverse foundations yield the same dynamics is both surprising

and compelling. More importantly, these commonalities show that if we interpret *evolution* in this broader sense that includes both biological and social change, then evolutionary game theory can serve as a powerful tool.

The rest of this chapter provides an outline of the crucial concepts for our application of evolutionary game theory to language change. First, we offer a general overview of the dimensions along which games vary. We focus on two simple games, and develop intuitions about expected behavior. Second, we define a set of solution concepts that can be applied to predict behavior in the two simple games. We discuss the limitations of solutions concepts for predicting behavior. Third, we supply game dynamics and examine the trajectory of the population under those game dynamics for the two simple games. We then we define signaling games as a natural model of communication and present the simplest case. Finally, we present a generalized means of describing the dynamics of arbitrarily complex signaling games.

3.2   A typology of games

There are several major dimensions along which games can vary.[2] Here we focus on several dimensions that are relevant for the case of communication and thus language change. The first dimension we will consider is the order which players make their decisions. If all players make a decision at the same time, then the game is a *simultaneous* game. If players make their decisions in a particular order, then the game is a *sequential* game. Communication is sequential insofar as we separate out the transmission and interpretation of a signal. Thus, we will be concerned with sequential games.

The second dimension along which games vary is whether or not players have pri-

---

[2]See Dixit et al. (2004) for a gentle introduction to game theory, Tadelis (2013) for a thorough but balanced introduction, and Osborne and Rubinstein (1994) and Fudenberg and Tirole (1991) for more advanced mathematical treatments.

vate information that affects others' payoffs. If players do have private information then the game is one *incomplete information*, otherwise it is one of *complete information*. Intuitively, this is the crucial aspect of communication. Speakers have some private information that hearers do not. We cannot read minds, thus we have to communicate. This asymmetry of information is crucially tied to the third dimension along which games vary. Namely, if the set of actions available to the different players differs or if the payoff from the same action differs, then the game is considered *asymmetric*. If the actions are identical for all players and if the outcome of actions are independent of who takes them then the game is *symmetric*. There are clearly two distinct roles in the act of communication, thus we will be concerned with asymmetric sequential games of incomplete information.

In this section we introduce to simple simultaneous games of complete information. We develop the intuitions required for the application of equilibrium concepts and dynamic analysis before moving on to the more complicated asymmetric sequential games of incomplete information known as *signaling games* that will be used in subsequent analysis.

The first simple example we consider is parallel to the case of choosing which side of the road to drive on, and is often referred to as a *coordination game*. There are two players who we will refer to as *Row* and *Column*, for reasons that will become clear later on. Each player has a choice between two options, *A* and *B*. We we will refer to these options as *pure strategies*.[3] That is, the set of pure strategies available to *Row* is $S_R = \{A, B\}$, and the set of actions available to *Column* is $S_C = \{A, B\}$. The set of all combinations of *Row* and *Column* strategies constitute the *strategy profiles* of the game. That is, for each $s_i \in S_R$ and $s_j \in S_C$, the tuple $\langle s_i, s_j \rangle$ constitutes a strategy profile.

Now, we can capture the intuition that both players prefer to avoid crashing by defining utility functions over outcomes for both players. *Row*'s utility function is a function from

---

[3]This is in contrast to *mixed* strategies, which specifiy a probability distribution over pure sender strategies, $\sigma = p_1 s_1 + ... + p_k s_k$, where $\sum_i p_i = 1$.

	A	B
A	$1, 1$	$0, 0$
B	$0, 0$	$1, 1$

Table 3.1: Coordination Game

strategy profiles to real numbers $U_R : \langle s_i, s_j \rangle \to \mathbb{R}$. Likewise *Column*'s utility function is a function from strategy profiles to real numbers $U_C : \langle s_j, s_i \rangle \to \mathbb{R}$. In fact, we can drop the subscripts entirely, if we define the two utility functions in the following way.

$$U(s_i, s_j) = U_R(s_i, s_j) = U_C(s_j, s_i) = \begin{cases} 1 & \text{if } s_i = s_j \\ 0 & \text{else} \end{cases} \tag{3.1}$$

Note that utilities are only important insofar as they represent an ordinal ranking over outcomes. In other words, there is nothing particularly special about 0 and 1 as opposed to say 3.277 and 110 other than the fact that in both cases the first is less than the second. However, there is something important about the fact that $U(A, A) = U(B, B)$. This simply reflects the fact that either rule of the road is equally useful for avoiding collisions.

With these utility functions defined, we can present the game in a slightly more compact *normal form* as in Table 3.1. The strategies for *Row* and *Column* are listed to the left and above respectively, hence the names. By convention, for each cell of Table 3.1, the payoff for *Row* is listed first, followed by the payoff for *Column*. The reason why coordination games are symmetric is that both U_R and U_C define symmetric matrices. Intuitively, we could rotate Table 3.1 around its diagonal and nothing would change; the payoffs for one player are the transpose of the other. In fact, we could represent the payoffs in this symmetric game as a matrix that captures this symmetry.

$$U(s_i, s_j) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.2}$$

40

	H	T
H	$0,1$	$1,0$
T	$1,0$	$0,1$

Table 3.2: Matching Pennies

With these definitions in place, we return to the reasoning above. Namely, both players want to coordinate on playing either A or B. If one player chooses A, then the other should as well. If one player chooses B, then the other should as well. But, there is no guarantee that the other player will choose A or B. Again, simply knowing that both would be good outcomes does not guarantee that both players will choose one or the other.

We can modify the game in a simple but substantial way by altering the payoff structure. That is, imagine the case where the utility functions of the sender and receiver are defined as the following.

$$U_R(s_i, s_j) = \begin{cases} 1 & \text{if } s_i \neq s_j \\ 0 & \text{else} \end{cases} \tag{3.3}$$

$$U_C(s_j, s_i) = \begin{cases} 1 & \text{if } s_i = s_j \\ 0 & \text{else} \end{cases} \tag{3.4}$$

That is, the payoffs of *Row* are diametrically opposed to that of *Column*. This game, called *matching pennies* has the following simple rules. Both players choose between heads H or tails T. One player wins if the coins match, and the other player wins if the coins do not match. The resulting game is summarized in Table 3.2. Matching pennies is a constant-sum asymmetric game, the components of every cell in Table 3.2 sum to a constant. Except in the trivial case where all outcomes yield the same utility for both players, constant-sum games are asymmetric. That is, we cannot represent the payoffs as a single value without losing crucial information.

Unlike the coordination game, players want to anti-coordinate with each other. That is, if *Row* plays H, then *Column* would prefer to play H, but if *Column* plays H then *Row* would prefer to play T. This poses another interesting problem. If both players have opposing interests, how should they act? It is clearly not feasible to pick one strategy or the other, as this would leave either player open to exploitation by the other. It would seem then that the best both players can do is to flip a coin and play whichever side comes up.

Before turning to defining equilibria, we need to note one last aspect of utility. As a case in point, let us assume that when playing matching pennies as *Row*, we know that our opponent has a certain probability of playing H, and thus a certain probability of playing T. Let p be the probability that our opponent plays H. Given that we are never certain of which strategy our opponent will play, we want to know the *expected utility* of choosing one strategy or another. With slight abuse of notation, this is just the following.[4]

$$E[U_R(H)] = pU_R(H, H) + (1 - p)U_R(H, T) = (1 - p) \qquad (3.5)$$

Expected utility is crucial in determining behavior in cases where there is uncertainty due to probabilistic behavior.

3.3 Equilibria as solution concepts

A game by itself does not constitute a model in the technical sense. It is a mathematical structure that describes the players preferences and strategies, but it does not predict their behavior. To do so it must be supplemented with a *solution concept* that predicts the conditions under which particular outcomes will obtain. Here we define two related solution concepts, note how they apply to the simple games we defined above, and note their

---

[4]This is just the expected value of a random variable. To take a simple example, imagine if we flipped a fair coin, counting heads as a 1 and tails as a 0, then the expected value of the coin flip would be $\frac{1}{2}$

limitations for predicting behavior.

3.3.1 Nash Equilbria

Perhaps the most widely used solution concept is that of a *Nash equilibrium*, which specifies when players have an incentive to unilaterally deviate from a given strategy profile. We provide a definition and apply the concept to our simple games.

For a given strategy profile where s_i is the strategy of a particular player, let s_{-i} be the set of strategies for all other players. We have the following definition.

Definition 1. A strategy profile $\langle s_i^*, s_{-i}^* \rangle$ is a *Nash equilibrium* if and only if:

- $\forall i, \forall s_i \in S_i$, such that $s_i \neq s_i^*$, $E[U(s_i^*, s_{-i}^*)] \geq E[U(s_i, s_{-i}^*)]$

This simply states that no players can do better by individually changing his or her strategy from the equilibrium profile. No one has an incentive to change, so everyone keeps doing what they are doing, hence the term equilibrium. That is, every player's current strategy is a *best response* to everyone else's. Note that this best response need not be unique to constitute an equilibrium. This follows from the fact that the inequalities in the definition are not strict. We can, however, impose uniqueness in these best responses by requiring that the inequalities in the definition be strict. A *strict Nash equilibrium* results, meaning that players can only ever do worse if they unilaterally deviate from the equilibrium.

We can now apply this solution concept to the games we described above. First, we look at the coordination game. We already noted that coordinating on one strategy or the other seems to be an intuitive solution. The conditions for $\langle A, A \rangle$ and $\langle B, B \rangle$ to be Nash equilibrium are the following.

$$E[U(A, A)] \geq E[U(B, A)] \tag{3.6}$$

$$E[U(B,B)] \geq E[U(A,B)] \tag{3.7}$$

In fact, looking at the payoffs in Table 3.1, both of these outcomes meet the definition of a strict Nash equilibrium. Both players can only do *worse* by deviating from the equilibrium.

There is one last kind of equilibrium to consider. Imagine that both players choose A with probability p and B with probability $(1-p)$. The expected utility for both pure strategies are the following.

$$E[U(A)] = pU(A,A) + (1-p)U(A,B) = p \tag{3.8}$$

$$E[U(B)] = pU(B,A) + (1-p)U(B,B) = (1-p) \tag{3.9}$$

Note that these are equal to each other when $p = \frac{1}{2}$. Now suppose that an agent plays a mixed strategy σ that strikes this balance by playing each strategy half of the time. We can show that this mixed strategy is a Nash equilibrium by comparing how another mixed strategy would fare against it. First, note that the expected utility of σ against itself is $\frac{1}{2}$.

$$\begin{aligned}
E[U(\sigma,\sigma)] &= \frac{1}{2} \cdot \frac{1}{2}U(A,A) + \frac{1}{2} \cdot \frac{1}{2}U(A,B) + \frac{1}{2} \cdot \frac{1}{2}U(B,A) + \frac{1}{2} \cdot \frac{1}{2}U(B,B) \\
&= \frac{1}{2}
\end{aligned} \tag{3.10}$$

Second, consider the expected utility of any alternate strategy σ' that plays A with probability p and B with probability $(1-p)$. Note that this includes the pure strategies as extremes, A where $p = 1$ and B where $p = 0$.

$$E[U(\sigma', \sigma)] = \frac{1}{2} \cdot pU(A, A) + \frac{1}{2} \cdot pU(A, B) + \frac{1}{2} \cdot (1 - p)U(B, A)$$
$$+ \frac{1}{2} \cdot (1 - p)U(B, B) \tag{3.11}$$
$$= \frac{1}{2}$$

That is, if the other player plays according to strategy σ, then no matter what the other does, they will receive a fixed expected utility. This means that if the other player plays each option evenly, there is nothing to be gained from switching from the same strategy. That is, any mixed strategy will do just as well.

$$E[U_R(\sigma, \sigma)] \geq E[U_R(\sigma', \sigma)] \tag{3.12}$$

Thus, there are two pure Nash equilbria and one mixed Nash equilibria in this coordination game. This mixed Nash equilibrium is the unsatisfying solution mocked by Aristotle above. That is, by committing to neither, both players are worse off; each player receives an expected utility of $\frac{1}{2}$ compared to the expected utility of 1 at the pure strategy equilibria. But intuitively, this indecision is precarious, any slight reason to choose one or the other would suffice to break the symmetry and motivate one solution or the other. We find that this intuition is indeed the case when we turn to dynamics below.

In contrast to the coordination game, there are no pure strategy equilibrium for matching pennies. For any pure strategy profile, one of the players will have an incentive to change her strategy: for $\langle A, A \rangle$ and $\langle B, B \rangle$, *Row* will have an incentive to deviate; $\langle A, B \rangle$ and $\langle B, A \rangle$, *Column* will have an incentive to deviate. In fact the mixed strategy where both players split the difference between heads and tails is the unique Nash equilibrium. To see this note that the mixed strategy yields an expected utility of $\frac{1}{2}$ and that any alternate mixed strategy also yields an expected utility of $\frac{1}{2}$. Note that since any mixed strategy yields the

same payoff against the mixed strategy, then it is not strict.

3.3.2 Evolutionarily Stable Strategies

While the concept of a Nash equilibrium is particularly useful, we we are interested not just in the behavior of two individuals, but rather the aggregate behavior of a population. Thus, we might want a broader notion of equilibrium. In particular, we might ask whether an entire population is susceptible to change, rather than the behavior of two individuals. We define a criterion that captures this level of description, note its limitations, and apply it to our simple games.

At the population level, the most relevant concept is that of an *evolutionarily stable strategy* (Maynard Smith and Price 1973, Maynard Smith 1982). In a symmetric game where $U(s^*, s)$ is the payoff to an agent using strategy s^* against an agent using strategy s, we have the following definition.

Definition 2. A strategy s^* is an evolutionarily stable strategy (ESS) if and only if, for all alternate strategies s:

- $U(s^*, s^*) \geq U(s, s^*)$
- If $U(s^*, s^*) = U(s, s^*)$ then $U(s^*, s) > U(s, s)$

Note that the first condition is the same as that of a Nash equilibrium. The second condition limits evolutionarily stable strategies to a proper subset of Nash equilibria. A straightforward interpretation of this definition can be had by imagining a population composed entirely of individuals playing s^*. Suppose some small proportion, $1 \gg \epsilon > 0$, of individuals playing the alternative s is introduced into the population. The following inequality holds if and only if s^* is an ESS.

$$(1 - \epsilon)U(s^*, s^*) + \epsilon U(s^*, s) > (1 - \epsilon)U(s, s^*) + \epsilon U(s, s) \qquad (3.13)$$

For any sufficiently small influx of the alternate strategy, the expected payoff of the incumbent strategy in the resulting population is strictly greater than that of the alternate. Intuitively, s^* is stable because selection will carry the population back to playing s^*. Loosely speaking, a strategy is evolutionarily stable if it is resistant to invasion by alternate strategies. That is, when an entire population plays the incumbent strategy the incumbents do at least as well as an alternate. If the alternate does as well, then the incumbent strategy should do better than an alternate in an all-alternate population.

While evolutionarily stable strategies are a useful refinement of Nash equilibria, they only offer particular kinds of insight. That is, under the standard *equilibrium methodology* (Huttegger and Zollman 2013), the state of a population is often justified by some special status. For example, the persistence of particular states of the population is due to it being an evolutionarily stable strategy. However, this kind of explanation looses some of its explanatory bite if there either multiple evolutionarily stable strategies, or none. If we apply the solution concept to the games we described above, these problems becomes clear. First, we look at the coordination game. We can use the reformulation of the criterion above to give the conditions for $\langle A, A \rangle$ and $\langle B, B \rangle$ to be evolutionarily stable strategies.

$$(1 - \epsilon)U(A, A) + \epsilon U(A, B) > (1 - \epsilon)U(B, A) + \epsilon U(B, B) \qquad (3.14)$$

$$(1 - \epsilon)U(B, B) + \epsilon U(B, A) > (1 - \epsilon)U(A, B) + \epsilon U(A, A) \qquad (3.15)$$

That is both of the pure strategy Nash equilibria for the coordination game are evolutionarily stable strategies if $\frac{1}{2} > \epsilon$. Assuming that $1 \gg \epsilon > 0$, this is guaranteed, but this condition also has another interpretation. That is, $\frac{1}{2} > \epsilon$ defines the *invasion barrier*, the influx required to move the population from one equilibrium to the other. This makes intuitive sense, if the population of the United States moved to the United Kingdom, and

insisted on maintaining the same traffic patterns, then everyone would best be served by adopting the majority pattern.

We can also consider the condition for the mixed Nash equilibrium to be an evolutionarily stable strategy.

$$(1 - \epsilon)U(\sigma, \sigma) + \epsilon U(\sigma, A) > (1 - \epsilon)U(A, \sigma) + \epsilon U(A, A) \tag{3.16}$$

Remembering from above that the expected utility of the mixed strategy against itself and any strategy against the mixed strategy are both $\frac{1}{2}$, we know that the mixed strategy state is not evolutionarily stable. Note that the conditions for B to invade the mixed state are the same as for A. The invariability of the mixed strategy state validates our intuitions about its instability.

Turning to asymmetric games, such as matching pennies, where players have different roles, we are actually concerned with two populations. To determine the stability of the two populations we can consider a *symmetrized* version of the game, where individuals have strategies for both roles. A strategy in the symmetrized game is thus a strategy profile of the asymmetric game. In the case of symmetrized asymmetric games, the criterion for stability is actually simpler. A strategy profile is evolutionarily stable if and only if it constitutes a strict Nash equilibrium in the original asymmetric game (Selten 1980). This means that the mixed strategy in matching pennies is not an evolutionarily stable strategy, the game simply does not have one. Thus, the justification of a particular state of the population as an evolutionarily stable strategy is off the table.

Maynard Smith (1982:8) noted both of these problematic conditions, but considered them to be ultimately unproblematic insofar as they are obvious exceptions. However, the obviousness of exceptions becomes less clear as games become more complex. Huttegger and Zollman (2013) provide a compelling example of where intuitions fail and focusing on

equilibria is particularly misleading. The solution to these limitations is to use both static equilibrium concepts along with an explicitly dynamic perspective.

3.4 Evolutionary game dynamics

While the term suggests a dynamic interpretation, equilibria, including evolutionarily stable strategies, are static solution concepts. They tell us whether a strategy is resistant to invasion or innovation, but tell us nothing about how the population got there in the first place, or where it might go next. To understand how a population changes over time, a particular set of *game dynamics* must be supplied.[5] These provide a description for how a population, or populations, changes from one point in time to the next.

In what follows we will focus on the replicator dynamics. The reasons for doing so are twofold. First, they are the most extensively studied dynamics and share a number of mathematical properties with a larger set of game dynamics (Hofbauer and Sigmund 1998). This makes establishing results and connecting them to a broader class of dynamics more straightforward. They have also found broad application in modeling economic (Samuelson 1997, Cressman 2003) and social behavior (Skyrms 1996, 2004, 2010). Second, they have a natural interoperation as particular form of learning (Börgers and Sarin 1997, Fudenberg and Levine 1998), which we will return to later on. We begin by defining the form of the replicator dynamics for symmetric and asymmetric games. We then outline general methods for assessing the stability of particular states. With these definitions and methods in place, we return to our simple games from above.

The intuition behind the replicator dynamics is that more successful strategies increase

---

[5]See McElreath and Boyd (2008) for a general introduction to the mathematical modeling of social evolution, Gintis (2000) for a comprehensive introduction to evolutionary game theory, and Weibull (1997) and Hofbauer and Sigmund (1998) for more advanced mathematical treatments. For a general introduction to non-linear dynamical systems see Strogatz (2014) and Hirsch and Smale (1974) for a more mathematical treatment. See Sandholm (2010) for a thorough and insightful grounding of evolutionary game dynamics in individual behavior.

in frequency. In particular, strategies that are more successful than the population average increase in share of the population. For the case of a symmetric game, let $\mathbf{x}$ be a vector that represents the composition of the population, where x_i represents the proportion of s_i, and the utility function is represented by the matrix $\mathbf{A}$. The change of x_i in the population can be given as the following.

$$\dot{\mathbf{x}}_i = \mathbf{x}_i((\mathbf{Ax})_i - \mathbf{x}^{\mathbf{T}}\mathbf{Ax}) \tag{3.17}$$

We begin by interpreting the second part of the equation. The first portion simply provides the expected utility of s_i given the current composition of the population. The second portion simply provides the average expected utility given the current composition of the population. The difference between these two is how much better a strategy s_i does than the average. If s_i does better than the average, then its share of the population will increase; if s_i does worse than the average, then its share will decrease. This increase or decreases is weighted by the current proportion of strategy in the population.

As a simple case in point, let us return to the coordination game from above. To start, let p be the proportion of the population playing A. Then the population vector is given by the following.

$$\mathbf{x} = \begin{pmatrix} x_A \\ x_B \end{pmatrix} \tag{3.18}$$

The payoff matrix, is given by the following.

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.19}$$

Since there are only two strategies in the symmetric game, we only need to keep track

of the proportion playing one, $x_A = 1 - x_B$. For simplicity, we will keep track of the proportion playing A, so we can drop the subscript, and note that it evolves according to the following replicator dynamic.

$$\dot{x} = x(1 - x)(2x - 1) \tag{3.20}$$

The *rest points* of the replicator dynamic for the coordination game are the points where this equation is equal to zero. That is, they are the points where the motion of the system is at rest. The rest points of the dynamic are exactly the states that correspond to Nash equilibria. Namely, when $x = 0, 1, \frac{1}{2}$.

We can define the parallel case for asymmetric games, where there are two populations denoted by $\mathbf{x}$ and $\mathbf{y}$, along with two payoff matrices $\mathbf{A}$ and $\mathbf{B}$

$$\dot{\mathbf{x}}_i = \mathbf{x}_i((\mathbf{Ay})_i - \mathbf{x^T Ay}) \tag{3.21}$$

$$\dot{\mathbf{y}}_i = \mathbf{y}_i((\mathbf{Bx})_i - \mathbf{y^T Bx}) \tag{3.22}$$

For matching pennies, we define the following components.

$$\mathbf{x} = \begin{pmatrix} x_H \\ x_T \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} y_H \\ y_T \end{pmatrix} \tag{3.23}$$

The payoff matrix, is given by the following.

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.24}$$

Again, simplifying our notation by omitting subscripts, these yield the following coupled

replicator dynamics.

$$\dot{x} = x(1 - x)(1 - 2y)$$

$$\dot{y} = y(1 - y)(2x - 1)$$

(3.25)

Note that the rest points of the replicator dynamics for matching pennies include more than the single Nash equilibrium we noted. That is, the point $(x, y) = (1, 1)$ where both populations only play heads is a rest point of the dynamics. However, this is certainly not an outcome we would expect. Thus, we need more information than just the rest points. We are, however, interested in the relation between the entire state space and these rest points. In particular, we are interested in whether the populations will move towards or away from particular states. That is, we want to know about the stability of the rest points.

A bit more formally, for a state space $\mathbf{x} = \{x_1, x_2, ..., x_n\}$, a solution trajectory through the state space over time $\mathbf{x(t)}$ is governed by the differential equations defined by the game dynamics $\dot{\mathbf{x}} = \{\dot{x}_1, \dot{x}_2, ..., \dot{x}_n\}$. The rest points of the system are states where the differential equations vanish; any trajectory starting at a rest point will remain there at rest. For a rest point $\mathbf{x}^*$, if $\mathbf{x(0)} = \mathbf{x}^*$ then $\mathbf{x(t)} = \mathbf{x}^*$ for all times $t > 0$, since $\dot{\mathbf{x}} = 0$. The criteria we listed above, then, require the following definition.

Definition 3. A rest point $\mathbf{x}^*$ is asymptotically stable if and only if:

1. For every neighborhood V of $\mathbf{x}^*$, there is a neighborhood V' of $\mathbf{x}^*$ such that if $\mathbf{x(0)} \in V'$ then $\mathbf{x(t)} \in V$ for all $t > 0$.

2. For some $\epsilon > 0$, for all $\mathbf{x(0)}$ such that $\epsilon > \|\mathbf{x(0)} - \mathbf{x}^*\|$, then $\mathbf{x(t)} \to \mathbf{x}^*$ as $t \to \infty$.

The first condition requires that all trajectories that start near a rest point stay near it. Small perturbations from the rest point do not lead the system away from the rest point, but rather to some nearby state. If only this first condition is met, then the rest point is

weakly or *Liapunov* stable. The second condition requires that all points near to the rest point converge to it in the limit. The set of all states that converge to a rest point constitute its basin of attraction. If both these conditions are met, then the rest point is *strongly* or *asymptotically* stable. If neither of these conditions is met, then the rest point is *unstable*.

Now that we have specified the relevant rest points, we want to know whether they are stable. That is, we want to know if the game dynamic will carry the population towards the rest points or away from them. In a single dimension, we evaluate the derivative of the dynamic at the rest points. If it is positive, then the rest point is unstable. If it is negative, then the rest point is asymptotically stable. The intuitive notion here is that rest points with a positive derivative correspond to "hills" and rest points with a negative derivative correspond to "valleys". An object may be at rest at the top of a hill. But, if we give it a small push, it will roll down one of the sides of the hill. An object in a valley between two hills will also be at rest. And, if we give it a small push, it will return to this position. All we are doing when we evaluate the derivative is finding out whether the population will be carried away from the rest point or back to it. We are finding out whether it is a "hill" or a "valley" under the game dynamic.

The derivative of the replicator dynamic of the coordination game is given by the following. Note that if we evaluate this at each of the rest points, we get the expected results.

$$\lambda = -6x^2 + 6x - 1 \tag{3.26}$$

Both of the evolutionarily stable strategies are asymptotically stable, whereas the mixed strategy equilibrium is unstable.

In more than one dimension, we can gain insight into the behavior of the system near a rest point by considering its *linearization*. That is, just like taking the derivative of function in one dimension gives us a line, we can do the same thing for more dimensions (Hirsch

and Smale 1974). To do so we examine the eigenvalues of the Jacobian, which is matrix of all first-order partial derivatives of a set of vector-valued function. $\mathbf{f} = \{f_1, .., f_m\}$, $\mathbf{x} = \{x_1, .., x_n\}$

$$\mathbf{J} = \frac{d\mathbf{f}}{d\mathbf{x}} = \begin{bmatrix} \frac{\partial f_1}{x_1} & \cdots & \frac{\partial f_1}{x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{x_1} & \cdots & \frac{\partial f_m}{x_n} \end{bmatrix} \tag{3.27}$$

For the case of matching pennies, we have two dimensions, and thus the following Jacobian and eigenvalues.

$$\mathbf{J} = \begin{bmatrix} (2x - 1)(2y - 1) & -2x(1 - x) \\ 2y(1 - y) & -(2x - 1)(2y - 1) \end{bmatrix} \tag{3.28}$$

$$\lambda_1 = \sqrt{12x^2y^2 - 12x^2y + 4x^2 - 12xy^2 + 12xy - 4x + 4y^2 - 4y + 1}$$
$$\lambda_2 = -\sqrt{12x^2y^2 - 12x^2y + 4x^2 - 12xy^2 + 12xy - 4x + 4y^2 - 4y + 1} \tag{3.29}$$

For any of the rest points along the edge of the state space that are not equilibria we have at least one positive eigenvalue, thus they are all unstable. Turning to the mixed strategy equilibrium, we have purely imaginary eigenvalues, $\lambda_1 = -\frac{i}{2}, \lambda_2 = \frac{i}{2}$. These suggest that the mixed strategy is a *center*, around which the system orbits. However, we have to be cautious in interpreting these results.

