Signalling Games, Sociolinguistic Variation and the Construction of Style

Heather Burnett
Laboratoire de Linguistique Formelle
CNRS-Université de Paris-Diderot

May 24, 2016

Abstract

This paper develops a formal model of the subtle meaning differences that exist between grammatical alternatives in socially conditioned variation (called *variants*) and how these variants can be used by speakers as resources for constructing personal linguistic styles. More specifically, this paper introduces a new formal system, called *social meaning games (SMGs)*, which allows for the unification of variationist sociolinguistics and game-theoretic pragmatics, two fields that have had very little interaction in the past. Although remarks have been made concerning the possible usefulness of game-theoretic tools in the analysis of certain kinds of socially conditioned variable linguistic phenomena (Clark, 2014; Dror et al., 2013), a general framework uniting game-theoretic pragmatics and quantitative sociolinguistics has yet to be developed. This paper constructs such a framework through giving a formalization of the *Third Wave* approach to the meaning of variation (see Eckert, 2012, for a review) using (modified) *signalling games* (Lewis, 1969) and a probabilistic approach to speaker/listener beliefs of the kind commonly used in the *probabilistic pragmatics* framework (see Goodman and Lassiter, 2014; Franke and Jäger, 2016, for recent overviews).

1 Introduction

This paper develops a formal model of the subtle meaning differences that exist between grammatical alternatives in socially conditioned variation (called *variants*) and how these variants can be used by speakers as resources for constructing personal linguistic styles. The range of empirical phenomena that the proposed model aims to capture is exemplified through the grammatical alternations in (1)-(3). In terminology commonly used in the field of *variationist sociolinguistics* (Labov, 1963, 1966; Weinreich et al., 1968, et seq.),

alternations such as those shown below are called sociolinguistic *variables*, and they can be, among other things, phonetic in nature (like (1)-(2)) or morpho-syntactic (as in (3)).

(1) (ING)

a.	I'm working on my paper.	[iŋ]
b.	I'm workin' on my paper.	[in]

(2) /t/ release

a.	I want a glass of wa[t^h]er.	released /t/
b.	I want a glass of wat[ɾ]er.	flap

(3) Discourse *like*

a.	What are we doing tonight?	bare
b.	Like, what are we doing tonight?	like

More specifically, this paper introduces a new formal framework, called *social meaning games (SMGs)*, which allows for the unification of variationist sociolinguistics and *game-theoretic pragmatics* (see Benz et al., 2004, for an overview), two fields that have had very little interaction in the past[1]. Although remarks have been made concerning the possible usefulness of game-theoretic tools in the analysis of certain kinds of socially conditioned variable linguistic phenomena (Clark, 2014; Dror et al., 2013), a general formal framework uniting game-theoretic pragmatics and quantitative sociolinguistics has yet to be developed. This paper gives such a framework, called *Social Meaning Games* (SMGs). This paper constructs such a framework through giving a formalization of the *Third Wave* approach to the meaning of variation (see Eckert, 2012, for a review) using (modified) *signalling games* (Lewis, 1969) and a probabilistic approach to speaker/listener beliefs of the kind commonly used in the *probabilistic pragmatics* framework (see Goodman and Lassiter, 2014; Franke and Jäger, 2016, for recent overviews).

The paper is laid out as follows: in section 2, based on results from sociolinguistic perception studies, I observe (following others) that the use of one grammatical variant versus another can induce inferences on the part of the listener about the kinds of properties that characterize the speaker. I further argue that, in addition to being triggered by particular linguistic forms, these inferences share certain non-trivial properties with other kinds of

[1]Note that I am speaking here of the field of variationist (quantitative) sociolinguistics. There is a rich (and developing) tradition of work within the game-theoretic paradigm on some other sociolinguistic/pragmatic topics such as the formal modelization of politeness and social networks (see Van Rooy, 2003; Mühlenbernd and Franke, 2012; Mühlenbernd, 2013; Quinley and Mühlenbernd, 2012, among many others). Additionally, there is already some work aiming at integrating formal semantics/pragmatics and variationist sociolinguistics using non-game-theoretic methods (Smith et al., 2010; Acton, 2014, 2016; Beltrama, 2016, for example). However, (to my knowledge) there is no account within the game-theoretic paradigm of the kinds of phenomena that have been the focus of empirical work within the variationist tradition (to be described below).