The trouble lies with the fact that the center is not *hyperbolic*, because the real parts of its eigenvalues are all equal to zero. The *Hartman-Grobman theorem* states that if a rest point is hyperbolic, then the linearization at that point gives an accurate picture of the dynamics. However, when the rest point is not hyperbolic, then there is no guarantee that the linearization yields an accurate picture. The simplest means of testing the linearization
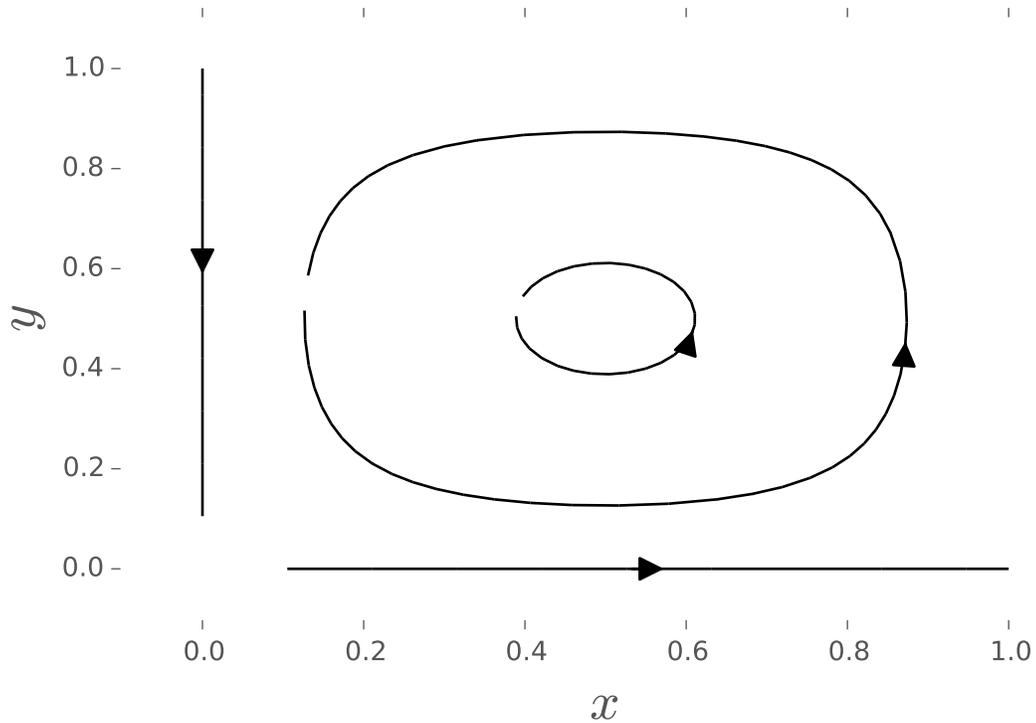
Figure 3.1: Phase diagram of replicator dynamics for matching pennies

is to compare the result to numerical simulations. For example, in Figure 3.4 we have numerically constructed a *phase diagram* of the replicator dynamics. And indeed, we find that the linearization accurately predicts the behavior of the system near the rest point. The populations circle along closed orbits around the center in a counter-clockwise manner.

In what follows we will largely use numerical simulations to explore the behavior of more complicated systems, but with reference to the concepts we have introduced in this section. This will allow us to explore more and more complicated systems.

3.5 Signaling games

Now that we have outlined both the static and dynamic tools used to understand behavior in games, we turn to the asymmetric sequential games of incomplete information, *signaling*

games, that we will use subsequently to understand both communication and language change. First, signaling games are introduced. The relevant structures and definitions are presented along with a canonical example. Second, we consider the solution concepts from the previous sections. This allows us to determine the existence and characteristics of equilibria. Finally, we consider how signaling evolves according to the replicator dynamics.

It should be noted that there are multiple traditions that have contributed to the development of signaling games. In the Economic tradition the work of Harsanyi (1967, 1968a,b) on games of incomplete information and Spence's (1973) model of job signaling have been central. In the Philosophical tradition, the work of Lewis (1969) has been the most influential. His response to the Quinean skepticism that meaning could arise by convention has had the most direct impact on Linguistics. This account, however, rests on the assumption that speakers and hearers have perfectly-aligned interests. This assumption will figure in the exposition of this section, mostly for the simplicity it affords us in the examples. In the next section we turn towards the consequences of loosening it.

Signaling games offer an intuitive model of communication between agents. A signaling game consists of two players, a sender and a receiver. The sender has some private piece of information, $t \in T$, drawn according to some commonly known probability distribution, δ. The piece of information can be thought of as some fact about the state of the world. For example, we might think of it as information that the sender wants to convey to the receiver. The sender chooses a message, $m \in M$, to send to the receiver. The receiver does not know what state of the world actually holds and must choose an action, $a \in A$, as an interpretation of the message sent. That is, the receiver is faced with the problem of inferring the state of the sender given the message.

The outcome of the game is determined by the state of the sender, the message sent, and the action taken by the receiver. The sender and the receiver have preferences over these outcomes, which are given by utility functions, U_S and U_R respectively, which map

outcomes to real numbers: $U_S, U_R : T \times M \times A \to \mathbb{R}$. Note that the message sent can figure into the utility functions. For example, if one message is costlier than another, then the utility can be adjusted to reflect this. In what follows we will only consider costless or *cheap talk* signaling where all message incur the same cost, and thus do not affect the structure of the utility functions.

As a simple example of a signaling game, consider the case where there are two states, two messages, and two actions: $T = \{t_0, t_1\}$, $M = \{m_0, m_1\}$, and $A = \{a_0, a_1\}$. We can represent the structure of the game in *extensive form* as in Figure 3.2. The uppermost node, labeled δ, determines the likelihood of either of the two states occurring. The lines from the uppermost node represent one state or the other obtaining. The nodes labeled S indicate the points in the game where the sender makes a decision regarding which message to send. The nodes labeled R indicate the points where the receiver must make a decision regarding how to interpret the message. The dashed lines between receiver nodes indicates that the receiver is uncertain as to which state holds after hearing a given message. The nodes at the bottom represent the outcome and the players' preferences, as determined by the utility functions.
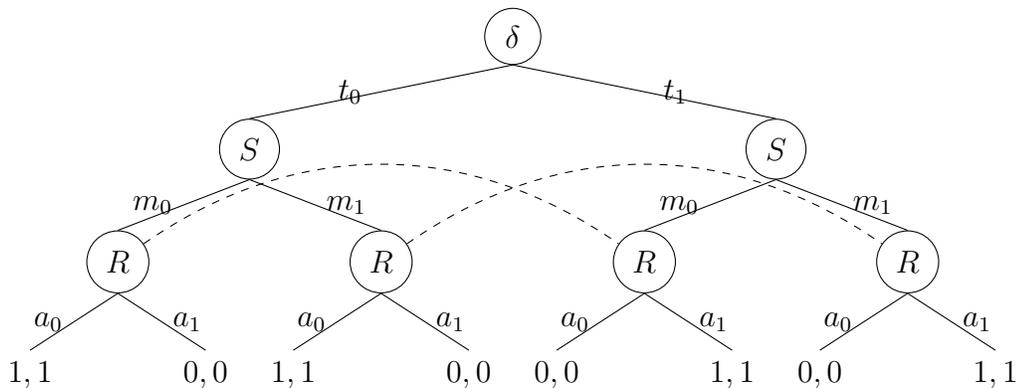
Figure 3.2: Lewis signagling game with two states, two messages, and twos actions.

As alluded to above, with the exception of the topmost node and the leaves, all other nodes are *decision nodes*. That is, they represent those points in the game where a particular

player must make a decision. An *information set* consists of a set of decision nodes for a given player, which that player cannot distinguish. For example, the receiver is in an information set after hearing m_0. She is not sure whether she is in the node beneath state t_0 or t_1 and she must choose between a_0 and a_1. She is likewise uncertain after receiving message m_1. Again, her inability to distinguish the two states is indicated by a dashed line. An information set can also consist of a single decision node. For example, the sender is never uncertain about the state he is in. Thus the information sets for the sender only ever consist of a single node.

For each player, a *strategy* specifies which action to take at all *information sets* for a player.[6] We will refer to the set of all such sender strategies as $S : [T \rightarrow M]$, and the set of all such receiver as $R : [M \rightarrow A]$. All pure sender and receiver strategies are summarized in Figure 3.3. The sender and receiver strategies that combine to yield a one-to-one mapping from states to actions are called *signaling systems*. Thus, the strategy profiles $\langle s_1, r_1 \rangle$ and $\langle s_4, r_4 \rangle$ constitute signaling systems.

The set of possible combinations of sender and receiver strategies constitute *strategy profiles*. That is, a sender strategy in the set of possible sender strategies, $s \in S$, and a receiver strategy in the set of all possible receiver strategies, $r \in R$, yield a strategy profile $\langle s, r \rangle$. Each strategy profile determines the outcome of the game. Crucially, each player's utility function depends on the state of the sender and the action taken by the receiver. That is, their respective utilities are a function of state and action. As an example, consider the case where both sender and receiver prefer successful communication. Then they both receive their preferred payoffs if there is some correspondence between the sender's state and the receiver's action. For example, in Figure 3.2 both sender and receiver prefer the

---

[6]For simplicity, we only consider *pure* sender (receiver) strategies, functions from states to messages (messages to actions). *Mixed* strategies are a straightforward generalization. A mixed sender strategy specifies a probability distribution over pure sender strategies, $\sigma = p_1 s_1 + ... + p_k s_k$, where $\sum_i p_i = 1$. Similarly, a mixed receiver strategy specifies a probability distribution over pure receiver strategies, $\rho = q_1 r_1 + ... + q_k r_k$.
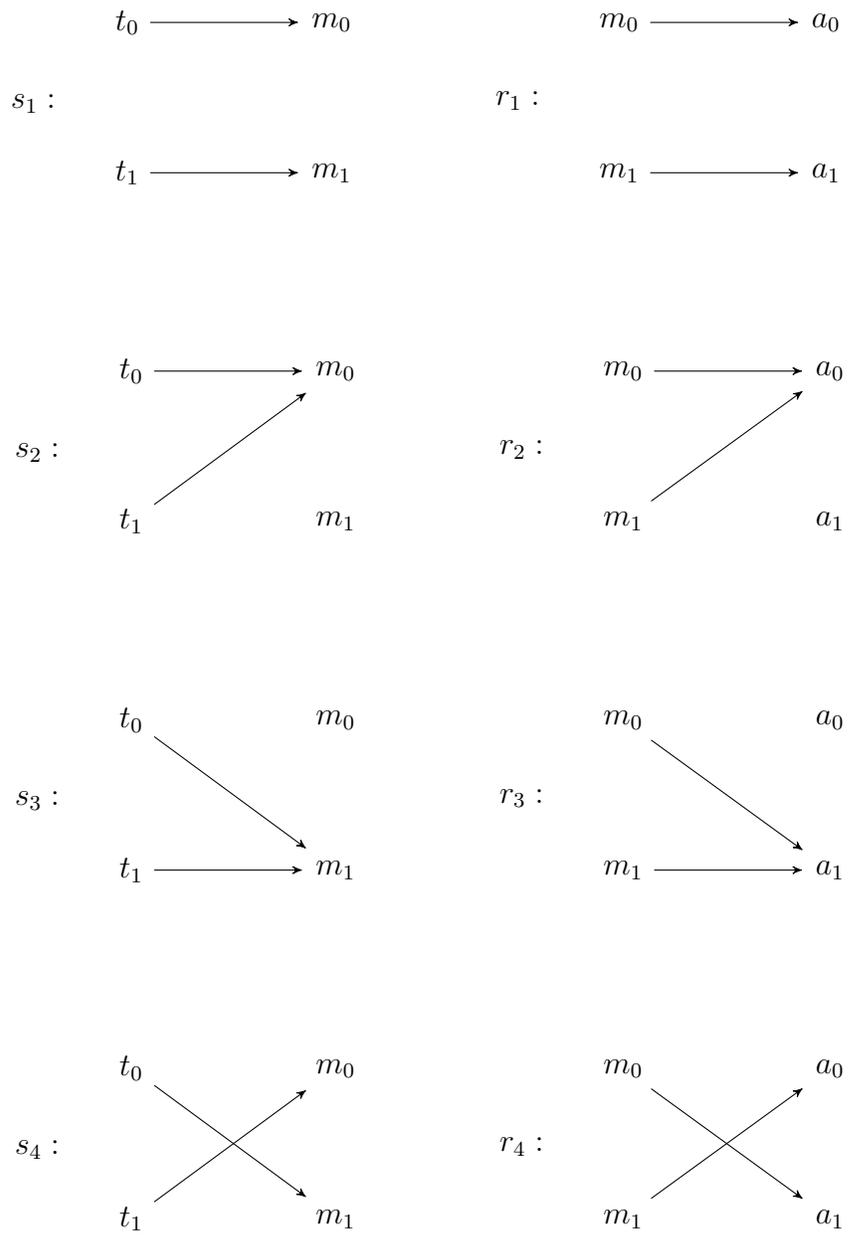
Figure 3.3: Sender and Receiver strategies for signaling game

	r_1	r_2	r_3	r_4
s_1	1	$\frac{1}{2}$	$\frac{1}{2}$	0
s_2	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
s_3	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
s_4	0	$\frac{1}{2}$	$\frac{1}{2}$	1

Table 3.3: Payoff matrix for signaling game

state and action to be the same in the following sense.

$$U_S(t_i, a_k) = U_R(t_i, a_k) = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{else} \end{cases} \tag{3.30}$$

Note that this reflects the assumption that both sender and receiver have the same preference over outcomes.

Given that the different states occur with certain probabilities, and we are interested in how particular strategies do on average, we consider the expected utility for a given strategy profile. This is simply the expected value of the utility functions given the two strategies, which can be given in the case of a discrete state space as in our example above.

$$E[U_S(s, r)] = \sum_t \delta(t) \cdot U_S(t, r(s(t)))$$
$$E[U_R(s, r)] = \sum_t \delta(t) \cdot U_R(t, r(s(t))) \tag{3.31}$$

For each possible state the sender and receiver strategies determine an outcome. $s(t)$ is the message the sender will employ and $r(s(t))$ is the action the receiver will take given that message. The expected utility is the sum of these outcomes weighted by the probability of the state that yields them. Assuming that the two states are equiprobable, we can construct the payoff matrix for the signaling game as in Table 3.3. Since payoffs are symmetric, only one number is presented.

60

There are several Nash equilibria in the game, but only the signaling systems $\langle s_1, r_1 \rangle$ and $\langle s_4, r_4 \rangle$ are evolutionarily stable strategies, because they are the only strict Nash equilibria. In fact, given the structure of the payoffs it is possible to show that these signaling systems are the unique asymptotically stable rest points of the system, which attract the entire interior of the state space (Hofbauer and Sigmund 1998:82). This means that one of the two signaling systems is almost guaranteed to evolve.

3.6   The behavioral dynamics of signaling

While the simplest signaling game can be analyzed in a fairly straightforward manner, we might be interested in larger, more complicated games. For example, we might wonder about the dynamics of signaling if the state space were continuous rather than discrete. If we were to approach this by first enumerating all sender and receiver strategies, the dimensions of the system would explode exponentially. One means of controlling for this increase suggested by Hofbauer and Huttegger (2015) is to consider behavioral strategies insofar as no information is lost in doing so (cf. Kuhn 1953).[7] The basic idea is that instead of there being a single sender population and a single receiver population, there are multiple sender and receiver populations. Each sender population corresponds to a particular state, and messages compete to be used in that state. Each receiver population corresponds to a message, and actions compete to be used in response to that message. We start off by defining the components necessary for our analysis. Once we have defined these components we can simulate the game dynamics.

---

[7]Erol Akçay (p.c.) has suggested a similar approach to the problem.
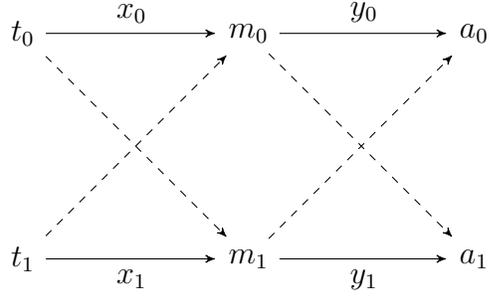
Figure 3.4: Signaling game in behavioral space

3.6.1 The simplest signaling game

To begin, we apply this new framework to the simplest non-trivial signaling game where we have two equiprobable states, two messages, and two actions. The structure of the behavioral replicator dynamics is determined by the structure in Figure 3.4. x_0 represents the probability that m_0 will be used in the t_0 sender population. Likewise x_1 represents the probability that m_1 will be used in the t_1 sender population. The dotted lines indicates probabilities that can be derived from others. All told, then, the behavioral replicator dynamic for this signaling game gives rise to a four-dimensional system.

We will generalize the parameters a bit to match the simpler games we used above. Namely, let the sender and receiver payoffs be the following matrices, with parameters α and β.

$$\mathbf{A} = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix} \tag{3.32}$$

$$\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.33}$$

These yield the general behavioral sender and receiver replicator dynamics.

$$\dot{x}_0 = x_0(1 - x_0)(y_0 + y_1 - 1)(1 - 2\alpha)$$

$$\dot{x}_1 = x_1(1 - x_1)(y_0 + y_1 - 1)(1 - 2\beta)$$

(3.34)

$$\dot{y}_0 = y_0(1 - y_0)\left(\frac{x_0 + x_1 - 1}{x_0 - x_1 + 1}\right)$$

$$\dot{y}_1 = y_1(1 - y_1)\left(\frac{x_0 + x_1 - 1}{x_1 - x_0 + 1}\right)$$

(3.35)

The case where $\alpha = \beta = 0$ corresponds to the lewis signaling game we introduced in the previous section, which is also related to the coordination game we started with. This is a signaling games with common interests. The case where $\alpha = \beta = 1$ corresponds to signaling with matching pennies. This is a signaling game with conflicting interests. The cases where $\alpha = 0, \beta = 1$ and $\alpha = 1, \beta = 0$ also offer an interesting comparison. These are games of at least partial common interest. We address each of these below.

3.6.2   Signaling under common interests

Since the receiver dynamics do not vary as we change the parameters, we omit them. Under common interests the sender dynamics are the following.

$$\dot{x}_0 = x_0(1 - x_0)(y_0 + y_1 - 1)$$

$$\dot{x}_1 = x_1(1 - x_1)(y_0 + y_1 - 1)$$

(3.36)

Typical results for a numerical solution are shown in Figure 3.6.2, where we plot x_0 against y_0 and x_1 against y_1, with circles represent the starting state of both of trajectories. That is, we represent the four-dimensional system as two two-dimensional systems overlaid
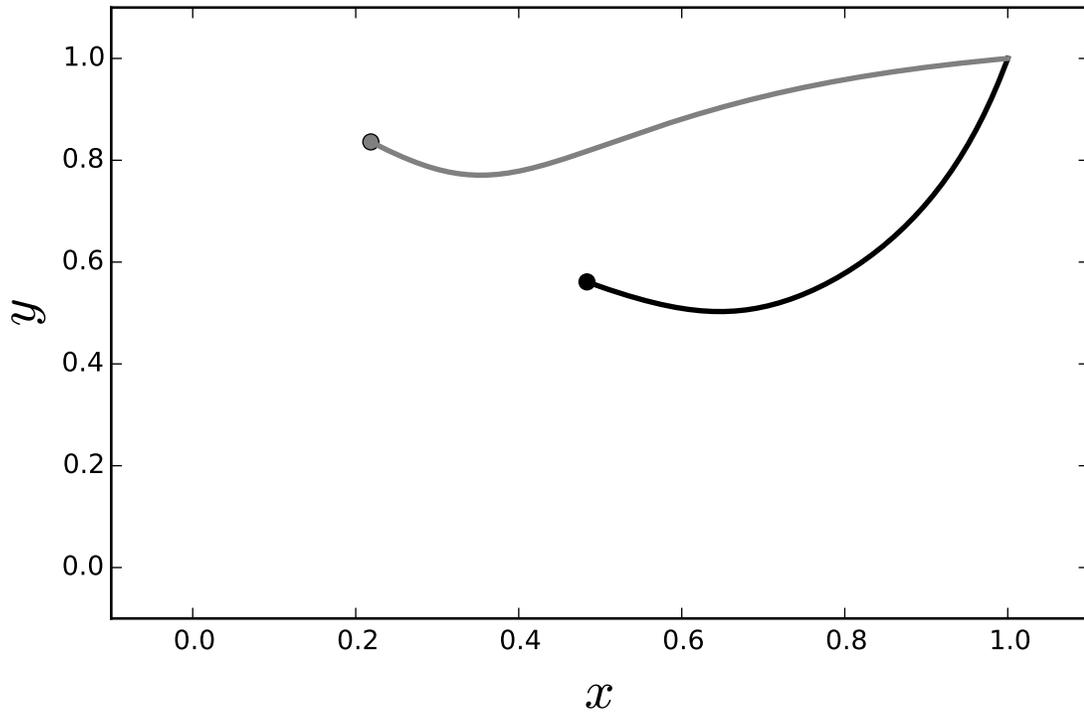
Figure 3.5: Solution trajectories for behavioral replicator dynamics for common interest signaling

on each other. The established result is exactly what we see. Namely, the system either converges to the origin or to the upper right-hand corner. Referring back to Figure 3.4 it is clear that these constitute the signaling systems of the signaling game as we described them. That is, these are the points that guarantee a one-to-one mapping between states and actions.
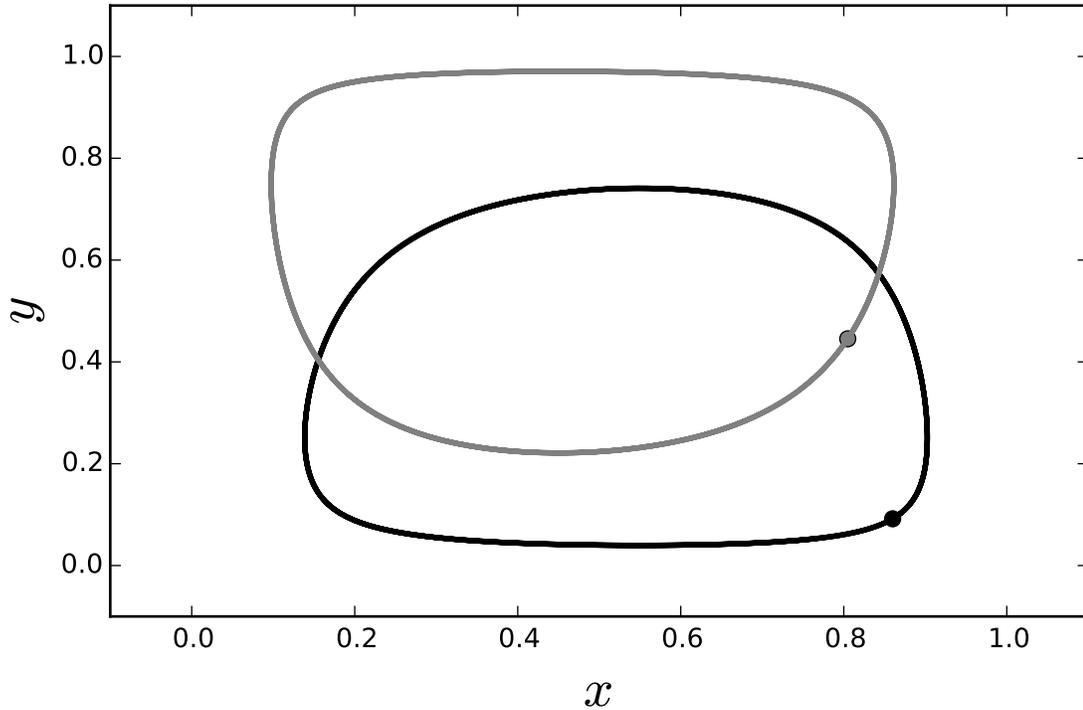
Figure 3.6: Solution trajectories for behavioral replicator dynamics for conflicting interest signaling

3.6.3 Signaling under conflicting interests

If we alter the payoff structure to reflect that of matching pennies, then the following behavioral sender dynamics result.

$$\dot{x}_0 = x_0(1 - x_0)(1 - y_0 - y_1)$$
$$\dot{x}_1 = x_1(1 - x_1)(1 - y_0 - y_1)$$

(3.37)

Typical results for numerical solution are shown in Figure 3.6.3. Interestingly, much like in the case of matching pennies, the system exhibits closed orbits.

An interesting consequence of this behavior is that despite the fact that the sender does not want to signal what action he is going to take, there is some amount of information

carried by the signal. This can quantified by calculating the *Kullback-Leibler divergence* or *information gain* due to the signal.

$$KL(m) = \sum_t P(t \mid m) log \left(\frac{P(t \mid m)}{P(t)} \right) \qquad (3.38)$$

The basic intuition behind this formula is that it allows us a way to compare the receiver's expectations prior to receiving the message, which is determined by the probability distribution over states, to the posterior distribution after having heard the message. The difference between these two distributions is the information gained by having received the message. The average information gain for the two signals is shown in Figure 3.6.3. As a point of reference, if the messages perfectly corresponded to states, as they do in a signaling system, then the average information gain would be equal to one. In other terms, the information gained from the signal would be one bit, because it would allow us to distinguish between two equiprobable states. The fact that signals still carry information, even under diametrically opposed interests is both interesting and surprising (cf. Sato et al. 2002, Wagner 2012).

Summary

In this chapter we outlined the basic framework that will be used subsequently. We noted the complementary role of static and dynamic methods for understanding the evolution of populations. We also presented a general method for describing the dynamics of signaling games and noted interesting properties of the simplest kind of signaling game. We now turn to our application of this framework to the functional cycle.
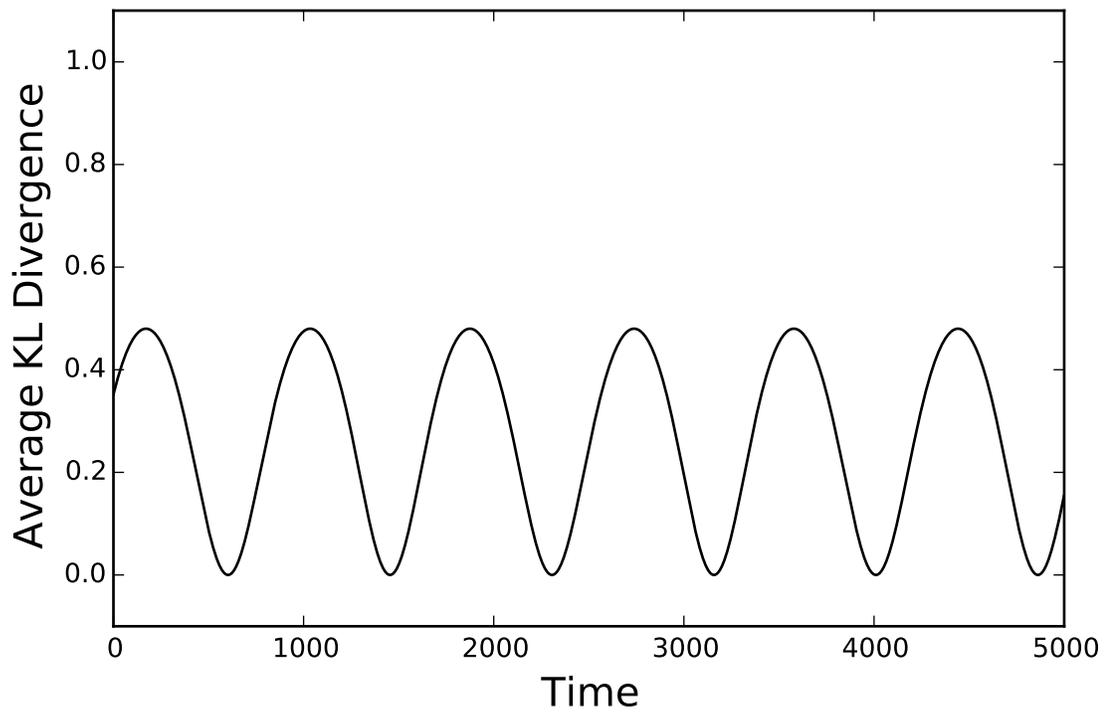
Figure 3.7: Average Kullback-Leibler Divergence for signals under conflicting interests

Chapter 4

Cycles

> [T]o say what a word means in a language is to say what it is in general optimal for speakers of that language to do with that word, or what use they are to make of it.
>
> –H.P. Grice (1989:299)
>
> I can't say 'It's cold here' and mean 'It's warm here' – at least, not without a little help from my friends.
>
> –David Lewis

Distinguishing between the formal and functional Jespersen cycles clarifies what needs to be explained. The functional cycle can occur independently of the formal cycle, so we need an explanation for it regardless. This motivates our focus in this chapter on the functional cycle. That is, we want to know why one form displaces another, taking over the meaning of plain negation. We also want to know why this new form can be displaced in further functional cycles. In particular, we want to explain why in the history of English we observe emphatic *ne...not* increase in frequency and displace pre-verbal *ne*. To do so, we build a mathematical model of the pragmatic pressures that lead to this transition.

First, we provide an interpretation of the notion that the incoming form is *emphatic* with

respect to the incumbent form, that it conveys some special meaning. Namely, we note that the incoming form is initially restricted to contexts where the proposition being negated has just been introduced into the discourse, but expands to contexts where it is merely inferable from the discourse, and eventually to contexts where the proposition is brand new to the discourse. Second, we discuss experimental evidence that suggests why the incoming form spreads across contexts in the way it does. Speakers have difficulty in separating out their own private knowledge from what is common knowledge between themselves and their interlocutors. Given this difficulty, speakers adopt a heuristic that biases them towards their own perspective when assessing how closely connected a negated proposition is to the discourse. This leads speakers to treat propositions as more connected to the discourse than is warranted. Third, we incorporate these facts into a signaling game, determine the equilibria and dynamics, and show how the number of signals used interacts with speaker bias. Finally, we fit the resulting model to data from the functional cycle in Middle English and discuss the implications of the fitted parameters in light of the experimental evidence.

The main contributions of this chapter are twofold. The first contribution is that we offer the first explicitly dynamic model of the functional cycle that explains why the discourse constraints change in the manner that they do. While previous accounts have noted the constraints on the different forms of negation in the functional cycle, they have not explained the increase in the incoming form beyond the somewhat circular claim that the incoming form is overused. Here we argue that the functional cycle is a byproduct of our cognitive limitations in tracking common knowledge. Importantly, while the the driving force of the cycle is rooted in the cognition of individuals, it leads to change because of the social interactions between individuals in a population. In Grice's terms, the optimal use of and response to different forms of negation are both moving targets. Their coupled movement is what underlies the functional cycle.

The second contribution of this chapter is that the model we present offers an information-

theoretic foundation for grammaticalization. For example, the intuitive notion of *bleaching* is simply the loss of information carried by a signal as it becomes obligatory. Grounding qualitative terms in this quantitative framework offers a new perspective on diachronic changes in how meaning is signaled. Moreover, it allows for a broader conception of meaning as the information carried by linguistic signals. This broader conception of meaning has the potential to unify our analyses of semantic, pragmatic, and sociolinguistic meaning.

4.1   Emphasis as activation

We begin with the intuitive notion of *emphasis*. We know it when we hear it, but what it amounts to is often left implicit in accounts of the functional cycle. There are two general functions that have been suggested as candidate interpretations. The first is that emphatic negation widens and strengthens negation to preclude exceptions (Kadmon and Landman 1993). That is, emphatic negation signals a stricter *standard of precision* for interpreting a proposition (cf. Austin 1962, Lewis 1970, Landman 1991, Krifka 1995). The second interpretation is that emphatic negation serves to deny an expectation, or mark the contradiction of a potentially implicit assertions (Detges and Waltereit 2002, Kiparsky and Condoravdi 2006). That is, emphatic negation signals the relationship between the proposition being negated and the preceding discourse. In what follows we focus exclusively on this second interpretation, but return to the first later on as an important point of comparison.

Broadly speaking, this second interpretation of emphasis depends on the joint attention of speakers and hearers (cf. Chafe 1974, Prince 1981). More precisely, emphatic negation has been found to signal that the proposition being negated is *activated* (Dryer 1996, Schwenter 2005, 2006). A proposition is *directly activated* if its contents have just been explicitly introduced into the discourse, so it is present to the joint attention of speakers and hearers. A proposition is *indirectly activated* if its contents can be inferred from the

discourse either via an entailment or implicature. A proposition is *non-activated* if its contents have not been explicitly introduced into the discourse or it cannot be inferred from the preceding discourse.

There are two important things to note about activation. First, activation does not entail belief, nor vice versa. Participants in a discourse believe propositions that are not activated, and not all of the activated propositions are believed. This distinguishes activation from the notion of *common ground* (Stalnaker 1978), which consists of the set of propositions that both interlocutors believe to be true, or at least accept as true for some purpose (Stalnaker 2002:715-720). Second, and related to this first point, emphatic negation can be used both to negate an activated proposition or restate it. For example, if the activated proposition is p then negation can be used to negate that p. If the activated proposition is $\neg p$ then negation can also be used to restate that $\neg p$.

Crucial for our purposes is the fact that activation can reasonably be identified in historical corpora with the use of modern translations. Moreover, activation has been shown to have the same effect diachronically in the histories of French (Hansen 2009, Grieve-Smith 2009), Italian (Hansen and Visconti 2009, 2012), and English (Wallage 2013): the incoming emphatic form in all of these languages is initially restricted to use with either activated or directly activated propositions. For example, Wallage (2013) shows that in Early Middle English *ne...not* is overwhelmingly restricted to negating activated propositions. But, over time it spreads to negating non-activated propositions as well.

The following examples from the *Penn Parsed Corpus of Middle English* (Kroch and Taylor 2000) cited by Wallage (2013) demonstrate the transition. First, the emphatic form is restricted to contexts where it is used to negate directly activated propositions. Where $p =$ "They are deceived", emphatic negation can be used to deny the explicitly stated proposition that p. In all of the examples that follow, p and $\neg p$, or both are bolded in the translation of the passage.

(1) Alle ðo men ðe   swinkeð on ðessere swinkfulle world, alle he   swinkeð for
     All the men that labour in this   toilsome   world, all they labour for
     sumere hope ðe hie habbeð, ðe hem oft   eaten ande beswinkð ... Ac ðo
     some   hope that they have,   that them often at   end deceives ... But those
     ðe swinkeð for ðessere eadi   hope, hie *ne*   bieð *naht* becaht
     that labour   for this   blessed hope, they NEG are   not deceived.
     "All the men who labor in this toilsome whorld, they all labor for some hope they

     have which often **deceives them** in the end...But those who labor for this blessed

     hope, **they are not deceived**."

     (CMVICES,33.385, 1200 CE)

The incoming form can also be used to restate a proposition that has already been explicitly

negated. For example, where $\neg p =$"You don't know yourself", it can be restated.

(2) If u   ne cnawest e   seolf ... If u   *ne*   cnawest *naut* e   seolf
     If you not know   the self ... If you NEG know   not   the self
     "If **you do not know yourself**...If **you do not know yourself**"

     (CMANCRIW, II.80.941-948, 1230 CE)

Subsequently, the use of the incoming form is extended to being used to negate propositions

that are only indirectly activated. For example, we might suppose that a virtuous religious

rite with all the sin-cleansing properties of baptism would have some post-mortem benefits.

If we take the proposition resulting from this inference to be $p =$"[That rite] opened to them

the bliss of heaven", then the incoming form can be used to negate the proposition resulting

from the inference.

(3) and te lage hadde to   alle te mihtes te haueð nu   fulluht for ðat clensede
     and the law had   then all the virtues that has   now baptism for that cleansed
     te man of sinne: swa doð nu   fulluht ac it *ne* openede hem *noht* te blisse
     the man of sin: as   does now baptism but it neg opened   them not   the bliss

of heuene alse fulcneng doð us.

of heaven as baptism does us.

"And **that rite had then all the virtues which baptism now has**, for that cleansed

man of sin even as baptism now does, but **it opened not to them the bliss of heaven**

as baptism does to us."

(CMTRINIT, 87.1165, 1225 CE)

Similarly, the incoming form can also be used to state a negative inference. If we suppose

that renouncing one's sins requires being done committing them, $\neg p =$"I cannot renounce

my sins", then the incoming form can be used to state the proposition resulting from this

negative inference.

(4)   Ich nam   noht giet sad of mine sines and forti   *ne* mai ich hie   *noht*

        I   not-am not yet sated of my   sins and therefore neg can I   them not

        forlete.

        renounce

    "**I am not yet sated of my sins** and therefore **I cannot renounce them**."

(CMTRINIT, 75.1028, 1225 CE)

Finally, the incoming form can be used to negate propositions that are entirely new to the

discourse. That is, *ne...not* is used to negate a proposition that is not readily identifiable

either directly or indirectly from what has come before. It is useful to note that all of these

examples come from roughly contemporaneous documents.

4.2   Experimental evidence regarding activation

While we observe this trend in historical corpora of English and other languages, this does

not explain why the incoming form spreads across the contexts in the way it does. To

understand why, it is useful to consider experimental evidence demonstrating particular

communicative biases on the part of speakers. Namely, speakers' private knowledge persistently influences how they signal meaning.

For example, Wu and Keysar (2007) had pairs of participants play a communication game in fixed roles of speaker and hearer. Speakers and hearers jointly learned names for a set of abstract shapes. Speakers then learned several names for additional shapes privately. The experimenters varied the number of shape names that the speaker and hearer learned together, what the experimenters called the *informational overlap* between participants. The participants then played a game where the speaker directed the hearer to select a target shape from a set. Across trials the target shapes were evenly distributed between shapes whose names were learned together, shapes whose names were learned privately by the speaker, and shapes that were new to both participants. Presumably, using a name to refer to a shape is only felicitous if the name of the shape was learned together by both speaker and hearer. But, surprisingly, in trials where the target shape's name was private knowledge, the name was the first thing speakers said in 5% of trials where there was a low informational overlap and in 28% of trials where there was a high informational overlap. That is, speakers relied on their private knowledge more where there was a greater degree of informational overlap.

Note that this use of the privately known names was not a result of speakers forgetting the context in which the names were learned. Heller et al. (2012) replicated these findings and showed that speakers were incredibly accurate at distinguishing between names learned together versus names learned privately. Rather, Wu and Keysar (2007:4) suggest that these results point to speakers using a kind of *overlap heuristic*: when the informational overlap between yourself and your interlocutor is sufficiently extensive act *as if* they have all the same information as you. In fact, this is kind of combined *co-presence* and *community membership heuristics* proposed by Clark and Marshall (1981) to resolve the problem of *common knowledge*: that everyone knows that p, that everyone knows that everyone knows

that p, *ad infinitum*.[1] This heuristic and the speaker bias that it creates are a means for solving the difficult task of keeping track of what is private versus common knowledge. Assuming that an interlocutor knows roughly the same things about the world reduces the cognitive burden and simplifies things a great deal.

However, given that speakers' use of private knowledge varied across conditions, we might wonder whether this bias is specific to certain contexts. For example, speakers might be able to pay closer attention to the discourse and keep better track of things. That is, there might be two modes of thinking regarding the discourse (Keysar et al. 2003, Kahneman 2011). However, this bias is not subject to conscious manipulation. Wardlow Lane et al. (2006) had pairs of participants play a communication game in fixed roles of speaker and hearer. Four shapes of varying size and color were presented to the participants. One shape was visible to only the speaker, blocked from the view of the hearer by an occluder. Speakers were instructed to communicate information about a target shape visible to both participants so the hearer could identify it. In the test conditions, the item that was visible only to the speaker was the same shape as the target item, but varied along some relevant dimension (e.g. size, color). For example if the target shape was a blue triangle then the shape that was only visible to the speaker was a green triangle, and none of the other shapes visible to both participants were triangles.

Wardlow Lane et al. (2006) found that speakers modified their description more in the target condition. That is, speakers were more likely to say "The blue triangle" to refer to the target shape if there was a green triangle visible only to them. Speakers' private information leaked into what they said and how they said it. This happened even though speakers had direct evidence that only they could see the shape that contrasted with the target shape.

---

[1]This technical term was introduced into the Philosophical literature by Lewis (1969), but has a long history under various names. Whereas Clark and Marshall (1981) use the term "mutual knowledge" to refer to common knowledge, as it is commonly used *mutual knowledge* only requires that everyone know that p. Thus anything that is common knowledge is also mutual knowledge, but not *vice versa*.

In fact, this over-modification happened to an even greater extent when speakers were explicitly instructed to conceal information about their privileged information. Speakers have a difficult time inhibiting their perspective even when they want to, suggesting that speaker bias is persistent fact about communication.

These experimental results show how speakers' private knowledge persistently influences how they signal meaning. In particular, they show that speakers tend to assume that their interlocutors are overwhelmingly similar to them. While this experimental evidence deals with the referential domain, the results can be extrapolated to the propositional domain and activation in particular. That is, activation is defined by the joint attention of speakers and hearers. But, neither speakers nor hearers know whether or not a proposition is actually being attended to by an interlocutor. This problem is a perfect candidate for solution by the kind of heuristic described above. Namely, speakers can simplify the problem by assuming that what their hearers attend to is sufficiently similar to what they attend to.

One potential concern about this kind of heuristic is that mistakes seem unlikely. If a proposition is directly activated, then it is directly activated because it has just been mentioned. Similarly, if a proposition is indirectly activated, then it is indirectly activated because there is an entailment or implicature that does so. A response to these objections, particularly the first, lies in the nature of activation. While we have described discrete categories, Dryer (1996:481-482) rightly conceives of both direct and indirect activation as continuous measures. For example, while the utterance of a proposition directly activates it, this activation decays over time. As he puts it, the proposition is *deactivated* as time goes on and it passes from the joint attention of speakers and hearers. Similarly, while the utterance of one proposition may indirectly activate another via an inference, some propositions may be more *accessible* than others via inference. Some inferences are natural, whereas others are non-sequiturs. In fact, we might take activation as a whole to be constituted by degrees of inferability. Recently uttered propositions are high on the scale, whereas brand

new propositions are at the bottom end.

So, speakers might treat a proposition as more activated than is warranted for several reasons. For example, suppose that a speaker keeps thinking about a proposition p, but her interlocutor does not. This means that p is being deactivated as it has passed from the joint attention of both participants. However, the speaker still dwelling on p is not aware of this, and only has her own perspective to consider. Thus, she may still treat p as highly activated even though it is not. Similarly, a speaker may easily infer a proposition p from the preceding discourse due to her attention to particular aspects of the discourse. But, her interlocutor may only make the same inference with great effort. Thus, a speaker may treat p as highly activated even though it is not.

If speakers have a tendency to overestimate activation and use *ne...not* more than is warranted by the actual degree of activation, then how will hearers respond? Given hearers' response, how will speakers respond in turn? What does this mean for the functional cycle? To answer these questions we translate these experimental findings into a mathematical model that can be used to investigate meaning over time. This model consists of two components. First, we define the *stage game* that describes the interactions between speakers and hearers and captures speakers' bias towards overestimating activation. Second, we define the *game dynamics* that describe how a population of speakers and hearers change over time while playing the stage game.

4.3   A signaling game model of emphasis

We start by defining the components of the stage game that we will use to analyze the functional cycle. We define the states, messages, and actions along with their interpretations. We then turn to the utility function of senders and receivers as they relate to the preferences of speakers and hearers. Once we define the game we can determine its equilibria and the

dynamics of how speakers use different forms of negation to signal the activation of the proposition being negated.

First, let the set of states $T : [0,1]$ be the degree of activation of the proposition being negated, where $t = 0$ indicates a brand new proposition and $t = 1$ indicates a proposition that was the last thing uttered. Second, let the set of messages that the speaker sends be a finite set $M = \{m_1, m_2\}$. We can think of these as the incumbent and incoming form in the functional cycle respectively. So, for English, m_1 is *ne* and m_2 is *ne...not*. Finally, let the set of actions $A : [0,1]$ be the action taken by the hearer to interpret the message. For example, an action a_i can be thought of as an initial guess by the hearer about the level of activation t_i of the negated proposition.

With these components defined, it is important to make a conceptual clarification about the role of states as degrees of activation. In signaling games, the state is taken to be some piece of private information that the sender has about the state of the world. This is not quite accurate given that we assume that speakers never know the actual degree of activation. That is, they can never peer inside hearers' heads to verify what propositions are being attended to at any given moment. At best, speakers have a subjective estimate of the degree of activation. However, this subjective estimate is systematically related to the actual degree of activation.

To see this, suppose that both speakers and hearers have some subjective estimate of the activation of a proposition, call them t_S and t_R. For example, both might take their estimate to be the approximate amount of attention they are paying to a given proposition. Now, given that activation is defined in terms of the joint attention of speakers and hearers, then the actual state of activation t must be some function of the two subjective estimates. If a speaker's estimate of a proposition is that it is not activated $t_S = 0$ because she is not attending to it, then by definition it is not activated $t = 0$. In contrast, if a speaker's estimate of a proposition is that it is highly activated $t_S = 1$, this does not mean that it is

indeed activated since the hearer's estimate could be lower $t_R < 1$ because the hearer is not attending to it to the same degree.

A simple way of capturing the relation between the subjective estimates and the actual degree of activation is that $t = min(t_S, t_R)$. The actual level of activation is the highest degree that both participants would agree to based on their own subjective estimates. This is arguably what underpins our intuition that a form is infelicitous because a hearer did not or could not have had a sufficiently high subjective estimate of activation. If this is the relation between the subjective estimates and the actual state, the speaker's estimate stands in a particular relation to the actual state. Namely, the speaker overestimates the degree of activation. To see why, note that if $t_S \leq t_R$ then $t_S = t$, and if $t_S > t_R$ then $t_S > t$. Together these imply that $t_S \geq t$. In other words, on average the speaker overestimates the actual degree of activation.

Now, if speakers only have access to their own subjective estimates of the state of activation, then it makes sense that their preference are determined by that estimate. As far as the speaker is concerned, t_S is the actual degree of activation of the proposition being negated. It makes sense then that the speaker would want the hearer to infer that degree of activation. In particular, we suppose that speakers prefer for hearers to infer the degree of activation that is closest to their subjective estimate. The following utility function satisfies this constraint, it is maximized exactly where $a = t_S$.

$$U_S(t, a) = 1 - (a - t_S)^2 \tag{4.1}$$

While this function captures the speaker's preferences, it also introduces a new and undefined parameter, the speaker's subjective estimate. A simple way to address this is to posit a general functional shape $f(t) = t_S$ that is defined in terms of the actual state and other parameters but captures speakers' tendency to overestimate the actual state by guarantee-

ing that $f(t) > t$. There are infinitely many functions that satisfy this constraint, but a particular simple functional form has a natural interpretation in our case.

In their seminal work on signaling games of *information transmission*, Crawford and Sobel (1982) introduce a bias parameter $b \geq 0$ into the utility function of senders that indicates how aligned the goals of senders and receivers are. For example, where $b = 0$ their preferences are perfectly aligned, but for $b > 0$ they diverge. We can apply this directly to our case if we think about the bias parameter b as speakers' tendency to overestimate activation. Then the following simple linear function allows us to incorporate the tendency to overestimate into the speaker's utility function.

$$t_S = f(t) = t + (1 - t)b \tag{4.2}$$

This yields the following utility function which satisfies the constraint that $t_S \geq t$.[2] It is maximized for an action $a = t + (1 - t)b$. This means that speakers prefer that hearers take an action slightly higher the actual state of activation.

$$U_S(t, a) = 1 - (a - t - (1 - t)b)^2 \tag{4.3}$$

By changing the variables we obtain a functional form that depends only on states, actions, and the new bias parameter. In fact, this bias parameter has a natural interpretation in terms of speaker bias as a measure of how good or bad speakers are at keeping track of common versus private knowledge. The case where $b = 0$ corresponds to speakers developing the ability to read minds and accurately assess the actual state of activation. For $b > 0$ speakers have a tendency to overestimate the degree of activation and prefer a higher action on the part of hearers.

---

[2]This differs from the formulation in Crawford and Sobel (1982) where $U_S(t, a) = -(a - t - b)^2$. Their form allows senders to prefer actions $a > 1$, which has no interpretation in our model. That is, the degree of activation and interpretation of the degree of activation are both constrained to the unit interval. We also add a constant so that all payoffs are positive for the dynamic analysis in the next section.

Now, we might wonder whether speakers have access to the action taken by hearers. That is, if speakers cannot peer into the heads of hearers to determine the state of activation, does it makes sense to assume that they can somehow infer the reasoning process indicated by the action hearers take to interpret the message? However, speakers have incredibly rich sources of feedback from hearers. For example, this feedback includes the amount of time hearers take to respond, facial expressions, and backchannel cues (e.g. "Mhmm" versus "Huh?"), as well as verbal responses such as as requests for clarification or continuations of the discourse. All together then, it seems reasonable that speakers can recover the action taken by hearers.

With speaker preferences defined, we can think about hearers. Suppose that hearers want to accurately infer how the proposition being negated relates to the prior discourse. For example, a hearer does not want to overestimate the degree of activation of a proposition and expend too much effort on trying to discern how it is connected to the discourse. Similarly, a hearer does not want to underestimate the degree of activation and miss out on information regarding how a proposition fits into the discourse. The following utility function satisfies this constraint, it is maximized exactly when $a = t$.

$$U_R(t, a) = 1 - (a - t)^2 \tag{4.4}$$

That is, hearers do best when they accurately infer the actual degree of activation. Now, we might wonder again whether it is reasonable to assume that hearers have access to the actual degree of activation. They cannot read minds any more than speakers. However, hearers gain information as they reason about different potential degrees of activation.

To see why this is the case, first assume that speakers and hearers have the same reasoning capacities. That is, they would agree on what potential degrees of activation make some kind of sense given the discourse. So, we assume that speakers and hearers can both

identify a particular set of degrees of activation as making sense. Note that this does not require that speakers and hearers expend the same amount of effort in identifying the degree of activation, but rather that speakers and hearers have the same reasoning capacities. Here we will suppose that there is some function of states given the discourse that defines whether both speakers and hearers can identify a particular degree of activation. Namely, let $g(t)$ be a convex function such $g(t) = 1$ if and only if t can be identified by both speaker and hearer. Further, suppose that the lowest identifiable degree of activation corresponds to the actual state of activation, for all $t_i < t$, $g(t_i) = 0$. This simply means that the subjective estimates of speakers and hearers serves as a lower bound on what degrees of activation both speakers and hearers can reasonably identify.

Now, suppose that hearers reason in the following manner when they take an action to interpret the speaker's message. The hearer takes an action a_i, corresponding to an initial guess that the degree of activation of the proposition is t_i. By the same reasoning for speakers above, hearers overestimate the actual degree of activation $t_R \geq t$, so it makes sense that they would compensate for this by choosing an action such that $t_i < t_R$. Given this action, there are two possibilities. The degree of activation for the proposition could be identifiable or not. If the initial guess of the degree of activation is identifiable $g(t_i) = 1$, suppose that the hearer reasons about lower and lower degrees of activation; likewise, if the initial guess of the degree of activation is not identifiable $g(t_i) = 0$, suppose that the hearer reasons about higher and higher degrees of activation. In both cases the hearer will consider a degree of activation t_j such that $g(t_j) = 1$ and for all $t_k < t_j$, $g(t_k) = 0$. That is, she will eventually recover the actual degree of activation. Note also that the amount of effort put into finding the actual degree of activation grows with the distance between the initial guess and the actual degree. So, if speakers and hearers reason in a sufficiently similar manner, then hearers will be able to recover the actual state of activation.

The preferences we have defined for speakers and hearers are a way to represent the

experimental evidence we described in the previous section. That is, they allow us to state, in mathematical terms, the stage game which is the shape of the interaction. But, this shape by itself does not make any predictions about the behavior of individual speakers and hearers playing the game or the trajectory of a population of speakers and hearers over time. To understand these we need to determine the equilibria of the stage game and how a population changes while playing the stage game under a particular game dynamics.

4.4   Equilibria of the signaling game

Now that we have defined the components of the game, we turn to analyzing its properties. Determining the equilibria of the game we described will allow us to address several questions about how signals are used in the functional cycle. Broadly speaking, we want to know two things. First, we want to understand the relationship between senders' bias towards overestimating activation and the use of signals at equilibrium. In particular, we want to know how large this bias can be while still allowing for multiple forms to be used at equilibrium. Second, we want to know whether particular equilibria constitute evolutionarily stable strategies. If so, we want to know the strategies that are evolutionarily stable. If not, we want to know what strategies could invade the population and disturb an equilibrium. Broadly speaking, this amounts to determining the conditions for the functional cycle to occur. We begin by defining speaker and hearer strategies, the expected utility of different strategies, and then determine the evolutionarily stable strategies of the game.

The set of speaker strategies is all potential mappings from the unit interval to a discrete set $S : [0, 1] \to M$. This is problematic given that the domain is uncountable. To simplify things we consider the following condensed representation. Let $\mathcal{P}_n(T)$ 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 sender uses the message m_i. A speaker's strategy

is then a function from this partition to messages $S : [\mathcal{P}_n(T) \to M]$. 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, consider the case of two messages $\mathcal{P}_2(T)$, where m_1 and m_2 correspond to *ne* and *ne...not* respectively. For $t \in (0, t_1)$ a sender will use *ne* and for $t \in (t_1, 1)$ a sender will use *ne...not*.