inferences commonly studied in formal pragmatics. I therefore propose that social meaning of the kind studied in this paper should be viewed as an instance of *pragmatic enrichment*, and that, as a consequence, a unified framework that can treat both social meaning and other kinds of meaning in context should be developed. In section 3, I consider what the properties of such a framework should be. In particular, based on the results of sociolinguistic production studies, I argue that the social aspects[2] of linguistic variation should be analyzed as instances of interactive rational language use. This is most obviously seen through studies of intra-speaker variation (also known as *style shifting*); however, following previous research, I argue that there are reasons to think that inter-speaker variation (a.k.a. *social stratification*) should also be analyzed as the result of interactive rational language use. I then give a brief description of one influential theory within sociolinguistics which aims to derive both style shifting and social stratification from (informal) principles governing rational use: *Third Wave variation theory* (TW). Based conclusions from TW that both interactivity and rationality characterize all the social aspects of variation, I propose that a game-theoretic approach can be useful in modelling this kind of linguistic communication.

This being said, game-theoretic tools are extremely general and have been used in the analysis of a wide range of economic, social and biological phenomena[3]. Thus, for such a proposal to have any substantive content, one must be more precise concerning the definition of the games (the players, the space of options, utility functions etc.) and what their solution concepts are. In section 4, I give a concrete proposal for how to integrate sociolinguistic variation into the broader framework of game-theoretic pragmatics: *social meaning games*. I first define the games, and then I give some illustrations of the kinds of predictions that this framework makes for quantitative patters of sociolinguistic variation, on the one hand, and the options for and constraints on the construction of personal linguistic styles, on the other. I further show how this framework can capture both similarities and differences between social meaning and other kinds of pragmatic enrichment discussed in section 2. Section 5 provides some concluding remarks and explores how the proposals made in this paper for social meaning and the construction of linguistic style could be extended to other aspects of stylistic performance.

[2]The main proposals in this paper are limited to modelling the aspects of linguistic variation that are determined by what sociolinguists call *social* or *external* factors. Patterns of linguistic variation are also determined by other factors which are not social/strategic in nature (ex. general cognitive factors associated with linguistic production and comprehension, as well as grammatical factors (what Labov calls *internal* factors). I will make some remarks concerning how the analysis of social factors given here could be integrated into a broader theory of linguistic variation and change; however, I will not discuss non-social factors in great detail here.

[3]See Osborne and Rubinstein (1994) for an introduction to this vast field.

2 Social Meaning as Pragmatic Enrichment

Suppose we are having a conversation and the person that we are talking to says (1-a) (repeated as (4-a)). What do we understand from this utterance?

(4) a. I'm working on my paper. [iŋ]
b. I'm workin' on my paper. [in]

From hearing (4-a), we can certainly conclude that the speaker is working on their paper. Intuitively, it also seems as if we might be able infer some additional thing(s) from (4-a), possibly something about the properties of the speaker, of the working event, or maybe even of both. Likewise, if our interlocutor says (1-b) (repeated as (4-b)), we will definitely understand from this sentence that they are working on their paper. But again, it seems as if we might want to infer something extra from this utterance, crucially something that is different from what we inferred from (4-a).

One of the most common ways in which the properties of these extra inferences have been investigated in both social psychology and variationist sociolinguistics is through the use of an experimental paradigm known as the *matched guise technique (MGT)* (Lambert, 1967). In a MGT experiment, participants listen to samples of recorded speech that have been designed to differ in very specific and controlled ways. Participants hear one of two recordings (called *guises*) which differ only in the alternation studied. After hearing a recording, participants' beliefs and attitudes towards the recorded speaker are assessed in some way, most often via focus group and/or questionnaire. All efforts are made to ensure that the two recordings *match* as possible, modulo the forms under study. Indeed, many recent studies (such as the ones described below) use digital manipulation of naturalistic speech recordings to ensure that any observed differences in inferences that participants draw in different guises are **directly** attributable to the variable under study, not to some other aspect of the voice of the speaker or of the content of their discourse.

In her 2006 dissertation and subsequent work, Campbell-Kibler (2006, 2007, 2008) performed an MGT study with 124 American college students using stimuli formed from the speech of 8 different speakers investigating how the use of the variable (ING) influences listener beliefs and perceptions. This study yielded a variety of complex patterns (to be further discussed below), and her results show that there exist certain consistent associations between linguistic forms (*-ing* vs *-in'*) and property attributions for the listeners who participated in the experiment. For example, all speakers were rated as significantly more **educated** and more **articulate** in their *-ing* guises than in their *-in'* guises[4]. In

[4]In Campbell-Kibler (2006), articulateness was presented as a binary property, and was positively selected by 27% of listeners in an *-ing* guise and 21% of listeners in an *-in'* guise ($p = 0.037$; χ^2 test). Education, on the other hand, was presented as a gradable property on a six point scale ranging from *not at all educated* to *very educated*. The mean value for education in the *-ing* guise was 3.98; whereas, it was 3.81 for the *-in'* guise ($p = 0.007$; ANOVA) (see Campbell-Kibler, 2006, 103).

other words, in this study we see the existence of relationships between linguistic variants and cognitive representations associated with education and eloquence, at least for the participants of Campbell-Kibler's study.