In fact, this kind of threshold strategy is extremely close to what we observe in historical data. That is, as the functional cycle proceeds, the incumbent form *ne* is not evenly distributed across activation contexts, but largely negates non-activated propositions. The probability of using the two negative forms is overwhelmingly conditioned by the activation of the proposition being negated. For example, for Midddle English from 1150-1250 CE the conditional probabilities are the following (Wallage 2013:12)

$$p(\textit{ne} \mid \text{NON-ACTIVATED}) = .85 \tag{4.5}$$

$$p(\textit{ne...not} \mid \text{ACTIVATED}) = .84 \tag{4.6}$$

This means that specifying speaker strategies in this manner is both a useful and empirically accurate abstraction. Interestingly, this also gives some empirical credence to the push-chain scenario conception of the functional cycle; *ne* actually appears to be pushed down the scale of activation by *ne...not*.

The set of hearer strategies is all potential mappings from the set of messages to the unit interval $R : M \to [0, 1]$. 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_1 would be the hearer's response to message m_1, in this case *ne*, and a_2 would be the hearer's response to message m_2, in this case *ne...not*.

Now that we have defined the set of speaker and hearer strategies, we can ask what

strategies constitute evolutionarily stable strategies. Given that signaling games are asymmetric, this amounts to identifying the strict Nash equilibria of the game. This can be done by determining what strategies jointly maximize the expected utilities of speakers and hearers.

$$E[U_S(s,r)] = \int_T \left(1 - (r(s(t)) - t - (1-t)b)^2\right) p(t)dt \tag{4.7}$$

$$E[U_R(s,r)] = \int_T \left(1 - (r(s(t)) - t)^2\right) p(t)dt \tag{4.8}$$

These are exactly analogous to expected utility in the discrete case, where we summed over all the possible states. Again, $r(s(t))$ is the receiver's respond to the sender's message and yields an action, which determines the utility for both sender and receiver given a state.

We estimate the prior probability over states as a *beta distribution* over the set of states, parameterized by two shape parameters α and β, and often written as $\mathcal{B}(\alpha, \beta)$. Figure 4.1 shows the distribution for several values of α and β, including the uniform distribution $\mathcal{B}(1,1)$. These two shape parameters gives us quite a bit of flexibility in modeling different potential prior probability distributions. We use them extensively in the analysis that follows, so we note two things that should offer an intuitive conceptual foothold. First, the expected value of a beta distribution is given by $\frac{\alpha}{\alpha+\beta}$. So, for example, the expected value of the uniform distribution is $\frac{1}{2}$ given that $\alpha = \beta = 1$, which is what we see in Figure 4.1. Second, if we fix $\beta = 1$ and let α vary, then the distribution will be more and more skewed to the right as α grows larger. This also follows from the fact that the expected value $\frac{\alpha}{\alpha+1}$ gets closer and closer to one as α grows. Note that the mirror image case would hold if we fixed $\alpha = 1$ and let β vary.

We use the historical data in Table 4.1 from Wallage (2013:12) that shows the number of propositions by activation to estimate the estimate the prior distribution over states. The
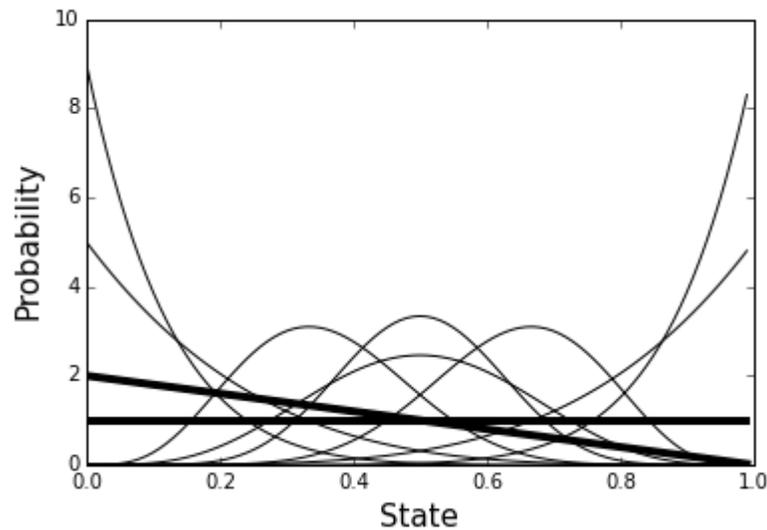
Figure 4.1: Beta distribution for various parameter values of α and β, including the uniform distribution $\mathcal{B}(1,1)$, and the empirical distribution $\mathcal{B}(1,2)$.

degree of activation in these examples is estimated using translations of the texts. However, there are reasons to be confident that activation can reliably be identified in historical corpora. These figures agree with similar estimates from contemporary corpora. For example, in a sample from a corpus of British English, Tottie (1991) finds that negation is only used 14% of the time with directly activated propositions. Likewise, in a corpus of American English, Thompson (1998) finds that negation is only used 5% of the time with directly activated propositions. These results suggest that the prior distribution is stable, with the preponderance of negation being used with brand new non-activated propositions. Intuitively, this distribution makes perfect sense: the majority of conversation is about introducing new information rather than treading the same old ground of what has already been said. If the prior distribution is indeed stable, then we can estimate it from the data pooled across time periods. A good fit to the data is the prior probability distribution $\mathcal{B}(1,2)$, also shown in Figure 4.1.[3]

---

[3]We treat each of the discrete categories as equal portions of the unit interval and find values of α and β such that $\int_0^{\frac{1}{3}} \mathcal{B}(\alpha,\beta)(t)dt \approx p(\text{NON-ACTIVATED})$, $\int_{\frac{1}{3}}^{\frac{2}{3}} \mathcal{B}(\alpha,\beta)(t)dt \approx p(\text{INDIRECTLY ACTIVATED})$, and

PERIOD	NON-ACTIVATED	INDIRECTLY ACTIVATED	DIRECTLY ACTIVATED
1150-1250	393	203	52
1250-1350	346	296	42
1350-1420	294	179	60
TOTAL	1033	678	154

Table 4.1: Distribution of sentence activation in PPCME (Kroch and Taylor 2000) from Wallage (2013)

To calculate the maxima of the expected utility functions, let $\langle s^*, r^* \rangle$ be an equilibrium strategy profile. In this case, the speaker strategy is defined by $s^*((0, t_1^*)) = m_1$ and $s^*((t_1^*, 1)) = m_2$. Likewise, the hearer strategy is defined by $r^*(m_1) = a_1^*$ and $r^*(m_2) = a_2^*$. We can determine the evolutionarily stable strategies by jointly maximizing speaker and hearer expected utility. That is, we solve a system of partial differential equations for t_1^*, a_1^* and a_2^*. For any amount of bias, the maximizing values are shown in Figure 4.2.[4] Along the horizontal axis is the amount of speaker bias, the vertical axis represents the point at which speakers partition the state space and the actions of hearers in response to the forms. For any value of the speaker bias b, the solid black line represents the point at which speakers partition the state space t_1^* and the dashed lines represent hearer responses to the different messages, a_1^* and a_2^*.

We are now in a position to answer our first question regarding the relationship between speaker bias and the use of different forms at equilibrium. Namely, if speakers are too biased when it comes to keeping track of common versus private knowledge then only a single message can be used in equilibrium. For example, we can read off of Figure 4.2 that if $b > \frac{1}{6}$ then only m_2 will be used in equilibrium.[5] When speaker bias is sufficiently

---

$\int_{\frac{2}{3}}^{1} \mathcal{B}(\alpha, \beta)(t)dt \approx p(\text{DIRECTLY ACTIVATED})$, where the probabilities are estimated from the totals in Table 4.1. Obtaining a better empirical estimate of the prior from contemporary data and intuitions is something we leave for future research.

[4]See Appendix A for the full calculations of the solution.

[5]In fact, for any number of messages n there exists a maximum amount of bias b_n such that all messages are used in equilibrium. For any number of messages Crawford and Sobel (1982) show that $b_n > b_{n+1}$. That is, for a given number of messages, there is a maximum amount of bias that allows for all messages to be used in equilibrium. As speakers' bias decreases, $b \to 0$, the number of messages that can be used in
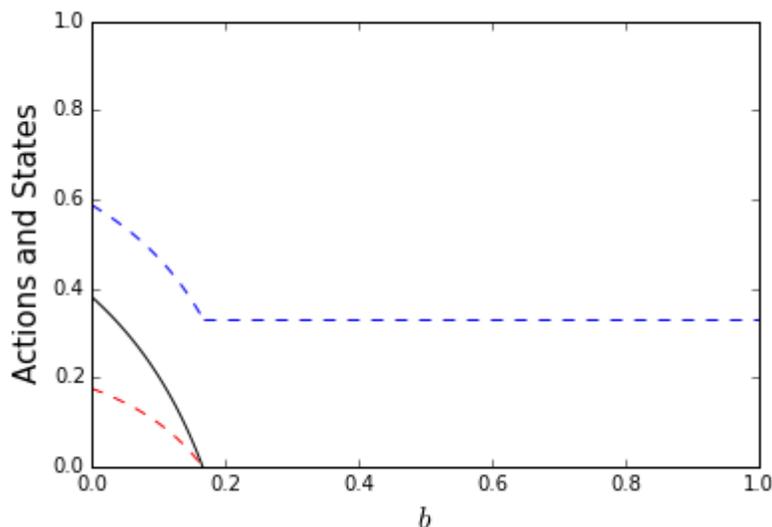
Figure 4.2: Equilibrium solution for two messages for values of bias

large this form carries no information about the activation of the proposition being negated. The *information gain*, or Kullback-Leibler divergence, from receiving the message is zero exactly when the message fails to shift the hearer's beliefs from the prior probability.

$$KL(m_2) = \int_0^1 log\left(\frac{p(t \mid m_2)}{p(t)}\right) p(t \mid m_2)dt \qquad (4.9)$$

If the posterior $p(t \mid m_2)$ is the same as the prior $p(t)$, as is necessarily the case when a single message is used, then the information gained is zero. This follows directly from the definition of information gain, the logarithm of one is zero. This offers a precise definition of bleaching as the loss of information as a signal spreads across states.

There are two important implications for the functional cycle. First, if speakers are sufficiently biased towards their own perspective when estimating the degree of activation for a proposition, then only a single form is stable. In particular, *ne* and *ne...not* cannot coexist, in fact, only *ne...not* can be used in equilibrium. This leads to our second question.

---

equilibrium goes to infinity. The closer the incentives of senders and receivers, the finer and more detailed the information senders want to signal. The number of messages available is the only limit on the amount of information conveyed when the interests of both parties are perfectly aligned.

Where speaker bias is sufficiently large, are equilibria where only a single form is used evolutionarily stable? In fact, we can show that a signaling equilibria is evolutionarily stable only if all available signals are used.

To see why this is the case, suppose that the amount of bias only allows for a single form to be used in equilibrium, call it m_1. Now, suppose further that there is an additional form that is not used in equilibrium, m_2. The action that hearers would take in response to this unused message can vary without affecting the expected utility of speakers or hearers. This means that the equilibrium is not strict, and therefore is not evolutionarily stable. In particular, a single-form equilibrium can be disturbed by the introduction of a new message, it is not *neologism-proof* (Farrell 1993). For example, suppose that a new form is only used for high degrees of activation, and that hearers respond by inferring a high degree of activation. Given speakers' bias, there will be additional states that speakers think warrant using the new form. In response to this increase, hearers will infer a lower degree of activation, meaning additional states will be used by speakers, and so on. This holds for any pair of messages m_i and m_{i+1}.

Importantly, this is just the functional cycle as we have been describing it. So, even though two forms may not be stable for a sufficiently large degree of speaker bias, a single form is never evolutionarily stable. The functional cycle can always be set in motion by the introduction of the appropriately conditioned form. This means that just as *ne* can be pushed out by *ne...not*, so too could *ne...not* be pushed out with the introduction of the right new form, one that is initially restricted to high degrees of activation.

4.5 The dynamics of the signaling game

While we can reason about how speakers and hearers might react to the introduction of a new form, this kind of equilibrium reasoning is essentially static. That is, it allows us to

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 examine how a population evolves from any starting state in general. To understand how speakers and hearers change over time, 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. First, we discuss the replicator dynamics as an appropriate evolutionary game dynamics for studying changes in meaning. Then, we simulate trajectories of a population interacting over time.

The replicator dynamics were originally introduced as an explicitly dynamic model of biological replication, but have since been shown to have deep connections with some of the most widely-studied models of learning. In particular, Börgers and Sarin (1997) prove that the expected behavior of agents playing an asymmetric game while learning according to a *linear reward-inaction* scheme (Bush and Mosteller 1955) is equivalent to the asymmetric replicator dynamics if the agents interact frequently and change their behavior slowly. That is, if individuals tend to do things that are more successful, then their expected behavior can be modeled by the replicator dynamics.

If we assume that speakers and hearers learn in this manner, there are a few conceptual clarifications to be made to justify the use of the replicator dynamics in modeling the functional cycle. First, we need to be sure that speakers and hearers interact frequently and change their behavior slowly. Both of these would seem to follow from the overall frequency of negation. Given that negation is one of the most frequently used forms in any language, it is safe to assume that speakers and hearers interact frequently and do not dramatically alter their use or interpretation of negation from one sentence to the next. Second, we assume that each individual acts as a speaker and hearer, but cannot introspectively reason about the impact of one on the other. That is, individuals cannot use their behavior as speakers to change their own behavior as hearers, nor vice versa. Third, while the repli-

cator dynamics can be used as a model of individual learning, we are interested in how the population as a whole changes. However, given that the *expected* behavior of individuals is equivalent to the replicator dynamics under this kind of learning, then the expected behavior of a population of individuals should be as well. That is, if averaging at the individual level yields the replicator dynamics, then so should averaging over averages. This is akin to treating the populations of speakers and hearers *as if* they were individuals.

We simulate the change in the proportion of the different forms over time for a population evolving according to the replicator dynamics from the same starting conditions.[6] In this case, the population starts off from a state where speakers only use m_2 for high degrees of activation and hearers respond by inferring a high degree of activation. Varying the bias parameter b yields the trajectories shown in Figure 4.3. Along the horizontal axis we have time, and along the vertical axis we have the proportion of m_2 used at any point in time. For sufficiently large amounts of speaker bias, the incoming form is guaranteed to replace the incumbent form, the amount of bias controls the rate at which this happens. For sufficiently small amounts of speaker bias, both forms are guaranteed to persist.

So, our numerical simulations agree with the predictions from the static equilibrium analysis. If speaker bias is too large, then only a single form is used. More specifically, the form that is used at equilibrium is the form that started off restricted to higher degrees of activation. Now, while these simulations yield qualitative information about the dynamics of the functional cycle, we are interested in how this model can be used to understand the details of the functional cycle in the history of English. We now turn to fitting the model to historical trajectory of negation in English.

---

[6]Here we use the discrete-time replicator dynamics for computational tractability, whereas the results presented by Börgers and Sarin (1997) hold for the continuous-time replicator dynamics. While the two dynamics are substantially similar for our purposes, we leave a comparison for future work. Note that we also discretize the set of states and actions for speakers and hearers, respectively. That is, for some n, we treat the set of states $T : \{t_0, ..., t_n\}$ and actions $A : \{a_o, ..., a_n\}$, where $t_i = a_i = \frac{i}{n}$. See Appendix B for full details of the numerical simulations.
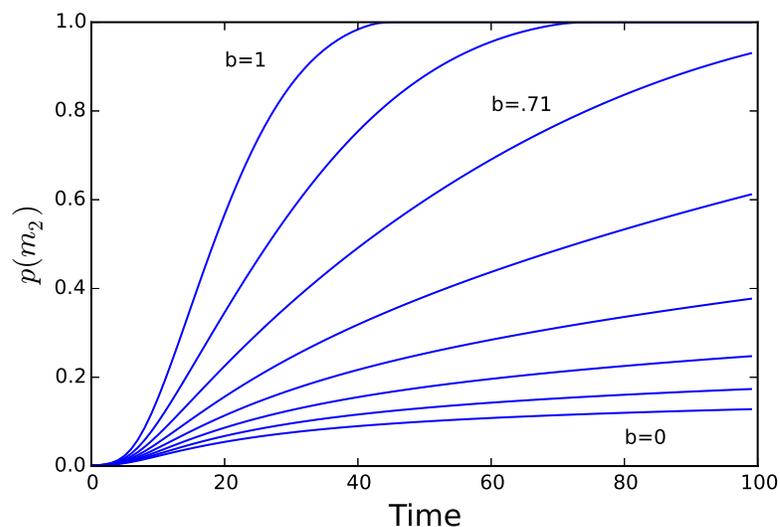
Figure 4.3: Proportion of m_2 under the discrete-time replicator dynamics for varying amounts of speaker bias

4.6 Modeling the functional cycle

Defining the stage game and the evolutionary dynamics allowed 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 used to explain the actual historical trajectory of negation in a concrete case. In particular, we are interested in what happens when we fit the model to data from the history of negation in English. First, we describe the data that we fit the model to. Second, we define the parameters of the dynamics that we fit. Finally, we evaluate these parameters in light of the experimental data presented above.

The data we use are drawn from the PPCME2 (Kroch and Taylor 2000). All tokens used are negative declaratives, excluding cases of contracted negation as well as cases that appear to be constituent negation.[7] Each circle represents tokens in a given year. The size of the circle represents the number of tokens. The height of the circle represents the

[7]We model the data used here after Wallage (2008), which makes a compelling argument for treating contracted negation, among other cases, separately. Many thanks go to Aaron Ecay for sharing the code for generating the queries.
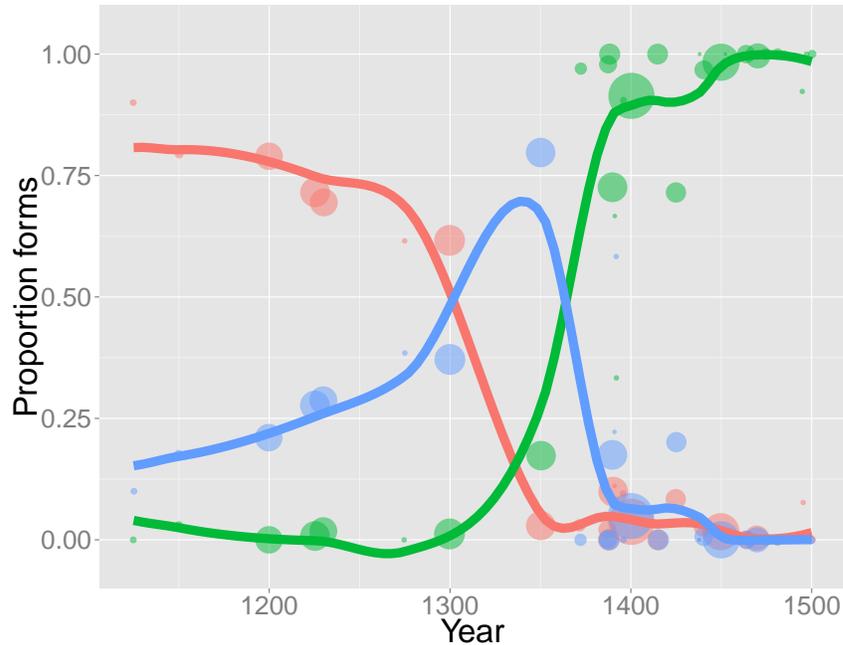
Figure 4.4: Proportion of *ne*, *ne...not*, and *not* in Negative Declaratives

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.

Now, given that we are interested in the functional cycle, we care about the transition from *ne* to *ne...not*. That is, we care about an incoming emphatic form displacing an incumbent form. So, the subsequent rise of *not* is the second transition in the formal cycle, but not a part of the functional cycle. How should we deal with *not* in our analysis? There are two possibilities. First, we could ignore *not* and simply fit the model to the proportions of *ne* and *ne...not*. The problem with doing so is that this attributes too much to small fluctuations in the proportions of *ne* and *ne...not* even if those fluctuations are not meaningful in any sense relevant to the model. It is unlikely that small changes in the 15th century are something that we want to model.

Second, we could ignore the distinction between *ne...not* and *not*, and treat them as if they were the same form. This alleviates the potential problem of attributing too much
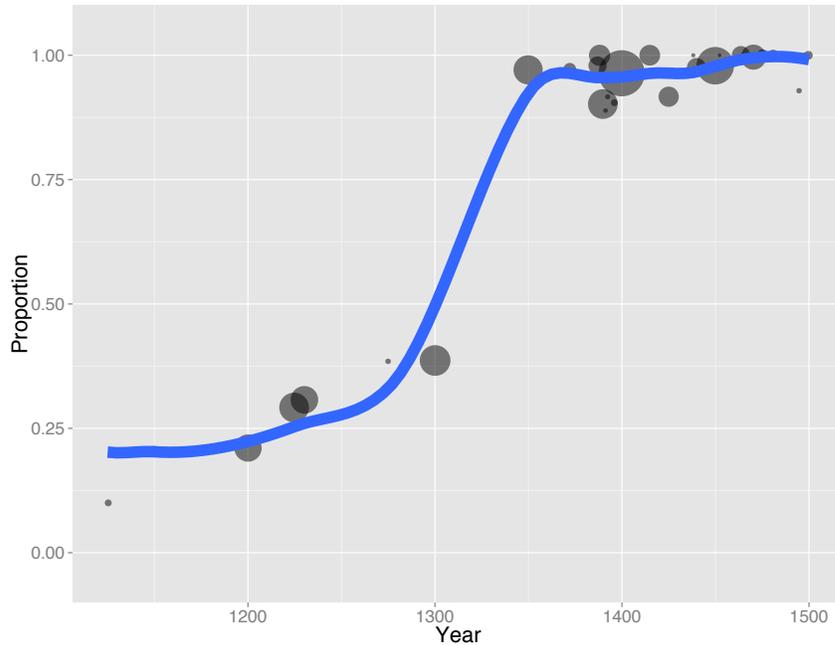
Figure 4.5: Proportion of *ne...not* and *not* versus *ne* over time

meaning to small fluctuations past a certain date. More importantly, it captures the contingency of the second transition of the formal cycle to purely post-verbal negation. That 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. Taking this route allows us apply the same model across languages without regard to subsequent contingent developments. The results of doing so are shown in Figure 4.5.

Taking the trajectory of forms in Figure 4.5 as the data we want to fit our model to, we need to specify the parameters of the model to be fit. In particular, we need to define the initial state of how speakers use the different forms and how hearers respond to them. 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 *ne...not* is fairly infrequent and largely restricted to high degrees of activation. Likewise, we know that hearers' response to *ne...not* is also

largely restricted to actions corresponding to high degrees of activation. We can 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. Namely, *ne* is the default form and does not carry any information above and beyond the prior, it roughly satisfies the following conditional distribution $p(t \mid ne) \sim \mathcal{B}(1,2)$. In contrast, *ne...not* is overwhelmingly used in states with high degrees of activation, that it satisfies the following conditional distribution $p(t \mid ne...not) \sim \mathcal{B}(\alpha, 1)$. The larger α is, the more skewed towards high degrees of activation is *ne...not*. Note that these two distributions along with the prior determine the initial proportion of *ne...not*. So, we only have a single parameter α to fit for the initial state of speakers.

Regarding hearers, we assume that the expected value of the responses to both forms correspond to the expected value of the conditional probability of states given the form. Intuitively, this corresponds to hearers starting off with a fairly accurate responses to the two forms. For *ne* this is satisfied by any distribution $\mathcal{B}(\alpha, \beta)$ such that $\alpha = \frac{1}{2}\beta$, which has an expected value $\frac{\frac{1}{2}\beta}{\frac{1}{2}\beta + \beta} = \frac{1}{3}$. Note that is the same as the expected value of the conditional probability of the state given the message $p(t \mid ne) \sim \mathcal{B}(1,2)$. So we take the conditional probability of actions given *ne* to be $p(a \mid ne) \sim \mathcal{B}(\frac{1}{2}\beta_1, \beta_1)$. All that β_1 does is to determine how concentrated the action is around the expected value. For *ne...not* let $\gamma = E[t \mid ne...not]$, then this is satisfied by any distribution $\mathcal{B}(\alpha, \beta)$ such that $\alpha = \left(\frac{\gamma}{1-\gamma}\right)\beta$, so we take the conditional probability of an action to be $p(a \mid ne...not) \sim \mathcal{B}(\left(\frac{\gamma}{1-\gamma}\right)\beta_2, \beta_2)$. Again β_2 determines how concentrated the action is around the expected value. So, we have two parameters β_1 and β_2 to fit for the initial state of hearers.

The last thing to note before fitting the model is the notion of time. That is, the replicator dynamics specify how populations change from one point in time to the next, but how these abstract units correspond to days or years is unspecified. In what follows we treat each of
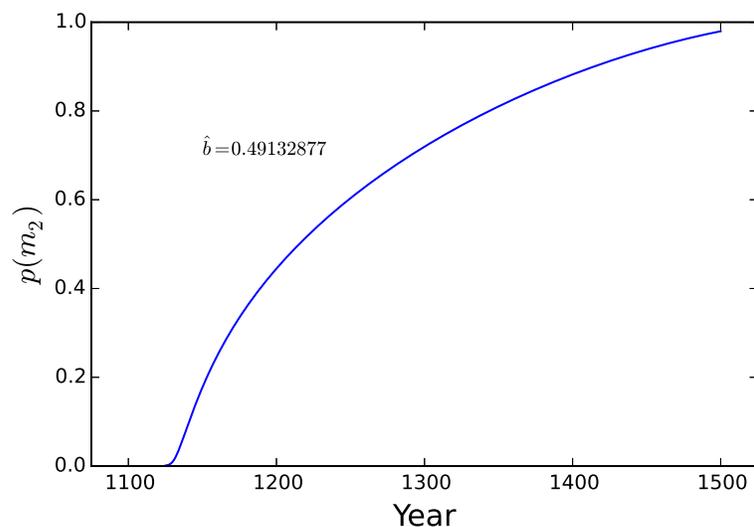
Figure 4.6: Predicted probability of *ne...not* over time for fitted model of functional cycle.

these abstract time units as a year, but leave open the possibility that another proportion may be more appropriate. One option would be to treat the ratio between years and abstract time units as another parameter to be fit in the model, but for now we leave this as an avenue for future research.

We fit the initial state parameters and bias parameter to the data.[8] We begin by visualizing the overall trajectory of the incoming form, then turn to the change in the meaning of the two forms over time. Figure 4.6 shows the predicted proportion of *ne...not* over historical time for the fitted model with the bias parameter $\hat{b} = 0.49132877$. Perhaps more importantly, we can actually inspect the inner workings of the model as they relate to the functional cycle.

In particular, we can examine how the information carried by the two forms changes over time. We gain insight into the functional cycle by considering how the meaning of *ne...not* changes over time as in Figure 4.7. The horizontal axis represents states and vertical axis represents the conditional probability of states given that *ne...not* was used. We

---

[8]See Appendix B for the full details of the starting states and the resulting fit.

96
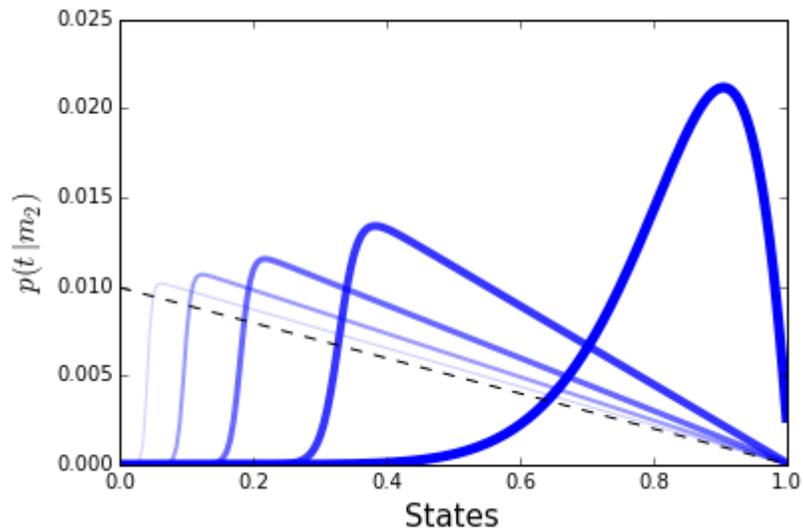
Figure 4.7: The emphatic form over time as given by the conditional probability of states given *ne...not*, where dashed line indicates prior probability distribution.

show this conditional probability at various points as the functional cycle proceeds. The dashed line indicates the prior probability distribution over states. The initial meaning of the incoming emphatic form is represented by the curve with the most rightwards skew. This indicates the point at which the incoming emphatic form carries the most information and is thus the most emphatic. But, as time goes on, *ne...not* spreads to more and more degrees of activation as it the form increases in frequency. We represent this with subsequent distributions that move more and more towards the prior distribution. As they do so, the form loses its emphasis, as indicated by the thickness of the line. When *ne...not* is the only form, it carries no information about activation beyond the prior. Visually speaking, at this point its emphasis has faded entirely.

We also gain insight in to the functional cycle by comparing the relative meaning of both forms. Comparing the meaning of *ne* at the outset of the cycle and *ne...not* at the end of the cycle is particularly informative. Figure 4.8 emphatically demonstrates the dynamics of the push-chain scenario. At the beginning of the cycle in 1125 CE, *ne* carries no information about the degree of activation, it coincides with the prior as indicated by the dashed line. In
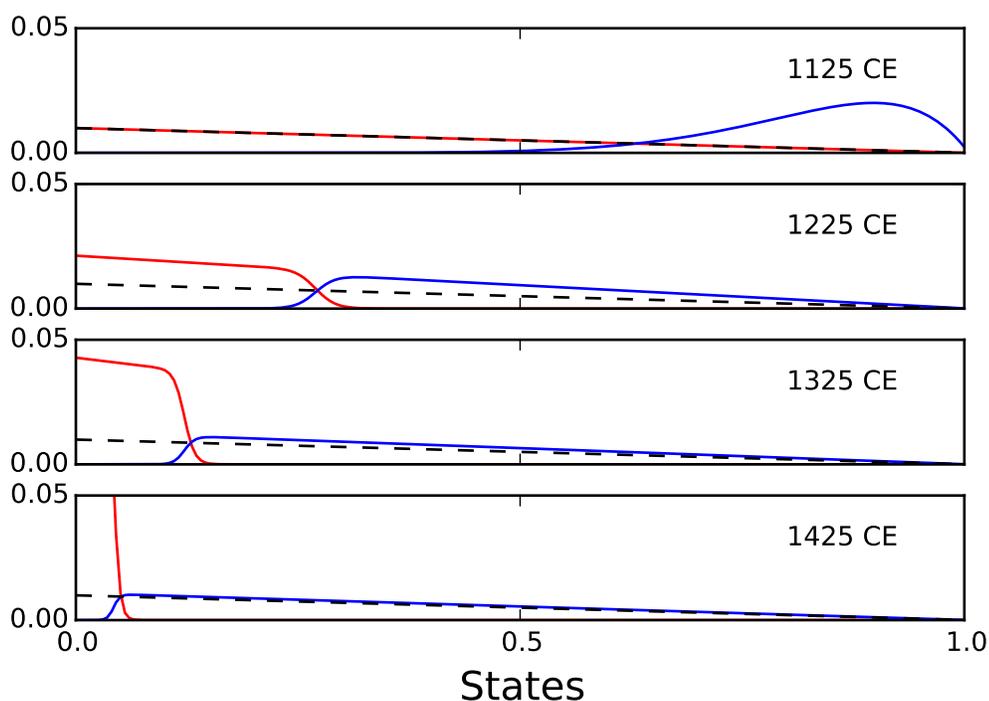
Figure 4.8: The push-chain of the functional cycle given by the conditional probability of states given form at various points in time, where dashed line indicates prior probability.

contrast, *ne...not* is overwhelmingly restricted to cases where the proposition being negated has a high degree of activation. Both of these facts are shown in the top panel of Figure 4.8. But, one hundred years later, *ne...not* has expanded to more states as it increases in frequency and *ne* is pushed to lower and lower degrees of activation. This is shown in the second panel of Figure 4.8. As the functional cycle proceeds, the old form is pushed lower and lower down the scale. Eventually, the incoming form has displaced the incumbent form and ceases to carry any information about the degree of activation.

So, the dynamics of the fitted model match our theoretical conceptions of the functional cycle as a kind of push chain. The incoming form pushes the incumbent form out, eventually taking its place and losing its emphasis. Given that the driving force behind this change was posited to be speakers' bias to overestimate activation, it is important to take a moment to evaluate the value of the fitted bias parameter, $\hat{b}$. Given that the incoming form replaces

the incumbent form, we would expect from our equilibrium analysis that at the very least $\hat{b} > \frac{1}{6}$, but this still leaves a fair amount of room for the parameter to vary. In fact the fitted value $\hat{b} = .4913287$ is well above this minimum.

Now, given this value, we might ask whether it is reasonable in light of the experimental results discussed above. To evaluate the parameter, we return to the results reported in Wu and Keysar (2007). Namely, when playing a communication game, speakers relied on private knowledge of the names of shapes in a proportion of trials. We can use this information to estimate the expected value of the bias parameter exhibited by speakers in the experiment.

To see this, first suppose that there are only two states corresponding to whether or not the name for a particular shape was learned privately or jointly. That is, knowing whether a name was learned privately or jointly is categorical. As Heller et al. (2012) found, this is a reasonable assumption given that participants were incredibly accurate at recalling the context of learning for shapes. Second, suppose that the utility functions for both speakers and hearers are the same form as above. Third, suppose that hearers take one action in response to names and another for descriptions that correspond to initial guesses about the status of the target shape. Then a speaker would only prefer the action taken in response to a name for a privately learned shape if $b > \frac{1}{2}$.

But, this preferences is not categorical. In fact, from the experimental results we only know the probability that $b > \frac{1}{2}$. However, we can estimate the expected value of the bias parameter. Let $p(b) \sim \mathcal{B}(\alpha, \beta)$ be a distribution over the unit interval. We find the parameters such that $\int_{\frac{1}{2}}^{1} p(b)db = p(b > \frac{1}{2})$, which in turn give the expected value of b. For example, where speakers use a privately known name in 5% of trials $E[b] = .1398$, and where speakers use a privately known name in 28% of trials $E[b] = .3549$.

So, there are a range of potential values of speaker bias that we can estimate from the experimental evidence. In both cases these are smaller than the fitted value of the

bias parameter for the functional cycle. However, there are good reasons to treat these experimental estimates as lower bounds. First, the fitted parameters deal with different domains. Where the experiments deal with the referential domain, the functional cycle deals with the propositional domain. It is certainly possible that speaker bias varies across these domains. In fact, we might even expect this. For example, referents often come along with some externally observable entity in the real world, whereas propositions often do not. The fact that propositions are in this sense more abstract may lead speakers to rely on their own perspective more. Second, the experiments found speaker bias even between strangers. These kinds of communicative biases are even more pronounced between people who know each other well (Savitsky et al. 2011).[9] Thus the degree of speaker bias in everyday life may be significantly larger than these experimental estimates suggest.

Careful experimentation will be needed to nail down how private and common knowledge are tracked in the propositional domain, and how this plays out in everyday life. However, the results are largely compatible with both the mechanics of the dynamic model of the functional cycle we have defined here.

Summary

Separating out the formal and functional cycles lightens the explanatory burden. By isolating the functional cycle we were able to identify what conditions the incoming form and reason about why those conditioning factors change over time. In particular, we argued that speakers have difficulty in keeping track of private versus common knowledge, which biases them towards overestimating the activation of propositions being negated. The tools

---

[9]For example, one day when I got home the first thing my wife said to me was, "I DID make an appointment." This struck me as out of the blue, but she said that she told me that she was feeling a little under the weather and debating whether her cold symptoms warranted a trip to the doctor. This conversation had happened several days prior and I had completely forgotten about it, but it was on her mind. In other words, her own subjective estimate of $p =$"I made a doctor's appointment." was greater than the actual degree of activation. In this case, I would say that the bias was fairly high $b \approx 1$.

used to model the functional cycle allow us to offer the first explanatory model of the dynamics of how meaning changes over time. Importantly, they also highlight the fact that while the driving force of the functional cycle is a byproduct of our cognitive limitations in tracking common knowledge, change comes about through the social interactions between individuals in a population. Thus explaining the functional cycle requires a model of how pragmatic competence shapes signaling over time.

Before moving on, we pause to consider two potential lines of research related to the model we have discussed here. The first deals with the alternative definition of emphasis offered at the outset. That is, emphatic negation widens and strengthens negation to preclude exceptions. This interpretation is appealing insofar as negative polarity items are often recruited to create emphatic forms and have exactly this effect (Kadmon and Landman 1993, Eckardt 2006). However, this approach to the functional cycle would have to do two things. First, it would have to specify what serves the role of speaker bias in driving the increase of an incoming emphatic form. Second, it would have to address the problem of over-prediction. That is, if new signals can be formed with the addition of any negative polarity item, then there will always be new forms available, and thus the functional cycle should always be occurring. The fact that we do not observe Jespersen's treadmill means there must be some kind of restriction on what can serve as a new emphatic form. One potential restriction is that the new form must be free from sortal restrictions. For example, both "I didn't move a crumb" and "I didn't eat a crumb" must be equally acceptable.

The other line of research has to do with the implications of this model of the functional cycle for referring expressions that are also sensitive to degrees of activation. Gundel et al. (1993) refer to the scale of sensitivity as the *giveness hierarchy*, which is roughly ordered by pronouns, demonstratives, definites, and indefinites. Pronouns are restricted to referring to entities that are directly activated, whereas indefinites can be used with any entity. Interestingly, similar diachronic patterns are observed as forms spread to lower degrees

of activation. For example, the Modern English definite *the* comes from the Old English demonstrative *se*.[10] However, there are at least two interesting implications of the model of the functional cycle for the stability within the givenness hierarchy. First, we would predict greater stability in these referential terms given the prior distribution over degrees of activation. If propositions are largely skewed towards being non-activated, then referents are largely skewed towards being activated. This change in the distribution largely counteracts any amount of speaker bias. Second, the generation of new pronouns, demonstratives, or definites is arguably a rare event. At least, it would seem rarer than a form of negation becoming associated with activation. We leave exploring both these lines of research for the future.

---

[10]I cannot help but note discussions along this line with Jon Stevens (p.c. March 26, 2010): "One long term goal of this sort of research could be to connect it up with facts about language learning and pragmatics (perhaps using game theoretic tools) to paint a larger picture of why grammaticalization phenomena are so pervasive across languages. As I alluded to in my vignette yesterday, a good model of semantic learning will likely interact with pragmatics in an interesting way; if such modeling techniques become sophisticated enough so as to model the acquisition of grammatical forms as well as content forms, then predictions will be made about the actuation and spread of bleaching, which could serve as a nice test of a model's plausibility."

Part II

The Formal Cycle

Chapter 5

Stability

> It may be urged that change in language is due ultimately to the deviations of individuals
> from the rigid system. But it appears that even here individual variations are ineffective;
> whole groups of speakers must, for some reason unknown to us, coincide in a deviation,
> if it is to result in a linguistic change.
>
> – Leonard Bloomfield (1927:445)

Distinguishing between the formal and functional Jespersen cycles simplifies the task of explanation. It allows us to disentangle two phenomena that overwhelmingly co-occur, and address them separately. For example, in the previous chapter we showed that the functional cycle in English can be explained in terms of the difficulty speakers have in keeping track of private versus common knowledge. Crucially, this explanation of the transition from pre-verbal *ne* to embracing *ne...not* rests on the way our pragmatic competence shapes the use and interpretation of linguistic signals over time.

However, it is important to distinguish between the logical relationship between the two cycles and the explanation of a particular historical change. That is, the functional cycle can occur independently of the formal cycle, so we need an explanation for cases

where it does occur independently. But, in the case of English, and many other languages, the functional cycle coincides with the first transition of the formal cycle. There is no guarantee that the model we described to address the functional cycle is the only or even the best explanation of the observed transitions of the formal cycle. That is, in any given case, pragmatic pressures might not explain the first transition of the formal cycle.

Here we consider other potential explanations for both of the transitions of the formal cycle. In particular, we examine the possible role of acquisition. The facts to be explained are the transition from pre-verbal *ne* to embracing *ne...not* and the transition from embracing *ne...not* to post-verbal *not*. Our goal is to understand whether the process of acquisition offers any insights into why the formal cycle takes place. More broadly, we want to test whether grammatical competence and the process by which it is formed are sufficient to explain the observed changes.

First, we present a model of syntactic acquisition that has several desirable theoretical properties. Second, we determine the dynamics of the model in a population over time. In particular, we show the conditions under which the acquisition dynamics lead to change or stable variation. Third, we outline the syntactic structures at various stages of the formal cycle. These structures allow us to explicitly state the conditions for the acquisition dynamics to lead to either of the transitions of the formal cycle. Finally, we fit the model of the acquisition dynamics to data from the formal cycle in Middle English and discuss the implications of the fitted parameters.

Acquisition can be taken as a cause of the formal cycle if only if the following qualitative and quantitative criteria are met. First, given the acquisition dynamics and grammatical structures underlying the formal cycle, we should predict the qualitative occurrence of both transitions. Second, if the acquisition dynamics predict the transitions, then the quantitative parameters of the dynamics fit to corpus data should be consistent with our theoretical assumptions. That is, the parameters of the model have some falsifiable empirical content

105

that can be tested using corpus data. If neither of these criteria are met, then acquisition cannot explain the transitions of the formal cycle.

In fact, we show that neither of the transitions of the formal cycle observed in the history of English can be explained by acquisition. First, for the grammatical structures posited to underly the formal cycle the acquisition dynamics predict stability rather than change. So, the first criterion cannot be met. Second, this also means that the second criterion cannot be met either; given that both transitions do occur, the parameters of the fitted model necessarily differ. In fact, it would seem that the only way acquisition could lead to either of the transitions would be, in Bloomfield's terms, a mass coincidence of deviation from the current system.

The main contributions of this chapter are twofold. First, we offer a general analysis of the acquisition dynamics, which clearly delineates the conditions for stability and change under certain kinds of parametric variation. Second, we make explicit the conditions for acquisition to play a role in either of the transitions of the formal cycle. In particular, one must show not only that acquisition predicts both transitions qualitatively, but that it also matches the quantitative trajectory of the change. It is important to note that these hold for any set of syntactic structures posited to underly the formal cycle. Demonstrating the role of acquisition requires demonstrating how the acquisition dynamics leads from one state to another.

5.1   A model of acquisition

In the most general sense, the process of language acquisition is some mapping from the initial state of the learner and the linguistic evidence provided to the learner to some terminal state, which is taken to be the grammatical competence of the speaker. We begin by introducing a model that can be used to describe this process. We then note some of the

properties that make it an appealing model of acquisition.

The *variational learning* model of acquisition (Yang 2000, 2002) consists of three basic components. First, there is a finite set of grammars that vary in a parametric fashion, in the sense of the *Principles and Parameters* framework (Chomsky 1981, 1995, Chomsky and Lasnik 1993). Second, a learner keeps track of a probability distribution over grammars, which we can think of as the weights of evidence the learner has for the various grammars. Third, a learner updates her distribution over grammars according to a learning scheme as she receives input from the linguistic environment.

To see this in detail, suppose that a learner is presented with sentences from the linguistic environment. The learner selects a grammar $G_i \in G = \{G_1, ..., G_n\}$ with probability p_i to analyze a sentence. There are two possible outcomes: either the grammar can analyze the sentence or it cannot. That is, the sentence is grammatical with respect to the selected grammar or it is not. The learner updates her distribution over grammars in the following manner, where $0 < \gamma \ll 1$ is a small learning parameter.

$$\text{If } G_i \to s \text{ then } \begin{cases} p'_i = (1 - \gamma)p_i + \gamma \\ p'_j = (1 - \gamma)p_j \qquad \text{for} j \neq i \end{cases} \tag{5.1}$$

$$\text{If } G_i \nrightarrow s \text{ then } \begin{cases} p'_i = (1 - \gamma)p_i \\ p'_j = (1 - \gamma)p_j + \frac{\gamma}{n-1} \quad \text{for} j \neq i \end{cases} \tag{5.2}$$

This is a *linear reward-punishment* scheme (Bush and Mosteller 1955):[1] grammars that are

---

[1]This learning scheme is similar to the linear reward-inaction scheme that yields the replicator dynamics (Börgers and Sarin 1997). It differs in that failures are not ignored, but rather punished. One compelling reason for treating learning differently across the domains of meaning and structure is the hypothesis space of each: the grammatical hypothesis space is heavily constrained, whereas the semantic hypothesis space, even under the Fodorian (1975) conception, is constrained but arguably unbounded. That is, even a finite set of innate concepts can be combined in the appropriate manner without end. Given this quantitative, if not qualitative difference between the domains, it is not clear how a learner would decide what aspects of a given hypothesized meaning to punish (cf. Quine 1960). But, see Smith and Yu (2008), Medina et al. (2011), Smith et al. (2011), Trueswell et al. (2013) for experimental evidence, and Yu and Smith (2007), Frank et al. (2009), Stevens et al. (2013) for computational approaches to the problem of learning meaning.

compatible with a sentence drawn from the linguistic environment are rewarded, whereas grammars that are not compatible with the sentence are punished. Both of these actions are reflected in the first line of the two possible outcomes. If the grammar can analyze the sentence, then its probability is bumped up by some small amount determined by the learning parameter. If the grammar cannot analyze the sentence, then its probability is knocked down by some small amount determined by the learning parameter. The second conditions allow for the weights over grammars to be redistributed while guaranteeing that all probabilities always sum to one.

Now, the probability that a learner attributes to a grammar changes according to how successful that grammar is in dealing with the linguistic environment. In fact, the long term distribution over grammars can be determined from the probability that a grammar will not be able to analyze a sentence and will thus be penalized. For two grammars, G_1 and G_2, let the *penalty probabilities* be c_1 and c_2, respectively. It can be shown that the expected value of the probabilities of the two grammars converge to the following limit values (Narendra and Thathachar 2012:111).

$$
\begin{aligned}
\lim_{t \to \infty} E[p_1(t)] &= \frac{c_2}{c_1 + c_2} \\
\lim_{t \to \infty} E[p_2(t)] &= \frac{c_1}{c_1 + c_2}
\end{aligned}
\tag{5.3}
$$

We should note that this result is about the expected behavior of an individual, rather than the actual behavior of that individual. So, while the expected value of the probabilities converges to these values, the actual probability in the mind of a given learner does not. As we will see below, the actual values in the mind of an individual are close to, but not necessarily equal to these values. This distinction has important implications for the dynamics of acquisition, which we return to in the next section. In particular, it requires us to make certain assumptions about the size of the population.
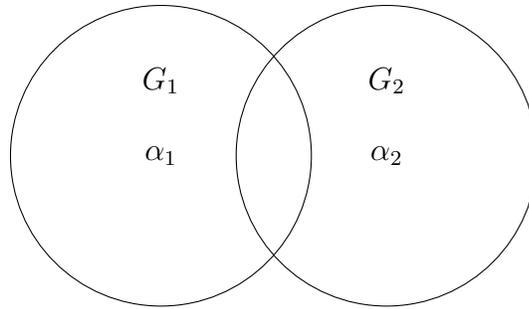
Figure 5.1: Two partially incompatible grammars

So, we know the expected behavior of a learner given the penalty probabilities of the grammars in question. We can determine the penalty probabilities in the following manner. First, suppose that the linguistic environment is composed of the output of two partially incompatible grammars. Let α_1 be the proportion of sentences generated at random by the first grammar G_1 that the second grammar G_2 cannot analyze; likewise let α_2 be the proportion of sentences generated at random by the second grammar G_2 that the first grammar G_1 cannot analyze. Note that while the assumptions underlying them are theoretical both α_1 and α_2 empirical estimates of both can be calculated from a corpus (e.g. Ingason et al. 2013:94-95). We can represent the relationship between the two grammars visually as in Figure 5.1 where the overlap of the two grammars indicate the sentences that are jointly analyzable by both grammars. Second, let the linguistic environment be composed of some distribution over the two grammars, call it $L = p_1 G_1 + p_2 G_2$. From this we can calculate the penalty probabilities as $c_1 = p_2 \alpha_2$ and $c_2 = p_1 \alpha_1$. The likelihood that the first grammar will not be able to analyze a sentence depends on the prevalence of the second grammar in the environment and the likelihood a sentence generated by that second grammar will be incompatible with the first grammar. The same reasoning holds for the second grammar.

We can get a sense for the learning process by simulating individual trajectories. This can be seen in Figure 5.2 where the horizontal axis represents time as additional sentences drawn from the environment and presented to the learner. The vertical axis represents the
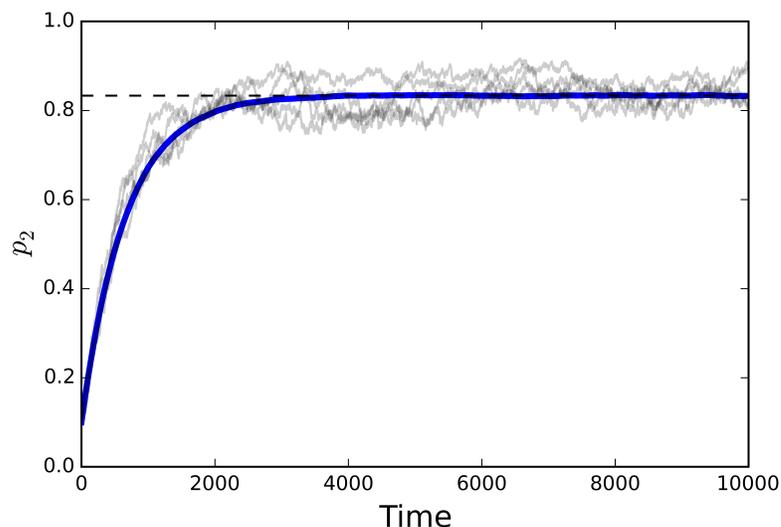
Figure 5.2: Probability of G_2 over time for various individual trajectories and expected motion of trajectories, where $L = \frac{1}{2}G_1 + \frac{1}{2}G_2$, $\alpha_1 = .1, \alpha_2 = .5, \gamma = .1$

weight of G_2 in the learner's mind. We can compare the expected motion averaged across several hundred individual trajectories in bold to several individual trajectories. Where the expected motion smoothly approaches the value predicted by the model, shown by the dotted line, individual trajectories continue to move around the value. Again, we return to this important distinction in the next section when we turn to the dynamics of the model.

However, before doing so, there are several important properties that are evident from the simulations presented in Figure 5.2. First, this learning model allows for the gradual adjustment of learners to the linguistic environment rather than abrupt changes (cf. Gibson and Wexler 1994, Hyams and Wexler 1993). Second, it allows learners to converge to distributions over grammars, rather than a single grammar (Kroch 1989). Third, the expected value of the probability of a grammar is directly proportional to its penalty probability (Narendra and Thathachar 2012:117), which means that learning is reasonably robust. These properties, along with the overall simplicity of the model make it an appealing starting point for investigating how acquisition might lead to change over time.

5.2   The dynamics of acquisition

We showed the properties of the variational learning model at the level of the individual; we now turn to the dynamics of learning in a population over time. First, we show how the expected change from one point in time to the next gives rise to a particular dynamics under particular assumption. Second, we show that the stable rest points of the dynamics are single grammars, with an important exception. We also show how these dynamics are closely related to the logistic model often taken as a proxy for changes in competing grammars.

Under the variational model, for the case of two grammars, we denote the expected value of the weight accorded to the grammar G_2 by a learner to be the following. That is, the average behavior of a learner converges to a probability determined by penalty probabilities and the prevalence of the two grammars in the linguistic environment.

$$p_2' = \frac{p_2 \alpha_2}{p_1 \alpha_1 + p_2 \alpha_2} \tag{5.4}$$

Now, suppose that this distribution in turn serves as the linguistic environment for the next generation of learners. We can determine the expected change in the average weight of p_2 from one generation of learners to the next as the following.

$$\dot{p}_2 = p_2' - p_2 = \frac{p_2 \alpha_2}{p_1 \alpha_1 + p_2 \alpha_2} - p_2 = p_2(1 - p_2) \frac{\alpha_2 - \alpha_1}{p_1 \alpha_1 + p_2 \alpha_2} \tag{5.5}$$

This *mean dynamics* follows the expected motion of the distribution over grammars in the population. Note that we are talking about the change in the expected behavior in the population. This means that we are modeling a fact about the population as a whole, which ultimately derives from individual learning. However, justifying the mean dynamics requires two important assumptions about the population.
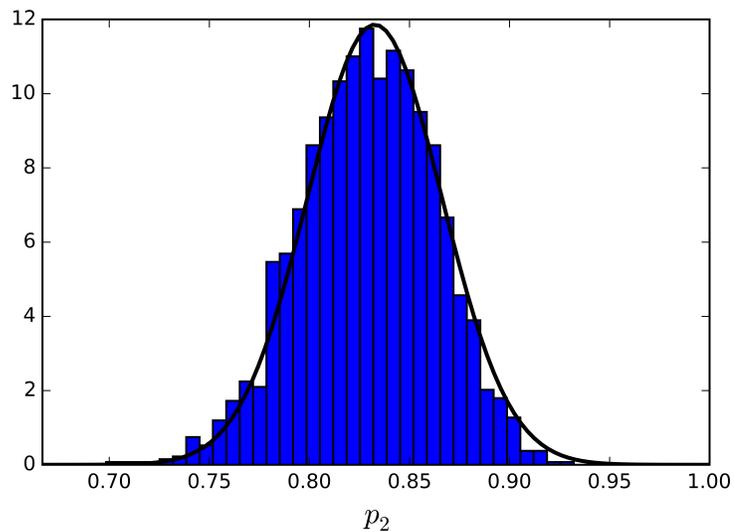
Figure 5.3: Distribution over p_2 for population of learners and fitted normal distribution

The first assumption that must be made is about the size of the population. As we noted in the previous section, learners converge to the limit values only in expectation. In fact, an individual learner is almost always either slightly above or slightly below this expected value, as can be seen in Figure 5.2. But, the distribution over these values in a population of learners at a given point in time is roughly normally-distributed around the limit value, as can be seen in Figure 5.3. In this case, the expected value in the population is close to the limit value. As the population grows, the expected value in the population gets closer and closer to the limit value. In the limit of an infinite population, the linguistic environment for the next generation of learners is the limit value.[2]

The second assumption that must be made is about the structure of the population. Namely, the continuous-time form of the dynamics requires that we assume that there are continuously overlapping generations of learners that contribute to the linguistic environ-

---

[2]This is often a necessary assumption for studying the mean dynamics of what is undoubtedly a stochastic process. See Chapter 10 of Sandholm (2010) for a detailed derivation of the mean dynamics as the limit of a stochastic process in infinite population. In the next chapter we relax the assumption of an infinite population, allowing for stochasticity in the change in the population over time.

ment.[3] Together these assumptions guarantee that the dynamics track the expected weights over the grammars in the minds of learners over time. Thus, the solutions to this *mean dynamics* is a model of the expected behavior in the population.