Other studies on different variables have yielded the same kinds of results. For example, in order to investigate the social meaning of the /t/ release variable (2), Podesva et al. (2015) performed an MGT study with 70 American participants (the majority in their early 20s) using stimuli formed from political speeches of 6 American politicians (Barak Obama, John Edwards, Nancy Pelosi, George W. Bush, Hilary Clinton, and Condoleezza Rice). As in Campbell-Kibler's study, the /t/ release study yielded a number of results concerning associations with released vs flapped/unreleased /t/: for example, John Edwards and Condoleezza Rice were rated as significantly more **articulate** in their released /t/ guises than in their flapped guise (i.e. when they say things like *wa*[t^h]*er*, rather than wat[ɾ]er[5].). On the other hand, Nancy Pelosi was rated as significantly less **friendly** and less **sincere** when she used released /t/, and Barak Obama was rated as significantly more **passionate** in his flapped guise than in his released /t/ guise.

The results concerning Pelosi and Obama in the /t/ release study serve to highlight an important feature of social meaning: depending on a variety of factors (to be further discussed below), it may be the case that use of a reduced or 'non-standard' variant triggers property attributions on the part of the listener that the speaker could find desirable (see also Trudgill, 1972; Rickford and Closs Traugott, 1985, among many others). In other words, even though a speaker who uses a flap may risk being perceived as less articulate than if they had used a released /t/, they also have a better chance of coming across as friendly, sincere or passionate with the non-standard variant. Therefore, depending on the persona that they are trying to construct in the context, it may be worth the speaker's while to risk being perceived as inarticulate in favour of being considered more authentic and solidary with their interlocutors.

Finally, we note that similar patterns can be seen beyond the domain of phonetic variation. For example, in another MGT experiment investigating the social meaning of discourse *like* (3), Maddeaux and Dinkin (2015) show that speakers using *like* in a sentence like (5-a) are perceived to be significantly less articulate and less intelligent by a group of 69 Canadian undergraduates than in their guise without *like* (5-b).

(5) a. I couldn't get, like, a peaceful sleep. (Toronto English corpus (2/r/f11) Cited from (D'Arcy, 2005, 21))

b. I couldn't get a peaceful sleep.

In sum, I suggest that we can conclude from these studies (and the many others like them) that, in addition to extra information derived through pragmatic processes that are

[5]These results are unsurprising given that articulateness has been associated with released /t/ in many other corpus/ethnographically-based studies, such as Bunin Benor (2001); Bucholtz (1996); Podesva (2006); Eckert (2008).

more familiar to researchers in formal pragmatics, listeners derive extra information from an utterance concerning the properties that hold of the speaker, and these inferences are based on the particular linguistic forms that the speaker has chosen to employ. In other words, I suggest that the inferences triggered by socially meaningful variants are kinds of implicatures, similar (although not identical to) scalar implicatures (6-a) or implicatures generated by expressions with expressive content (6-b) (see also McConnell-Ginet, 2011; Acton, 2014, 2016; Smith et al., 2010, for additional support for versions of this claim).

(6) a. Mary ate **some** of the cookies.
Extra inference: *Mary did not eat all of the cookies.*
b. That **bastard** Kaplan got promoted! (Kaplan, 1999, 9)
Extra inference: *The speaker does not like Kaplan.*

For some variables (such as (ING)) all or most listeners draw the same robust inferences no matter who the speakers are. However, in many cases, which property attributions a particular variant will trigger will depend greatly on which other properties are believed to hold of the speaker. This feature can already been seen in the discussion of Podesva et al.'s /t/ release study above. In particular, while these researchers found significant relationships between articulateness and released /t/ with Edwards and Rice, these results were found only with these two speakers. Likewise, in this experiment, flapping made only Nancy Pelosi sound more friendly and sincere; no significant effect of friendliness or authenticity was found with the other politicians. As (Podesva et al., 2015, 79) say,

> Listeners appear to be less likely to associate released /t/s with competence-based meanings in Pelosi's speech–such as articulateness, intelligence or authoritativeness–as illustrated by the quotation in [(7)] from one of our focus group respondents [...] The focus group participant's comment acknowledges that release can be used to sound more authoritative, but asserts that in Pelosi's voice the authoritativeness is not genuine.