For the case of two grammars, we can simplify the dynamics in the following manner. Let $s = \frac{\alpha_2 - \alpha_1}{\alpha_2}$ and $p_2 = p$, then we have the following. In population genetics s is referred to as the *selection coefficient* of α_1. For cases where $s > 0$, α_2 has a selective advantage and α_1 is selected against. For $s = 0$, α_1 and α_2 are selectively *neutral*.

$$\dot{p} = p(1-p)\frac{s}{1 - s(1-p)} \tag{5.6}$$

We can read the rest points off the resulting acquisition dynamics. There are two important cases that we will consider. First, if there is only a single grammar, then the weight over grammars is stable. That is, if $p = 0$ or $p = 1$, then the dynamics are, so to speak, at rest. This makes intuitive sense, a grammar cannot be considered if there is no evidence for it. Second, if there is a perfectly balanced amount of independent evidence for both grammars $s = 0$, then any distribution over grammars is a rest point. This also makes intuitive sense, if the balance of evidence is equally in favor of both grammars, then things should not change. We determine the stability of these sets of rest points in turn.

For the first set of rest points, we can evaluate the stability of a single grammar as a function of the selection coefficient, which captures the ratio of independent evidence in favor of one or the other grammar. To do so we evaluate the derivative of the dynamics at the two rest points constituted by a single grammar. If the derivative evaluated at the rest point is negative, then the rest point is *asymptotically stable*. That is, the dynamics will carry the population to the rest point. If the derivative evaluated at the rest point is positive, then the rest point is *unstable*. That is, the dynamics will carry the population away from

---

[3]A discrete-time dynamics might be more appropriate in allowing for a lag between when learners converge on a grammar and contribute to the linguistic environment, but this distinction does not affect the subsequent results.

the rest point. We can express these conditions as a function of the selection coefficient, s.

$$\left.\frac{\partial \dot{p}}{\partial p}\right|_{p=0} = \frac{s}{1-s} \tag{5.7}$$

$$\left.\frac{\partial \dot{p}}{\partial p}\right|_{p=1} = -s \tag{5.8}$$

Note that for $s \neq 0$, only one of these rest points can be asymptotically stable. The rest point $p = 0$ is asymptotically stable if and only if $s < 0$. That is, the population will eventually use grammar G_1 exclusively if only if there is more evidence for it than there is G_2. The rest point $p = 1$ is stable if and only if $s > 0$. That is, the population will eventually use grammar G_2 exclusively if and only if there is more evidence in favor of it.

In fact, these conditions for the stability of both rest points are rather general. If we assume that the selection coefficient is not zero, then there are no other rest points. So, these conditions amount to conditions for *global asymptotic stability*. Importantly, this means that no matter the initial distribution over grammars, if $s > 0$ then grammar G_2 will take over in the population. Thus, the acquisition dynamics are a kind of frequency-independent selection. We can get a sense for this fact by visualizing trajectories from the same low starting state of p for various values of s as in Figure 5.4. No matter how small the initial proportion or slim the margin of evidence, it is guaranteed to eventually go to completion.

Interestingly, the acquisition dynamics in these cases are closely related to the logistic models originally posited by Altmann et al. (1983) and Kroch (1989) to underly competing grammars. While there was no specific learning mechanism underlying the logistic model, it has both connections with the notion of biological competition as well as a straightforward application in terms of logistic regression (cf. Kroch 1989:4). However, it is obvious that the variational model provides some justification for this conception when we com-
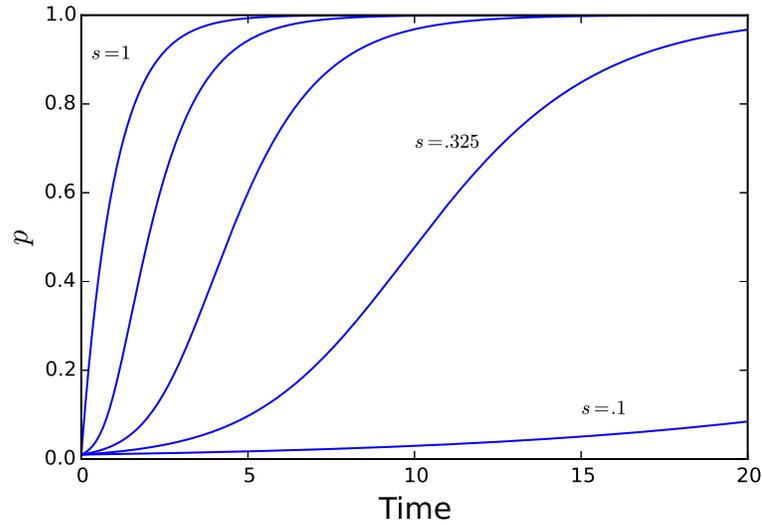
Figure 5.4: Proportion of G_2 over time for various ratios of evidence $s > 0$

pare the dynamics of logistic growth with the acquisition dynamics, where s is taken as the growth rate in the logistic model.

$$\dot{p} = p(1 - p)s \qquad (5.9)$$

In fact, the only difference is that the acquisition dynamics exhibits a varying growth rate as a function of the distribution over grammars in the population. We show the solution for the acquisition dynamics and the logistic model from the same starting point with the same growth rate in Figure 5.5. The solution to the acquisition dynamics predicts a faster initial rate of growth, but slows down to the same rate as the logistic as $p \approx 1$.

In many cases the predictions of the two underlying models may not be distinguishable. But, it is certainly possible that we might detect quantitative evidence for the acquisition dynamics. For example for selection coefficients $s \approx 1$ the acquisition dynamics are asymmetric, unlike the logistic, which is perfectly symmetric. This can be seen in the first few solutions in Figure 5.4. In the context of regression this could potentially lead to quanti-
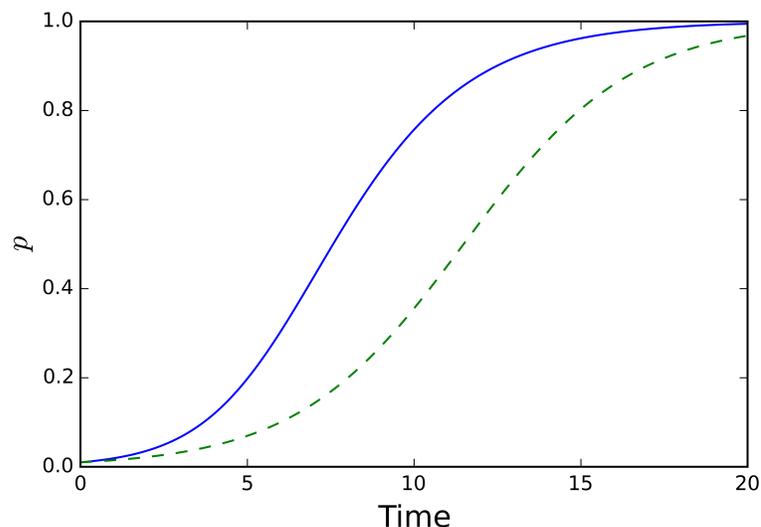
Figure 5.5: Solution to acquisition dynamics and logistic model (dashed) from same starting point with $s = .4$

tative patterns such as heteroscedastic residuals. For example, if we take the acquisition dynamics as a generative model, then fit of the logistic model may systematically under or overestimate the rate of change at different points. We leave investigating this possibility for future research.

For the second set of rest points, where both grammars have equal independent evidence $s = 0$, all distributions over grammars are rest points. All states are *weakly* or *lyupanov stable* in the sense that though the dynamics do not carry the population to a state, the also do not carry the population away from it. Now, we might wonder whether these seemingly knife-edge cases are likely if even possible. However, these cases have a natural interpretation and one that will be particular relevant to the formal cycle. That is, they describe cases where the difference between two grammars hinges on the expression of a single syntactic position. If two grammars correspond to two ways of expressing that position, then the output of each will be incompatible with the other. That is, they will be totally mutually incompatible grammars, rather than only partially incompatible grammars.
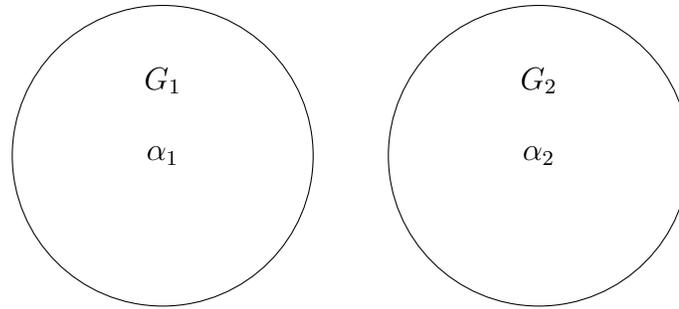
Figure 5.6: Two totally incompatible grammars

This can be visualized as in Figure 5.6, which stands in contrast with Figure 5.1. In other words, where two grammars vary parametrically at the appropriate level, the dynamics predict a kind of weak stability.

So, we determined the acquisition dynamics resulting from the variational model and showed how it exhibits a kind of frequency-independent selection. That is, in most cases only a single grammar is stable. We showed how the resulting dynamics resembles the logistic model of growth. We also noted a crucial exception to this rather robust behavior that leads to a weak kind of stability. With this in mind, we now turn to an analysis of the syntactic structures underlying the formal cycle.

5.3 The syntactic structures of the formal cycle

There are a range of ways of analyzing the syntactic structures underlying the formal cycle.[4] Here we focus on the analysis presented in Frisch (1997), which treats the formal cycle as the result of two independent morphological changes. First, we present the theoretical details of the analysis. We then note corpus evidence in favor of this treatment.

---

[4]For example, different analyses have suggested varying levels of detail in the number and realization of stages, ranging from three stages (Burridge 1983, Bernini and Ramat 1996, Haspelmath 1997, Frisch 1997, Zanuttini 1997, Horn 1989, Hoeksma 1997, Horn 2001, Roberts and Roussou 2003, van der Auwera and Neuckermans 1997, Mazzon 2004, Willis 2005, Lucas 2007, Jäger 2008, Wallage 2008), to four stages (Dahl 1979, Schwegler 1988, 1990, Schwenter 2005, 2006), up to five stages (Honda 2000, Beukema 1999, van der Auwera and Neuckermans 1997, Zeijlstra 2004).

NegP
Spec   Neg′
       Neg⁰   XP
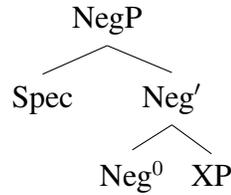
Figure 5.7: The structure of the Negative Phrase

NegP
XP   Neg′
∅   Neg   VP
      ne   ...
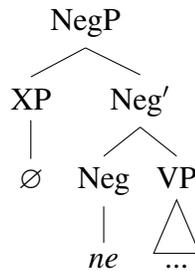
Figure 5.8: Stage one of the formal cycle according to morphological analysis

Frisch (1997) takes Pollock's (1989) analysis of negation as a starting point, assuming that negation constitutes its own phrase, with a fully projected structure like that in Figure 5.7: Neg^0 is the head of the phrase, Spec is its specifier, and XP is a sister phrase such as a verb phrase. In particular, Frisch assumes that this underlying structure is always present (Haegeman 1995). The formal cycle simply consists in changes to how the positions in this underlying structure are expressed.[5]

At the first stage of the formal cycle the negative head is expressed as *ne* whereas the specifier is a phonologically null operator ∅. The syntactic structure according to this morphological analysis can be seen in Figure 5.8. Sentential structure at stage one of the formal cycle in Old English is illustrated in Figure 5.9, where a strikethrough indicates successively upwards head movement. The result is purely pre-verbal negation.

The transition to the second stage in the formal cycle stems from a change in the realization of the specifier of the negative phrase. Namely, the specifier is no longer expressed

[5]This is a more localized version of the *cartographic approach* advocated by Rizzi (1997), which posits a universal syntactic structure. The locus of variation between languages under this conception is how that universal structure is expressed.

```
                    TP
              ┌──────┴──────┐
             DP             T′
              │        ┌─────┴─────┐
             ic   V+Neg+T        NegP
                      │        ┌───┴────┐
                  ne secge     ∅       Neg′
                                     ┌───┴───┐
                                  V̶+̶N̶e̶g̶     VP
                                           ┌─┴─┐
                                           V̶   ...
```
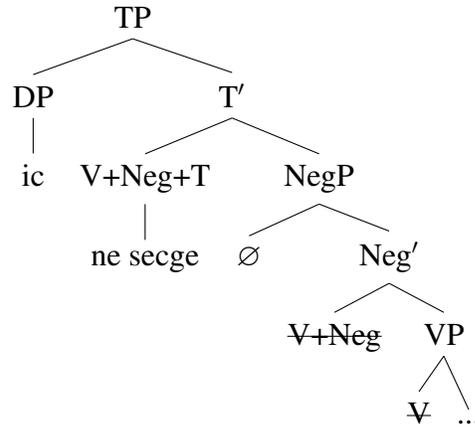
Figure 5.9: Sentential structure at stage one of the formal cycle according to morphological analysis

```
             NegP
          ┌───┴───┐
         XP      Neg′
          │    ┌───┴───┐
         not  Neg     VP
               │     ╱‾╲
              ne    ‾...‾
```
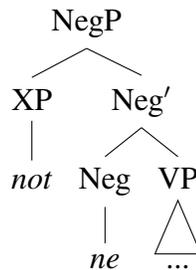
Figure 5.10: Stage two of the formal cycle according to morphological analysis

by a null operator, but instead by *not*, as can be seen in Figure 5.10. Sentential structure at the second stage of the formal cycle in Middle English is illustrated in Figure 5.11, which results in pre- and post-verbal negative elements.

The transition from the second stage to the third and final stage of the formal cycle is simply a a matter of whether the negative head is expressed via lexical content or by some null head ∅. The structure of the negative phrase at the third stage is shown in Figure 5.12, and sentential structure at this final stage in Late Middle English is shown in Figure 5.13.

Similar morphological approaches to the formal cycle are adopted by Roberts and Roussou (2003) and Zeijlstra (2004). The shared aspect of these morphological analyses is the assumption that the underlying syntactic structure remains stable, but the realizations of particular positions within that structure changes. Importantly, since this is the only locus
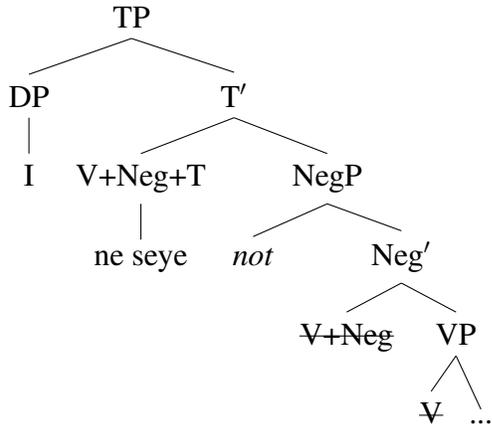
Figure 5.11: Sentential structure at stage two of the formal cycle according to morphological analysis
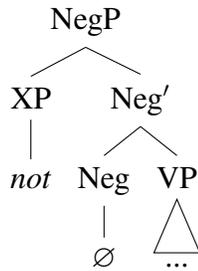
Figure 5.12: Stage three of the formal cycle according to morphological analysis
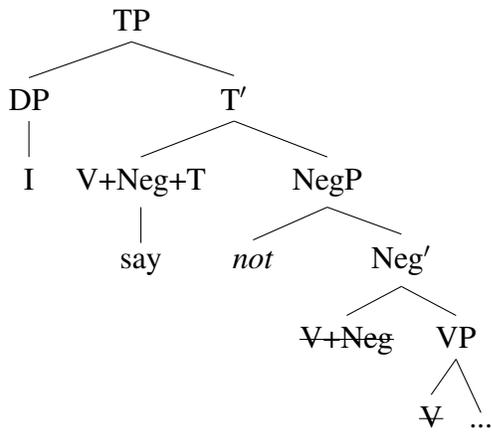
Figure 5.13: Sentential structure at stage three of the formal cycle according to morphological analysis

of change, the transitions that constitute the cycle are independent of each other. That is, the fact that both *ne* and *not* show up in the embracing *ne...not* form is simply the coincidental product of two forms waxing and waning at the same time.

Frisch (1997) tests this theoretical prediction using the trajectory of negation in Middle English in the Helsinki corpus of Middle English. He notes that if the two transitions are independent of each other, then the co-occurrence of *ne* and *not* in the negative phrase should be the product of the probabilities of each occurring.

$$P(ne...not) = P(ne)P(not) \tag{5.10}$$

That is, the probability of *ne...not* is the probability of two independent events *ne* and *not*. This is in fact what Frisch finds.[6] So the formal cycle can be conceived of as two changes in the expression of positions within an underlying structure. In what follows, we will assume that grammatical knowledge is represented as in Figures 5.8, 5.10, and 5.12. That is, under the assumption of a constant underlying structure, a grammar is characterized by the mapping from the syntactic positions in the negative phrase to lexical items. With these definitions in place, we turn to the actual trajectories of the transitions of the formal cycle in English.

5.4   Modeling the formal cycle

Now that we have stated the grammars underlying the stages of the formal cycle we can turn to modeling the transitions of the formal cycle. First, we note that given the structure of the grammars posited to underlie the formal cycle, we can treat each of the transitions separately. Second, we note that given the composition of the grammars we should ex-

---

[6]Calculating these probabilities requires some adjustment for things like instances of adverbial *not* among others. See Frisch (1997:32-47) for the details.

pected stability under the acquisition dynamics. Finally, we fit the acquisition model for the two transitions of the formal cycle and note that the parameters are indeed not what the acquisitions dynamics predict. However, we note what a successful explanation of the formal cycle would have to do.

The grammars underlying the formal cycle differ only in the expression of two syntactic positions. This has two important implications. First, if the two changes in how these positions are expressed are independent of each other, then we can treat them as such. That is, we can treat the two transitions of the formal cycle as independent events of competition between two grammars for expressing those positions. In what follows we take this approach, treating both the first and second transitions as cases of the acquisition dynamics with two grammars. Second, if the grammars involved in each transition differ only in the expression of one position, then they are totally mutually incompatible $s = 0$. This means that the acquisition dynamics predict stability in the case of both transitions, but this is certainly not what we observe. In fact, this alone suffices as a demonstration that acquisition as we have modeled it here cannot account for either of the transitions of the formal cycle. That is, given the description of the grammars underlying the stages of the formal cycle, acquisition cannot cause the transitions between them. This means that the qualitative criterion for acquisition serving as a cause of the formal cycle is not met.

However, it is useful to note what would have to be the case for acquisition to explain the empirical trajectories of the two transitions. That is, if we fit the acquisition dynamics to data, the fitted parameters tell us what we would need to find in order to take acquisition as the cause of the formal cycle. We fit the acquisition dynamics to the trajectory of the first transition modeled as the competition of two grammars for the specifier of the negative phrase. In this case we take G_1 as $G_\varnothing$ and G_2 as G_{not} to be grammars that determine how the specifier of the negative phrase is expressed. We take instances of *ne* to be compatible with $G_\varnothing$ and instances of *ne...not* and *not* to be compatible with G_{not}. The parameters to fit

are the initial state of the second grammar in the population and the selection coefficient s that captures the ratio of evidence in favor of the second grammar.

The last thing that needs to be specified is the notion of time. The solution to the acquisition dynamics is in abstract time units, but how these correspond to the actual time in days, months, or years is unspecified. We could fit this relationship as another parameter in the model, but this would be problematic if we were to find different values for the second transition. Instead, we stipulate a ratio between the units of the dynamics and years where one abstract time unit corresponds to five years. We take this to be a rough approximation of the time between when a learner is born an starts contributing to the linguistic environment, but leave it to further research to gain a better estimate the actual value.

The results of fitting the acquisition dynamics can be seen in Figure 5.14.[7] It is important to note, however, that regardless of how well the fitted model approximates the actual trajectory of the first transition, the parameters are not possible. That is, the second grammar cannot have an advantage over the first given that they differ only in the expression of a single syntactic position, yet this is exactly what would be required for acquisition to explain the first transition. This means that the quantitative criterion for acquisition serving as a cause of the formal cycle is not met. However, for another description of the grammars underlying the stages of the formal cycle, this parameter would need to match our theoretical predictions. That is, if the grammars were specified differently, the selection coefficient $\hat{s}$ would still have empirical content. We should be able to look at a corpus and use the grammatical descriptions to see if it is consistent.

We also fit the acquisition dynamics to the trajectory of the second transition modeled as the competition of two grammars for the head of the negative phrase. In this case we treat G_1 as G_{ne} and G_2 as $G_\varnothing$. We take instances of *ne* and *ne...not* to be compatible with G_{ne} and instances of *not* to be compatible with $G_\varnothing$. The results of fitting the acquisition

---

[7]See Appendix C for details.
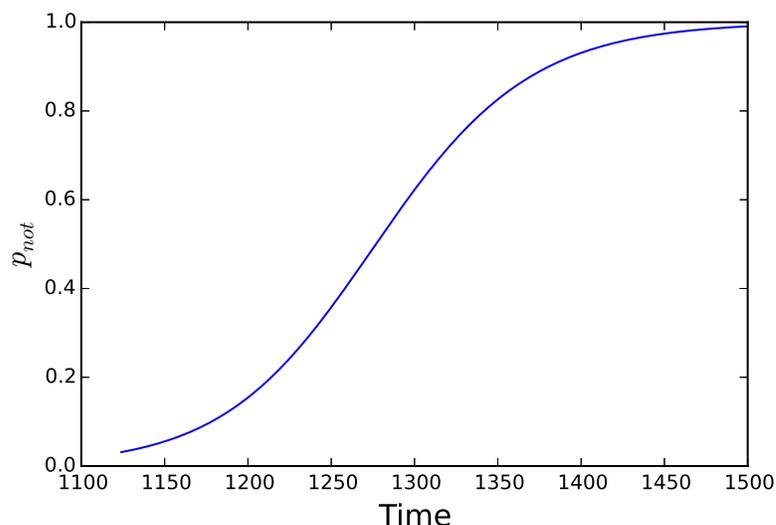
Figure 5.14: Proportion of G_{not} over time for the first transition of the formal cycle, $\hat{s} = 0.10291529$

dynamics can be seen in Figure 5.15.[8] Again, this means that our second criterion for acquisition serving as a cause of the formal cycle is not met. So, neither of the criteria for acquisition serving as a cause of the formal cycle have been met. That is, the acquisition dynamics do not predict either of the transitions, nor do the parameters of the fitted models agree with the corpus evidence predicted by the grammars.

It bears emphasis that these criteria are not specific to the acquisition dynamics we specified nor to the grammatical structures posited to underly the stages of the formal cycle. We could just as well adopt another model of acquisition or the grammatical description of the formal cycle (cf. Niyogi 2006). But, abandoning the appealing theoretical properties of the variational model and its acquisition dynamics seems a bit hasty. This is especially true given that we need some model to provide any explanation at all. There are, however, a wealth of options when it comes to grammatical descriptions of the formal cycle, as we noted above.

---

[8]We only fit the dynamics to data from the point where there are instances of *not* in all subsequent years, from 1300 CE onwards. Again, see Appendix C for details.
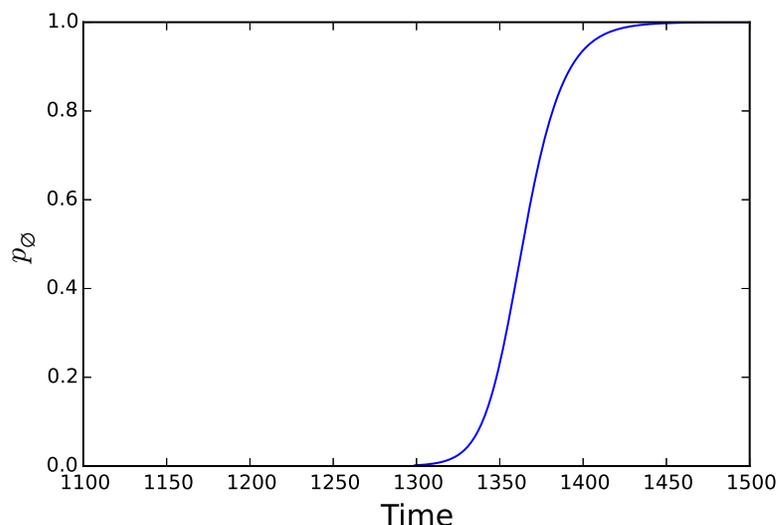
Figure 5.15: Proportion of $G_\varnothing$ over time for the first transition of the formal cycle, $\hat{s} = 0.34128455$

For example, Wallage (2008) makes a compelling corpus-driven argument for the treatment of the formal cycle as two interdependent morphosyntactic changes. In particular, Wallage treats the first transition as the addition of the post-verbal *not* as well as the change in the formal features of pre-verbal *ne* from an interpretable to an uninterpretable feature (Chomsky 1995). This more articulated approach would likely face the same problem regarding the first transition, but might offer insight into the second transition. But, it would also offer an interesting alternative insofar as it takes the locus of variation to be the properties and features of functional categories, according to the so-called *Chomsky-Borer conjecture*, (Baker 2008).

But, regardless, for acquisition to explain the formal cycle, both of the criteria we described above have to be met. Not only must both of the transitions be predicted, they must also be modeled in an empirically and theoretically consistent manner. To perhaps belabor the point, we can use the parameters of the fitted models of the transitions to predict the proportion of the different forms of the formal cycle in English over time. The result can
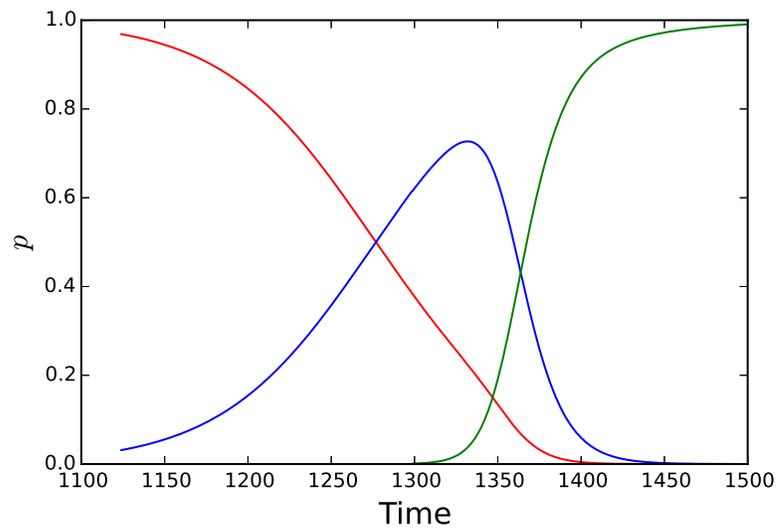
Figure 5.16: Proportion of *ne*, *ne...not*, and *not* predicted by the fitted parameters of the acquisition dynamics

be seen in Figure 5.16, and indeed the predicted forms are a close match to the empirical trajectories that we observe. It is tempting to take this as a reasonably good result. But, the parameter values that generate this result are on their face not compatible with the grammatical structures posited to underly the formal cycle. If we want to explain, rather than just describe historical changes we need models that get the picture right while simultaneously being self-consistent. That is, we not only need to be able to fit parameters, but also to make sure those parameters make sense given our theoretical assumptions about the grammatical knowledge that speakers acquire.

Summary

In this chapter we presented a model of syntactic acquisition, determined its predicted dynamics in a population over time, and fitted it to data from the formal cycle in Middle English. We found that neither the qualitative nor quantitative criteria for taking acquisition

as the cause of the formal cycle were met. All together then, it seems that acquisition cannot be taken as a cause of the formal cycle. If this is indeed the case, then it has important consequences for our understanding of the two transitions of the formal cycle.

Regarding the first transition from *ne* to *ne...not*, if acquisition cannot explain it, then use can. That is, given that this transition coincides with the functional cycle, then the explanation of the functional cycle put forward in the previous chapter is the only and necessarily the best explanation of the observed transition. Alternative analyses of the grammars underlying the formal cycle may change this, they must be both qualitatively and quantitatively accurate and consistent.

Regarding the second transition from *ne...not* to *not*, neither acquisition nor use can explain it. The transition does not coincide with another functional cycle, *not* is not restricted to specific contexts. This leaves us in the strange position of observing a change without an obvious cause. Absent some mass coincidence, what are we to make of the second transition of the formal cycle? One possibility is that this second transition is not the result of one mass coincidence, but rather the accumulation of many much smaller coincidences.

To see how this might be the case, consider the fact that the acquisition dynamics only predict a weak form of stability in the expected change of the expected behavior of learners in a population. For example, if we relaxed the assumption regarding the size of the population, then the linguistic environment provided to learners would differ slightly from the limit value. For the second transition, suppose that the proportion of $G_\varnothing$ in the actual linguistic environment is slightly higher than expected due to sampling errors. Now, suppose that it is slightly higher in the next generation as well due to sampling errors. If enough of these small coincidences compound over time, one grammar may replace another without ever having more evidence in favor of it.

Indeed, this possibility has been extensively studied in population genetics in terms of *genetic drift*. That is, when the selection coefficient is zero $s = 0$, as is the case in the

127

second transition of the formal cycle, change can come about due to random sampling. Or, in this case, random changes in the probabilities over grammars learned over time. This means that we the second transition might be the result of a series of small coincidences rather than a single improbable one. In the next chapter we turn to means of testing this possibility.

Chapter 6

Chance

> [T]o my imagination it is far more satisfactory to look at such instincts as...consequences
> of one general law leading to the advancement of all organic beings – namely, multiply,
> vary, let the strongest live and the weakest die.
>
> –Charles Darwin (1859)

In our analyses of historical change we have used mean dynamics that presuppose effectively infinite populations. This assumption allows us to smooth out chance occurrences and study the expected motion of what is arguably a stochastic process. However, as we noted in the previous chapter, this does not rule out the possibility that the changes we observe are actually due to chance. Here we relax the assumption of effectively infinite populations and determine its effect on the acquisition dynamics. While the acquisition dynamics predict stability, it is only a weak kind of stability stemming from the size of the population. Either of the transitions of the formal cycle could be the result of random sampling errors in a finite rather than an infinite population. This is particularly relevant for the second transition from *ne...not* to *not*, which we noted cannot be explained by either use or acquisition. In this chapter we investigate the possibility that the transitions of the

formal cycle are indeed due to chance.

First, we briefly discuss the conceptual role of drift in the history of population genetics. Whereas early approaches to population genetics from Darwin on have emphasized selection over drift, more recent work has developed tools for addressing both theoretical possibilities. Second, we demonstrate the dynamics of drift in finite populations. We show that change can indeed come about through drift. More importantly, we show that drift can lead to change that exhibits the qualitative trajectories so frequently observed in historical linguistics, and selection can lead to change that is decidedly not like what we observe historically. Third, we introduce a statistical method for testing whether we can reject the hypothesis of drift in favor of selection for a particular trajectory. We apply this test to both of the transitions of the formal cycle and show that while we can reject the possibility of drift in the transition from *ne* to *ne...not*, we cannot reject it in the transition from *ne...not* to *not*. Finally, we discuss these results in light of the previous chapters. Given that we can reject the role of drift in the first transition, this adds support for our model of the functional cycle. Given that we cannot reject the role of drift in the second transition, we can make sense of the varying times to completion of the transition across languages.

The main contributions of this chapter are twofold. First, we offer the first application of statistical methods for distinguishing between selection and drift in linguistic time series. This has important implications not just for the formal and functional cycles, but for linguistic change as a whole. The crucial fact is that the typical trajectories observed in linguistic change may arise from selection or drift. Simply put, we cannot distinguish between these two possibilities from simple visual examination. Rather, we need a means of testing competing hypothesis about the underlying cause of the change. Second, the application of these methods offers insight into the formal cycle, and the second transition in particular given that neither use nor acquisition offer an explanation. More broadly, it also offers constraints on the nature of stable variation.

6.1 Genetic drift

The nineteenth century saw both the discovery of the mechanics of genetic inheritance by Gregor Mendel and the formulation of the principle of natural selection by Charles Darwin. Yet, it was not until the early twentieth century that these two notions were reconciled and put on a rigorous mathematical foundation by the work of Fisher, Sewall Wright, and Haldane in what has become known as the modern *evolutionary synthesis* (Huxley 1942). As the term suggests, these foundations offer a coherent and compelling view of observed changes in biological populations.

The role of random drift in explaining these changes, however, was taken to be minimal in comparison to selection. The balance between selection and drift was revisited most notably by Kimura (1968) in his *neutral theory* of evolution, which emphasized the role of drift rather than selection at the molecular level. While the proper emphasis on of each has been the subject of intense debate, more recent work in population genetics has taken the balance between these two forces as an empirical matter, developing theoretical tools for distinguishing the potential role of each. In what follows, we adopt this approach to understanding changes in a population over time.

In particular, our goal will be to test the role of drift in both of the transitions of the formal cycle. Our first step towards this approach is to demonstrate the need for statistical methods for distinguishing the quantitative signatures of selection versus drift in linguistic time series. We do so by simulating the dynamics of change in a finite population to demonstrate some counter-intuitive possibilities regarding drift versus selection.

6.2 The dynamics of drift

Here we outline the dynamics of a simple model of change in a finite population. Using simulations of the *Moran model* (Moran 1958), we show that our intuitions about how drift

and selection look are not as useful as we might think. That is, we cannot simply look at the trajectory of a change and determine intuitively if it occurred due to selection or drift. This is true despite the fact that we often observe a similar qualitative *S*-shaped trajectory in the course of language change (Bailey 1973).

The *Moran model* describes the dynamics of selection and drift in a finite population. At each point in time one individual is chosen to reproduce and one individual is chosen to die, yielding a continuous-time Markov chain.[1] The probability that a variant is chosen to reproduce is proportional to its relative fitness. For example, consider a population composed of a particular number of two variants, $N = n_1 + n_2$. Let the selective advantage of the second variant be s, then the probabilities of each being selected to reproduce are given by the following.

$$p_1^{birth} = \frac{n_1}{n_1 + n_2(1 + s)} \tag{6.1}$$

$$p_2^{birth} = \frac{n_2(1 + s)}{n_1 + n_2(1 + s)} \tag{6.2}$$

For the case where the two variants are selectively neutral, $s = 0$, the two variants are chosen directly proportional to their respective frequencies in the population. Where $s > 0$, the second variant is selected for, so the second variant is slightly more likely to be chosen. This makes sense, if one variant has a selective advantage, then we would expect it to be more likely to reproduce.

While the probability of being chosen to reproduce is proportion to the relative fitness of the two variants, the probability of being chosen to die is directly proportional to just the

---

[1]While the Moran model assumes continuously overlapping generations as we did in the previous chapter regarding syntactic acquisition. the *Wright-Fisher model* can be taken as the discrete-time analogue where generations do not overlap and are sampled all at once. The same general point holds for the Wright-Fisher process as well though: we cannot determine the selective advantage of particular variants from the trajectory they create.

prevalence of the two variants in the population.

$$p_1^{death} = \frac{n_1}{n_1 + n_2} \tag{6.3}$$

$$p_2^{death} = \frac{n_2}{n_1 + n_2} \tag{6.4}$$

Given that a single individual is chosen for birth and death at each point in time, then the number of any variant either increases, decreases, or stays the same. For example, we can keep track of the number of the second variant in the population., n_2. This number will go up by one if an individual of the second variant is chosen to reproduce and an individual of the first variant is chosen to die. If the selections are the opposite, where an individual of the first variant is chosen to reproduce and an individual of the second variant is chosen to die, then the number of the second variant will decrease by one in the population. If the type of both individuals selected is the same, then there will not be any change in the number of either variants. The probability of all these outcomes is determined by the probability of selection for birth and death that we listed above.

In fact, we can specify the probability of each of these changes from one point in time to the next. In particular, we can show the probability of a particular change in the number of the second variant in the population. We list the probability that n_2 will increase, decrease, or stay the same.

$$p_{n_2, n_2+1} = p_2^{birth} p_1^{death} = \frac{n_2(1 + s)}{n_1 + n_2(1 + s)} \frac{n_1}{n_1 + n_2} \tag{6.5}$$

$$p_{n_2, n_2-1} = p_1^{birth} p_2^{death} = \frac{n_1}{n_1 + n_2(1 + s)} \frac{n_2}{n_1 + n_2} \tag{6.6}$$

$$p_{n_2,n_2} = 1 - p_{n_2,n_2+1} - p_{n_2,n_2-1} \qquad (6.7)$$

There are two important points where the population does not change at any subsequent points in time. Namely, if $n_2 = 0$ or $n_2 = N$, then the population will not change at any point moving forward. These are referred to as the *absorbing states* of the process, whereas all other states are *transient*. This follows from the fact that at the absorbing states the sampling probabilities for the two variants are either one or zero.

So, neutral selection in a finite population provides an interesting contrast to the acquisition dynamics we presented in the previous chapter. Indeed, if we loosen the assumption of an infinite population, then the acquisition dynamics makes a similarly strong prediction of no stable variation between grammars that are totally mutually incompatible. In other words, if acquisition is the only force acting on a language, then a single syntactic position will only ever be expressed by a single variant. We return to the implications of this prediction for observed stable variation below.

Now that we have specified the dynamics of the model, we can simulate trajectories of a population of size N for selection coefficient s. Before doing so though, it is useful to consider what our expectations would be about a population evolving under neutral versus positive selection. Intuitively, we would expect selection to exhibit a clear trajectory, where the variant that is being selected against is quickly driven from the population. In contrast, we would expect neutral selection to be a more random kind of change, not necessarily tending in direction or the other. That is, we would not expect any clear trajectory favoring one variant over the other. It is useful to dwell on these expectations a bit before considering the simulation results. We want a clear baseline to compare the results to.

When we do compare individual trajectories of the model for cases where $s = 0$ versus $s > 0$, we sometimes have the exact opposite results, as can be seen in Figures 6.1 and 6.2.
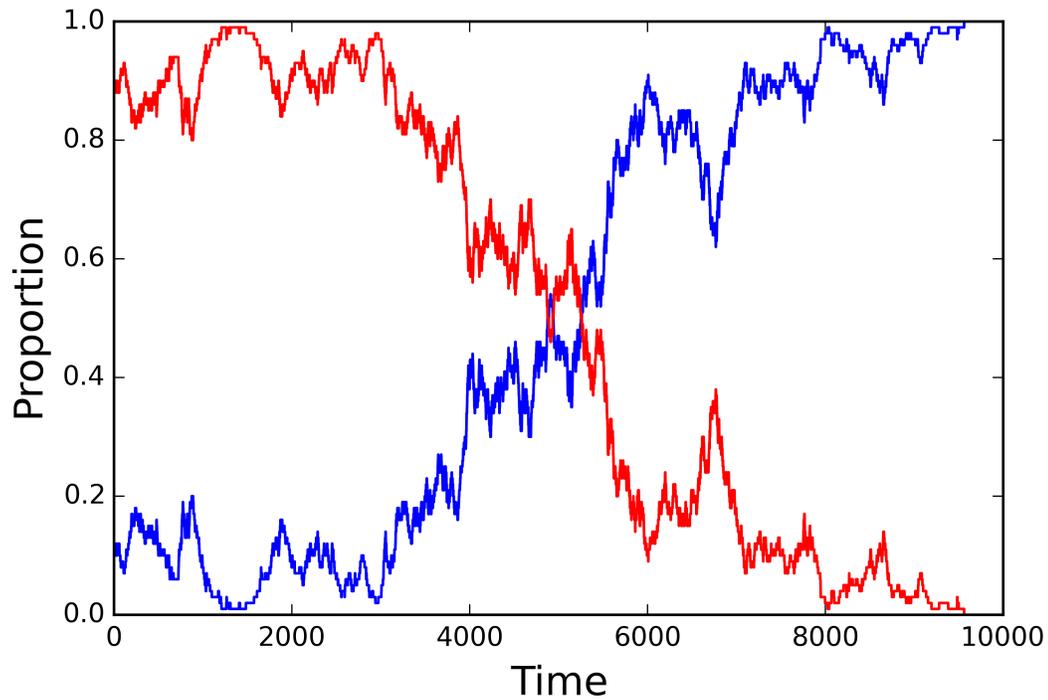
Figure 6.1: Proportion of forms over time in simulation of Moran model for $N = 100$ and $s = 0$

That is, in Figure 6.1 we show the trajectory of a population under neutral selection $s = 0$ that exhibits the characteristic S-shaped curve observed in language change. Likewise, in Figure 6.2 we show the trajectory of a population under positive selection $s = .1$ that exhibits a rather different kind of growth. The fundamental fact is that the underlying cause of a particular change cannot be read off its form, any particular S-shaped curve may be due to drift or selection.

Now, one objection to this point might be that the individual trajectories presented in Figures 6.1 and 6.2 are atypical. More often than not, selection and drift will match our expectations. This is certainly true, these trajectories were chosen particularly to emphasize the potentially counterintuitive result of selection and drift in finite populations. So, we might be tempted to conclude that the probability that a given change is due to drift given
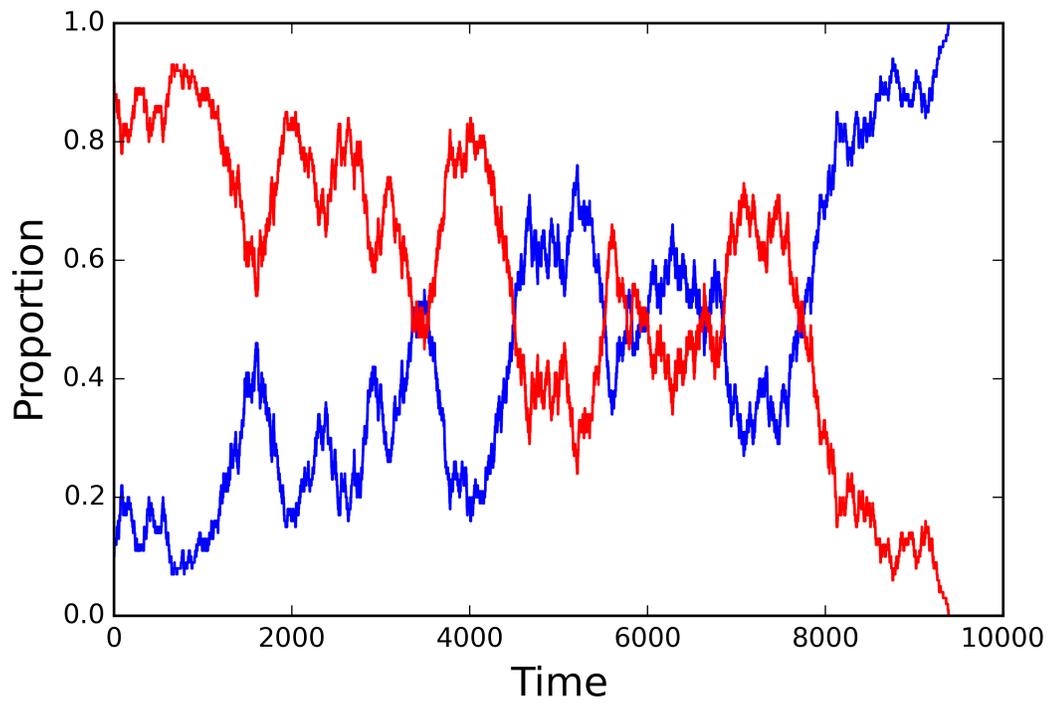
Figure 6.2: Proportion of forms over time in simulation of Moran model for $N = 100$ and $s = .1$

that it is *S*-shaped is actually fairly small.

However, this objection rests on several assumptions. First, it assumes that we have some prior expectation over drift versus selection. But, we do not have sufficiently developed causal models of linguistic change that would warrant any particular prior. In fact, in particular cases, we have very strong theoretical reasons for expecting drift, as was shown in the previous chapter regarding the transitions of the formal cycle. Second, the size of the population is not a parameter we know beforehand. That is, we can never be sure from the data itself what the probability of a particular curve should be under drift. Third, and perhaps most importantly, it is not clear what actually counts as being *S*-shaped or not. While we may have intuitions about what does or does not count, these would need to be clarified.

If all of these assumptions cannot be clearly articulated and justified, then it is reasonable to assume that we simply cannot know the cause of a particular change given its shape. If this is indeed the case, then we need some means of testing different hypotheses about the underlying causes of change. For the formal cycle, we want a means of testing for whether or not each of the transitions is due to drift or selection.

6.3   Modeling the formal cycle

If drift and selection can yield counter-intuitive trajectories in finite populations, then we need some means of distinguishing the two in linguistic time series. In what follows, we discuss the *fitness increment test* described by Feder et al. (2014) as a means for testing the hypotheses of drift versus selection. First, we begin by noting the motivation for the test, as well as some of details regarding its application. Second, we apply the test to the two transitions of the formal cycle.

The fundamental comparison to be made in distinguishing between drift and selection

is between two models. The first model assumes that there is no selection $s = 0$ and determines the population size N that would best explain the data. The second model assumes that there is selection $s > 0$ determines the selection coefficient and population size N that would best explain the data. Given that the first model can be taken as a special case of the second, the two can be compared using a likelihood ratio test. In this case, drift is our null hypothesis and selection is our alternative hypothesis.

However, Feder et al. (2014) note that doing so is not without complications. First, even using standard approximations to the Moran process (Kimura 1955a,b, Ewens 2012), calculating the likelihoods necessary to compare the two models is computationally intensive. Second, even if these values were simple to obtain, the relevant test statistic is not χ^2 distributed as is often assumed (cf. Wilks 1938). In fact, using the χ^2 distribution systematically underestimates the false positive rate, meaning that if we assume the test-statistic is χ^2 distributed, then we will reject the null hypothesis of drift more often than we would like to when it is true.

To address these problems Feder et al. (2014) propose an approximation to the Moran process that consists of the combination of a deterministic logistic process and a Gaussian noise process.[2] Under this approximation, the changes in the number of different forms over time have certain properties. Suppose we have measurements from several points in time of the two variants. Let p_i be the proportion of the second variant in the population at time t_i. We are interested in how these proportions change over time, so will look at the differences in proportion between different points of time, $p_i - p_{i-1}$. In particular, we rescale these *fitness increments* in the following manner.

$$Y_i = \frac{p_i - p_{i-1}}{\sqrt{2p_{i-1}(1 - p_{i-1})(t_i - t_{i-1})}} \tag{6.8}$$

[2]See Feder et al. (2014:521-522) for the mathematical details. Note that while this approximation technically only holds for the Moran process, in practice it works well for the Wright-Fisher process as well.

For the Gaussian approximation, under the null hypothesis of drift these scaled fitness increments are independent and approximately normally distributed around zero with a variance inversely proportional to the size of the population. In contrast, under the alternative hypothesis the increments are independent and approximately normally distributed, but with a non-zero mean and a different variance. For our purposes, we want to know whether the mean of the scaled fitness increments is greater than zero. That is, we want to know if the incoming variant has some selective advantage. This *fitness increment test* can be accomplished using a one-tailed *t-test*. So, under the Gaussian approximation, all we have to do to do is rescale the fitness increments and test if their mean is greater than zero.

Returning to the formal cycle, we are interested in determining whether we have sufficient evidence in favor of selection to reject the null hypothesis of drift in either of the two transitions. That is, we want to know not just whether a model with additional parameters fits the data better, but if it fits the data sufficiently better for us to reject the null hypothesis of drift. In the case of the formal cycle, we want to know whether we can reject the role of drift in the transitions from from *ne* to *ne...not* and from *ne...not* to *not*.

The overall trajectory of these transitions in Middle English is shown in Figure 6.3. However, as we noted in the previous chapter, we are really interested in two independent changes at different locations in the negative phrase. In what follows, we will treat the two transitions independently. The data relevant to the first transition is shown in Figure 6.4, where the contrast between *ne* versus *ne...not* and *not* indicates potential competition for what expresses the specifier of the negative phrase. The data relevant to the second transition are shown in Figure 6.5, where the contrast between *not* versus *ne* and *ne...not* indicates potential competition for what expresses the head of the negative phrase.

For each of the transitions we need to decide on how to group the data together to calculate the fitness increments and perform the fitness increment test. There are two general considerations in doing so. The first consideration is that we need to make sure the assump-
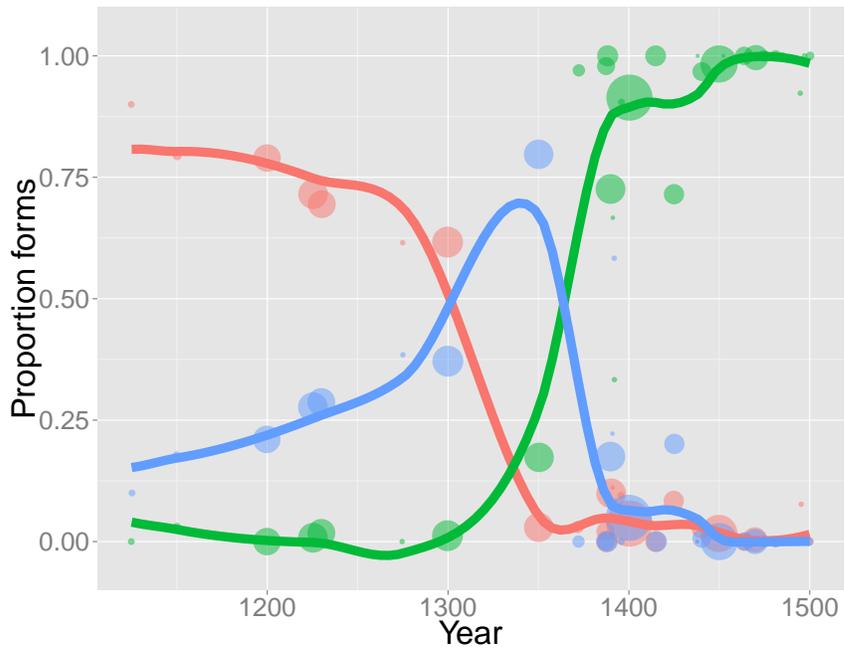
Figure 6.3: Proportion of forms of negation in Negative Declaratives
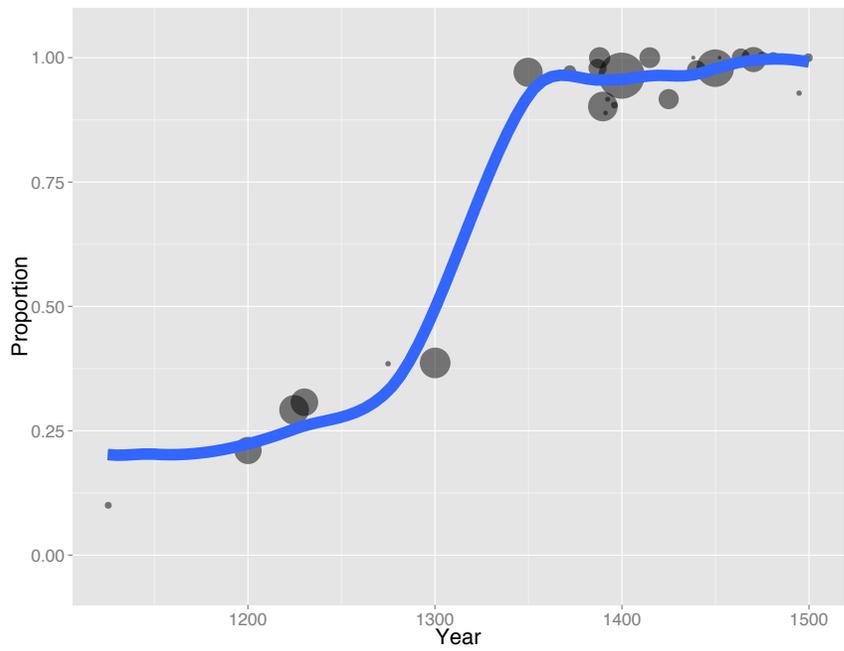
Figure 6.4: Proportion of *ne...not* and *not* versus *ne* over time
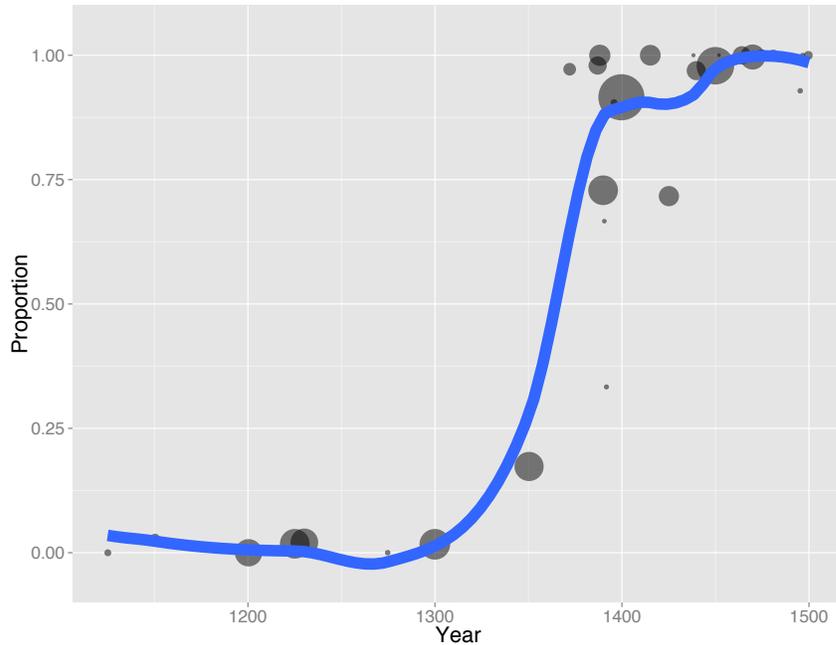
Figure 6.5: Proportion of *not* versus *ne* and *ne...not* over time

tions of the underlying Gaussian approximation are met. In particular, the approximation
cannot deal with absorption events before the last bin. If such an absorption event occurred,
then all subsequent bins should be the same. So, we need to make sure there are no bins
before the last one where only one variant is present. This basically serves as a limit on
how finely we can bin the data.

The second consideration is that we want the test to have as much statistical power
as possible, allowing us to reject the null hypothesis when it is indeed false. In practice,
the fitness increment test shows reasonable power when the increments include around a
thousand data points. To guarantee that bins have approximately equal numbers of data
points we bin by quantiles. This allows for the width of bins to vary, and we treat the
midpoint of each bin as the time of the sample measurement. Note that the scaling of the
fitness increments takes these time differences into account as well.

So, for each transition we do three things. First, we bin the data by quantiles and
perform the fitness increment test on the resulting bins. In what follows we use the finest

partition of the data that give us almost one thousand data points per bin but meets the assumptions of the Gaussian approximation. Namely, there are no absorption events and the fitness increments are normally distributed. Second, we perform the fitness increment test on the rescaled fitness increments to determine if we can reject the null hypothesis of drift. Third, if we can reject the null hypothesis then we numerically estimate the mostly likely selection coefficient and population size under the alternative hypothesis of selection; if we cannot reject the null hypothesis of drift then we numerically estimate the most likely population size without selection.

For the first transition we bin the data into six bins and confirm that there are no absorption events prior to the last bin and that the rescaled fitness increments are approximately normal according to the *Shapiro-Wilk test* ($p = 0.2050$). Given that these conditions are met, we perform the fitness increment test and note that we can reject the null hypothesis of drift ($\bar{Y} = 0.0278$, $t(5) = 2.6394$, $p = 0.0288$). The selection coefficient and population size that best explain the data under the alternative hypothesis of selection can be numerically inferred as $\hat{s} = 0.01913$ and $\hat{N} = 15900$.[3] While the selection coefficient from fitting the acquisition dynamics to the first transition are not directly comparable, it is interesting to note that in a finite population the selection coefficient is about five times smaller. This is a potentially important point to keep in mind when fitting deterministic mean dynamics to data.

For the second transition we bin the data into five bins and confirm that there are no absorption events prior to the last bin and that the rescaled fitness increments are approximately normal according to the *Shapiro-Wilk test* ($p = 0.1300$). Given that these conditions are met, we perform the fitness increment test and note that we cannot reject the null hy-

---

[3]Many thanks to Josh Plotkin and Mitchell Johnson for help installing and understanding the code, which can be found at `https://github.com/mnewberry/tsinfer`. Technically speaking, the population parameter reported here conflates another nuisance parameter regarding the time scale of the change. We leave interpreting the implications of the population parameter for future research.

pothesis of drift ($\bar{Y} = 0.0624$, $t(5) = 1.7021$, $p = 0.0820$). The population size under the null hypothesis of drift that best explains the data can be inferred numerically as $\hat{N} = 506$. We cannot reject the null hypothesis, so it makes sense that the population would have to be fairly small to account for the rapid change in the second transition.

So, we can reject the null hypothesis of drift in the first transition, but not in the case of the second transition. This first result makes sense. given the model of the functional cycle presented in Chapter 4. That is, the first transition of the formal cycle cannot be explained by random drift in syntactic acquisition, but it can be explained as an instance of the functional cycle due to pragmatic pressures. This second result interesting given that the second transition is the more dramatic of the changes, however, there two important things to note.

First, we have not shown that the second transition is due to drift. Rather, we have simply shown that we cannot reject the possibility. This could be due to the fact that the second transition is actually due to drift, or that we have simply failed to reject the null hypothesis even though it is not true. In this regard, we should note that the fitness increment test actually loses the power to detect selection under certain circumstances when selection is particularly strong (Feder et al. 2014:Figure 2 and 514-515). However, the power of the test depends on the number of data points per bin and the length of the time series in relation to the size of the population in complicated way. Determining whether the failure to reject the null hypothesis for the second transition is due to this is something we leave for future research.[4] For now though, given that we have no other alternative explanations to put forward, we lose nothing by simply noting that the second transition is consistent with random drift in syntactic acquisition.

Second, if the failure to reject the null hypothesis is indeed because the second tran-

---

[4]Although, these results do not change if we alter the number of bins that we divided the data into. We can never reject the null hypothesis. See Appendix D for the details.

sition is actually due to drift, then this offers an interesting explanation for the differing time courses across languages. For example, while the second transition quickly follows the first in the history of Middle English Wallage (2008) as well as Middle High German Jäger (2008). However, even in Middle Low German, where the second transition does go to completion we cannot reject drift using the data cited in Breitbarth (2009:109) ($\bar{Y} = 0.02644$, $t(3) = 0.8011$, $p = 0.2408$). Moreover, the second transition can take several hundred years, as is the case in the history of French (Martineau and Mougeon 2003) and Dutch (Burridge 1993). Indeed, some Flemish dialects still retain the embracing form (van der Auwera and Neuckermans 1997, Zeijlstra 2004). The fact that languages differ so widely in the amount of time spent between the first and the second transition has often been taken as a puzzling. However, when the second transition is viewed as a stochastic process, the varying amount of time makes perfect sense.

Now, claiming that the second transition in all of these languages may be due to drift obscures quite a bit of linguistic detail. But, it offers both theoretical and empirical leverage. For example, if we can rule out drift quantitatively in the second transition in a particular language, then we have a compelling reason to dive into theoretical analysis of the change in question. Ultimately, such an analysis should yield explanations of the observed change similar in form to our model of the functional cycle in Chapter 4 or the acquisition dynamics we presented in Chapter 5. From the other direction, if we have compelling theoretical reasons to believe that the second transition in a language happened for a particular reason, and we can build a model that explains its trajectory, but we cannot reject the null hypothesis of drift, then this shows us some of the limitations of the quantitative methods we have applied here.

At worst then, claiming that the second transition of the formal cycle may be due to drift is both empirically plausible and scientifically useful. Absent a better explanation, it makes sense of what has often been taken as a puzzling fact. More importantly, it clarifies

what would be needed to provide a better explanation.

Summary

In this chapter we applied statistical methods developed in population genetics to test the hypothesis of selection and drift in both transitions of the formal cycle in Middle English. We found that we could reject the null hypothesis of drift in the first but not the second transition. The result for the first transition makes sense. In fact, given the explanation of the formal cycle we offered in Chapter 4, we would be surprised if we could not reject drift in the first transition of the formal cycle. The result for the second transition also makes sense. The model we presented in Chapter 4 does not offer an explanation of the second transition. Moreover, given the syntactic structures posited to underly the formal cycle, the acquisition dynamics presented in Chapter 5 do not offer an explanation either. If neither acquisition nor use can explain the transition, then it might just be due to random drift.

Beyond the formal cycle, these results have important implications for stable linguistic variation in general. The acquisition dynamics allow for stable variation only when grammars are perfectly mutually incompatible. That is, when $s = 0$, the two grammars are perfectly balance and will not change over time. However, if we relax the assumption of an effectively infinite population that underlies this mean dynamics, we know that one or the other variant will win it. This follows from the fact that the absorbing states of the stochastic process occur where only one or the other variant is used. So, if acquisition is the only force acting on forms over time, then we will not observe linguistic variation at a particular level. Namely, there can never be variation between how a particular syntactic location is expressed.

The fact that we do observe stable variation suggests that there are other forces acting over the distribution of forms in the linguistic environment. Indeed, in some cases stable

variation has been observed over centuries, as is the case for the apical and velar variants of (ING) (Labov 1994): *hunting* versus *huntin'*. If not for some countervailing force, this variation should have arguably been extinguished, if these variants differ only in how they express some progressive aspectual head in the sense of *distributed morphology* (Embick and Noyer 2007).

Broadly speaking, it would seem that the most likely candidate for a countervailing force to the monomorphic effect of syntactic acquisition is meaning, which we take to include semantic, pragmatic, and sociostylistic information. That is, the stability of (ING) arguably stems from the fact that it signals style: *hunting* is formal whereas *huntin'* is not. So, while two variants may compete with each other to express the same syntactic position, they might stably coexist if they find complementary informational niches. This is exactly what we see with the specialization of the two (ING) variants to particular styles. As a slogan then, we might say: no stable variation without information. Note that this does not mean that stable informational variation is inherently stable. However, it would seem that it is more likely that two meanings might specialize rather than pool together. Of course, this is an empirical matter. One potential line of research to investigate such developments would be to employ word-embeddings to construct measures of similarity between the meanings of different variants over time.

Chapter 7

Conclusion

The main contribution of this dissertation is the demonstration that we need articulated models of both pragmatic and grammatical competence to construct causal models of language change. We summarize the contributions of each of the chapters towards this goal.

Chapter 2 offered the first distinction between the formal and functional Jespersen cycles that have so often been conflated. This terminological distinction simplified the explanatory burden and made clear what there is to be explained and how we should go about the explaining. We began with the functional cycle. Chapter 3 outlined the tools from evolutionary game theory that are so well suited to modeling how information is signaled in a population of boundedly rational agents over time. Chapter 4 applied these tools to model the functional cycle, incorporating experimental evidence regarding the limits of our pragmatic competence. We then turned to the formal cycle. Chapter 5 outlined the dynamics of a model of syntactic acquisition and demonstrated that it could not account for either of the transitions of the formal cycle. Chapter 6 applied methods from population genetics to show that while we can reject the possibility of drift in the the first transition of the formal cycle, we cannot do so in the second transition.

Simply put, it is our pragmatic competence that plays the central role in explaining

the dynamics of change in this case. Without taking it into consideration we are left with descriptions of different states of a language without a clue as to why a transition from one to another occurred. It should be noted that this does not mean that pragmatic competence will always occupy such a role. Rather, we should be mindful of the full range of factors that influence language change.

Appendix A

Equilibrium analysis

The full details of the analysis can be found at:

`https://github.com/christopherahern/dissertation`

Appendix B

Dynamic analysis

The full details of the analysis can be found at:

`https://github.com/christopherahern/dissertation`

Appendix C

Acquisition analysis

The full details of the analysis can be found at:

`https://github.com/christopherahern/dissertation`

Appendix D

Drift analysis

The full details of the analysis can be found at:

   `https://github.com/christopherahern/dissertation`

Bibliography

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

Aumann, Robert J. 1985. What is game theory trying to accomplish. *Frontiers of economics* 35.

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

van der Auwera, Johan, and Annemie Neuckermans. 1997. Jespersen's cycle and the interaction of predicate and quantifier negation. In *Dialectology meets typology. Dialect grammar from a cross-linguistic perspective*, ed. Bernd Kortmann, 453–478. Mouton de Gruyter.

Bailey, Charles-James N. 1973. *Variation and linguistic theory*. Center for Applied Linguistics.

Baker, Mark. 2008. The macroparameter in a microparametric world. In *The limits of syntactic variation*, vol. 132, 351. John Benjamins Publishing.

Bardsley, Nicholas, Judith Mehta, Chris Starmer, and Robert Sugden. 2010. Explaining focal points: Cognitive hierarchy theory versus team reasoning*. *The Economic Journal* 120(543):40–79.

Bernini, Giuliano, and Paolo Ramat. 1996. *Negative sentences in the languages of Europe: A typological approach.* Mouton de Gruyter.

Beukema, Frits. 1999. Five ways of saying no: the development of sentential negation in English in a Government and Binding perspective. In *Negation in the history of English*, ed. Ingrid Tieken-Boon van Ostade, Gunnell Tottie Ostade, and Wim van der Wurff, 9–26. Mouton de Gruyter.

Björnerstedt, Jonas, and Jörgen W Weibull. 1996. Nash equilibrium and evolution by imitation. In *The Rational Foundations of Economic Behaviour*, ed. K. J. Arrow, E. Colombatto, M. Perelman, and C. Schmidt, 155–181. St. Martin's Press.

Bloomfield, Leonard. 1927. Review of Jespersen's, The philosophy of grammar. *Journal of English and Germanic Philology* 26:444–446.

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

Breitbarth, Anne. 2009. A hybrid approach to Jespersenrqs cycle in West Germanic. *The Journal of Comparative Germanic Linguistics* 12(2):81–114.

Burridge, Kate. 1983. On the development of Dutch negation. In *Linguistics in the Netherlands 1983*, ed. Hans Bennis and W.U.S. van Lessen Kloeke, 31–40. Dordrecht: Foris.

———. 1993. *Syntactic change in germanic: aspects of language change in germanic with particular reference to middle dutch*, vol. 89. John Benjamins Publishing.

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

Bybee, Joan. 2003. *Phonology and language use*, vol. 94. Cambridge University Press.

Chafe, Wallace L. 1974. Language and consciousness. *Language* 111–133.

Chomsky, Noam. 1980. *Rules and Representations*. New York: Columbia University Press.

———. 1981. *Lectures on government and binding: The pisa lectures*. 9, Walter de Gruyter.

———. 1986. *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

———. 1995. *The minimalist program*. Cambridge University Press.

Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of english*. Harper & Row.

Chomsky, Noam, and Howard Lasnik. 1993. The theory of principles and parameters. In *Syntax: An international handbook of contemporary research*, ed. Joachim Jacobs, Arnim Von Stechow, and Wolfgang Sternefeld, 506–569. Mouton de Gruyter.

Clark, Herbert H, and Catherine R Marshall. 1981. Definite reference and mutual knowledge. In *Psycholinguistics: Critical concepts in psychology*. Cambridge University Press.

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

Cressman, Ross. 2003. *Evolutionary dynamics and extensive form games*, vol. 5. the MIT Press.

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.

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

Darwin, Charles. 1859. *On the origin of the species by natural selection*. Murray.

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

Dixit, Avinash, Susan Skeath, and David H Reiley. 2004. *Games of strategy*. WW Norton & Company Inc., New York.

Dryer, Matthew S. 1996. Focus, pragmatic presupposition, and activated propositions. *Journal of pragmatics* 26(4):475–523.

Ecay, Aaron. 2015. A multi-step analysis of the evolution of english do-support. Ph.D. thesis, University of Pennsylvania.

Eckardt, Regine. 2006. *Meaning change in grammaticalization: an enquiry into semantic reanalysis*. Oxford University Press.

Eigen, Manfred, and Peter Schuster. 1979. The hypercycle. *New York*.

Embick, David, and Rolf Noyer. 2007. Distributed morphology and the syntax/morphology interface. *The Oxford handbook of linguistic interfaces* 289–324.

Ewens, Warren J. 2012. *Mathematical population genetics 1: Theoretical introduction*, vol. 27. Springer Science & Business Media.

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

Feder, Alison F, Sergey Kryazhimskiy, and Joshua B Plotkin. 2014. Identifying signatures of selection in genetic time series. *Genetics* 196(2):509–522.

Fodor, Jerry A. 1975. *The language of thought*, vol. 5. Harvard University Press.

Frank, Michael C, Noah D Goodman, and Joshua B Tenenbaum. 2009. Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science* 20(5):578–585.

Frisch, Stefan. 1997. The change in negation in Middle English: a NEGP licensing account. *Lingua* 101(1):21–64.

Fudenberg, Drew, and David K. Levine. 1998. *The theory of learning in games*. MIT press.

Fudenberg, Drew, and Jean Tirole. 1991. *Game theory*. MIT press.

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

Gibson, Edward, and Kenneth Wexler. 1994. Triggers. *Linguistic Inquiry* 407–454.

Gintis, Herbert. 2000. *Game theory evolving*. Princeton University Press.

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

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

———. 1989. *Studies in the way of words*.

Grieve-Smith, Angus. 2009. The spread of change in french negation. Ph.D. thesis, University of New Mexico.

Gundel, Jeanette K, Nancy Hedberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 274–307.

Haegeman, Liliane. 1995. *The syntax of negation*. 75, Cambridge University Press.

Hansen, Maj-Britt Mosegaard. 2009. The grammaticalization of negative reinforcers in old and middle french: a discourse-functional approach. *Current trends in diachronic semantics and pragmatics* 227–251.

Hansen, Maj-Britt Mosegaard, and Jacqueline Visconti. 2009. On the diachrony of" re-inforced" negation in french and italian. *Grammaticalization and pragmatics: Facts, approaches, theoretical issues* 137–171.

———. 2012. The evolution of negation in French and Italian: Similarities and differences. *Folia Linguistica* 46(2).

Harsanyi, John C. 1967. Games with Incomplete Information Played by "Bayesian" Players, I–III: Part I. The Basic Model. *Management Science* 14(3):159–182.

———. 1968a. Games with Incomplete Information Played by "Bayesian" Players Part II. Bayesian Equilibrium Points. *Management Science* 14(5):320–334.

———. 1968b. Games with incomplete information played by lqBayesianrqplayers, part III. The basic probability distribution of the game. *Management Science* 14(7):486–502.

Haspelmath, Martin. 1997. *Indefinite pronouns*. Clarendon Press.

Heller, Daphna, Kristen S Gorman, and Michael K Tanenhaus. 2012. To name or to describe: Shared knowledge affects referential form. *Topics in cognitive science* 4(2): 290–305.

Hirsch, Morris W., and Stephen Smale. 1974. *Difierential Equations, Dynamical Systems, and Linear Algebra*. Academic Press.

Hoeksma, Jacob. 1997. Negation and negative concord in Middle Dutch. In *Negation and polarity: Syntax and Semantics*, ed. Danielle Forget, Paul Hirschbuller, France Martineau, and Maria-Luisa Rivero, 139–156. John Benjamins.

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

Hofbauer, Josef, and Karl Sigmund. 1998. *Evolutionary Games and Population Dynamics*. Cambridge University Press.

Honda, Isao. 2000. Negation: a cross-linguistic study. Ph.D. thesis, State University of New York at Buffalo.

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

Horn, Laurence. 2001. Flaubert triggers, squatitive negation, and other quirks of grammar. In *Perspectives on negation and polarity items*, ed. Jack Hoeksma, Hotze Rullmann, Victor Sanchez-Valencia, and Ton Van der Wouden, 173–200. John Benjamins.

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

Huttegger, Simon M., and Kevin J.S. Zollman. 2013. Methodology in biological game theory. *The British Journal for the Philosophy of Science*.

Huxley, Julian. 1942. *Evolution. the modern synthesis.* George Alien & Unwin Ltd.

Hyams, Nina, and Kenneth Wexler. 1993. On the grammatical basis of null subjects in child language. *Linguistic Inquiry* 421–459.

Ingason, Anton Karl, Julie Anne Legate, and Charles Yang. 2013. The evolutionary trajectory of the icelandic new passive. *University of Pennsylvania Working Papers in Linguistics* 19(2):11.

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

Jäger, Agnes. 2008. *History of german negation*. John Benjamins.

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

Kahneman, Daniel. 2011. *Thinking, fast and slow*. Macmillan.

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

Keysar, Boaz, Shuhong Lin, and Dale J Barr. 2003. Limits on theory of mind use in adults. *Cognition* 89(1):25–41.

Kimura, Motoo. 1955a. Solution of a process of random genetic drift with a continuous model. *Proceedings of the National Academy of Sciences of the United States of America* 41(3):144.

———. 1955b. Stochastic processes and distribution of gene frequencies under natural selection. In *Cold spring harbor symposia on quantitative biology*, vol. 20, 33–53. Cold Spring Harbor Laboratory Press.

———. 1968. Evolutionary rate at the molecular level. *Nature* 217(5129):624–626.

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

Klima, Edward S. 1964. Negation in english. In *The structure of language*, ed. Jerry Fodor and Jerrold Katz, 246–323. na.

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

Kroch, Anthony, and 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.

Kuhn, Harry Waldo. 1953. *Lectures on the theory of games*. Princeton University Press.

Labov, William. 1994. *Principles of Linguistic Change: Internal Factors*. Oxford: Blackwell.

———. 2010. *Principles of Linguistic Change: Cognitive and Cultural Factors*, vol. 3. Oxford: Wiley-Blackwell.

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.

———. 1970. General semantics. *Synthese* 22(1):18–67.

Lucas, Christopher. 2007. Jespersens cycle in arabic and berber1. *Transactions of the Philological Society* 105(3):398–431.

Martineau, France, and Raymond Mougeon. 2003. A sociolinguistic study of the origins of ne deletion in european and quebec french. *Language* 79(1):118–152.

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

Maynard Smith, John, and George Price. 1973. The Logic of Animal Conflict. *Nature* 246: 15.

Mazzon, Gabriella. 2004. *A history of english negation*. Pearson Longman.

McElreath, Richard, and Robert Boyd. 2008. *Mathematical models of social evolution*. University of Chicago Press.

McMahon, April MS. 1994. *Understanding language change*. Cambridge University Press.

Medina, Tamara Nicole, Jesse Snedeker, John C Trueswell, and Lila R Gleitman. 2011. How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences* 108(22):9014–9019.

Mehta, Judith, Chris Starmer, and Robert Sugden. 1994a. Focal points in pure coordination games: An experimental investigation. *Theory and Decision* 36(2):163–185.

———. 1994b. The nature of salience: An experimental investigation of pure coordination games. *The American Economic Review* 658–673.

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.

Moran, Patrick Alfred Pierce. 1958. Random processes in genetics. In *Mathematical proceedings of the cambridge philosophical society*, vol. 54, 60–71. Cambridge Univ Press.

Narendra, Kumpati S, and Mandayam AL Thathachar. 2012. *Learning automata: an introduction*. Dover.

Nash, John F. 1950a. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences* 36(1):48–49.

———. 1950b. Non-cooperative games. Ph.D. thesis, Princeton University.

———. 1951. Non-cooperative games. *The Annals of Mathematics* 54(2):286–295.

Niyogi, Partha. 2006. *The computational nature of language learning and evolution.* MIT press Cambridge, MA:.

Osborne, Martin J, and Ariel Rubinstein. 1994. *A course in game theory.* MIT press.

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

Pollock, Jean-Yves. 1989. Verb movement, universal grammar, and the structure of ip. *Linguistic inquiry* 365–424.

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

Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In *Radical pragmatics*, ed. Peter Cole, 223–255. Academic Press.

Quine, Willard V. 1960. *Word and object.* MIT Press.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of grammar*, 281–337. Springer.

Roberts, Ian, and Anna Roussou. 2003. *Syntactic change: A minimalist approach to grammaticalization.* Cambridge University Press.

Rosenthal, Robert W. 1979. Sequences of games with varying opponents. *Econometrica: Journal of the Econometric Society* 1353–1366.

Samuelson, Larry. 1997. *Evolutionary games and equilibrium selection.* The MIT Press.

Sandholm, William H. 2010. *Population games and evolutionary dynamics.* MIT press.

Sato, Yuzuru, Eizo Akiyama, and J Doyne Farmer. 2002. Chaos in learning a simple two-person game. *Proceedings of the National Academy of Sciences* 99(7):4748–4751.

Savitsky, Kenneth, Boaz Keysar, Nicholas Epley, Travis Carter, and Ashley Swanson. 2011. The closeness-communication bias: Increased egocentrism among friends versus strangers. *Journal of Experimental Social Psychology* 47(1):269–273.

Schelling, Thomas C. 1960. *The Strategy of Conflict*. Harvard university press.

Schlag, Karl H. 1998. Why imitate, and if so, how?: A boundedly rational approach to multi-armed bandits. *Journal of Economic Theory* 78(1):130–156.

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

———. 1990. *Analyticity and syntheticity: A diachronic perspective with special reference to romance languages*. Mouton de Gruyter.

Schwenter, Scott A. 2005. The pragmatics of negation in Brazilian Portuguese. *Lingua* 115(10):1427–1456.

———. 2006. Fine-tuning Jespersen's cycle. In *Drawing the boundaries of meaning: Neo-Gricean studies in pragmatics and semantics in honour of Laurence R. Horn*, ed. Betty T. Birner and Gregory Ward, 327–344. John Benjamins.

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.

———. 1957. *Models of man; social and rational*. Wiley.

Skyrms, Brian. 1996. *Evolution of the social contract*. Cambridge University Press.

———. 2004. *The Stag Hunt and the Evolution of Social Structure*. Cambridge: Cambridge University Press.

———. 2010. *Signals: Evolution, learning, and information*. Oxford University Press.

Smith, Kenny, Andrew DM Smith, and Richard A Blythe. 2011. Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science* 35(3):480–498.

Smith, Linda, and Chen Yu. 2008. Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition* 106(3):1558–1568.

Spence, M. 1973. Job market signaling. *The quarterly journal of Economics* 87(3):355–374.

Stalnaker, Robert. 1978. Assertion. In *Syntax and Semantics*, ed. Peter Cole, 315–32. New York: New York University Press.

———. 2002. Common ground. *Linguistics and philosophy* 25(5):701–721.

Stevens, Jon, John Trueswell, Charles Yang, and Lila Gleitman. 2013. The pursuit of word meanings.

Strogatz, Steven H. 2014. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview press.

Tadelis, Steven. 2013. *Game theory: an introduction*. Princeton University Press.

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

Thompson, Sandra A. 1998. A discourse explanation for the cross-linguistic differences in the grammar of interrogation and negation. In *Case, typology and grammar: In honor of barry j. blake*, 309–341.

Tottie, Gunnel. 1991. *Negation in English speech and writing: a study in variation*. Academic Press.

Trueswell, John C, Tamara Nicol Medina, Alon Hafri, and Lila R Gleitman. 2013. Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive psychology* 66(1):126–156.

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

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

Von Neumann, John, and Oskar Morgenstern. 1944. *Theory of games and economic behavior*. Princeton university press.

Wagner, Elliott O. 2012. Deterministic chaos and the evolution of meaning. *The British Journal for the Philosophy of Science* 547–575.

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

———. 2013. Functional differentiation and grammatical competition in the English Jespersen Cycle. *Journal of Historical Syntax* 2(1).

Wardlow Lane, Liane, Michelle Groisman, and Victor S Ferreira. 2006. Don't talk about pink elephants! Speakers' control over leaking private information during language production. *Psychological science* 17(4):273–277.

Weibull, Jörgen W. 1997. *Evolutionary game theory*. MIT press.

Wilks, Samuel S. 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics* 9(1):60–62.

Willis, David. 2005. A minimalist approach to Jespersens Cycle in Welsh. In *Grammatical Change: Origins, Nature, Outcomes*, ed. Dianne Jonas, John Whitman, and Andrew Garrett, 93–119. Oxford University Press.

Wu, Shali, and Boaz Keysar. 2007. The effect of information overlap on communication effectiveness. *Cognitive Science* 31(1):169–181.

Yang, Charles. 2000. Internal and external forces in language change. *Language variation and change* 12(03):231–250.

———. 2002. *Knowledge and learning in natural language*. Oxford University Press.

Yu, Chen, and Linda B Smith. 2007. Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science* 18(5):414–420.

Zanuttini, Raffaella. 1997. *Negation and clausal structure*. Oxford University Press.

Zeijlstra, Hedde. 2004. Sentential negation and negative concord. Ph.D. thesis, Netherlands Graduate School of Linguistics.