(7) **Participant 1:** Um, I don't want to sound, say, fake, but she just sounded-like she was-it was just very-it contrasted from her other speech, so I-um, she sounded like she was trying to be more authoritative in her speech, perhaps.
071116-1140-2-3 (Focus group participants: *WF, WF, WF)

Thus, because (as Podesva et al. hypothesize) speakers find Pelosi's use of released /t/ to be pretentious and fake, using the flap makes her sound more sincere and creates a positive evaluation. In other words, these results show that social enrichment is dependent on speaker identity, but also (more importantly) on listeners' **interpretations** of speakers' linguistic performances. I will call this property of social meaning LISTENER ORIENTATION.

That social enrichment is listener oriented can also be seen in Campbell-Kibler's study of

(ING). In addition to a fair number of interpretative effects being limited to the speech of a single speaker or to a proper subset of speakers[6], Campbell-Kibler (2006, 2008) finds that listeners can assign different interpretations to the exact same linguistic performances of a single speaker. For example, after hearing recordings by the speaker (pseudo)named Elizabeth in which she discusses groups of people to which she herself does not belong, listeners in this MGT study were divided on which properties to attribute to this speaker. As shown in the table reproduced (from Campbell-Kibler, 2008, 646) as Table 1, after hearing the *-in'* guise, 30% of speakers thought Elizabeth was compassionate; however, another 17.5% of speakers thought that she sounded condescending in her *-in'* guise. Unsurprisingly, the selections for compassionate and condescending were virtually disjoint, with only a single speaker selecting both of these properties (Campbell-Kibler, 2008, 645). On the other hand, when speakers heard Elizabeth say *-ing* (which some focus group members describe what they think that she would "naturally' say (Campbell-Kibler, 2008, 648)), no speakers think that she sounds condescending and only 7.4% of speakers selected *compassionate*.

	% listeners selecting checkbox		
Checkbox label	-in	-ing	sig.
compassionate	**30.0**	7.4	0.022
condescending	**17.5**	0.0	0.005

Table 1 – Compassionate/condescending selections (Elizabeth's "other" recordings)

These results (and others like these) showing listener orientation highlight the differences between other kinds of enrichment and social enrichment of the kind discussed here: as Campbell-Kibler says (p.639)[7],

> The audiences for whom we perform on a day-to day basis are not obligated to accept our accounts of ourselves, even if they share a common ground with us regarding the basic meaning of our semiotic choices. As a result the process of constructing linguistic (and other social) performances is not like encoding a secret message, where we can trust that the recipient is seeking to uncover exactly the message we intend to send. Instead, social performance is more like choosing a name for a child: we may study name books and quiz friends about childhood memories of insulting nicknames, but once the name is chosen, we ultimately have no control over what someone gets called on the playground-that is, what interpretations others assign to our chosen resources.

Enrichment of the kind that we find with scalar implicatures (6-a), which proceeds by strengthening the truth conditional information of an utterance **is** often considered to occur as a kind of 'secret message' transmission, where the hearer tries to reason cooperatively to try to figure out what the speaker meant in the context (Lewis, 1969; Grice, 1975). Likewise, enrichment of the kind that we find with expressives (6-b) (ex. *bastard*) has been

[6]See (Campbell-Kibler, 2006, chapter 4) for a summary.

[7]See also the recent discussion in Levon (2014).

Kind of Implicature	Orientation	Example
Scalar	Objective (satisfaction depends on the world)	***Some*** *of the cookies*
Expressive	Speaker (satisfaction depends on the speaker)	*That* ***bastard*** *Kaplan*
Social	Listener (satisfaction depends on the listener)	(ING), /t/ release, like

Table 2 – Orientations of Pragmatic Implicatures

argued to be *speaker oriented* (Kaplan, 1999; Potts, 2007, among others); i.e. the listener's action is to simply attribute a subjective belief to the speaker; they certainly don't get to decide what belief is. Thus, I suggest that social meaning fills an interesting gap in our typology of pragmatic implicatures[8], as shown in Table 2.

Finally, I note that social enrichment shares another notable property with (some) other kinds of pragmatic enrichment: it shows (what I will call) GRAMMATICAL DEPENDENCE. More specifically, as has been observed since the beginnings of variationist sociolinguistics (ex. Labov, 1966), the environments in which social enrichment is possible, and what the nature of that enrichment is, are defined by the phonological or morpho-syntactic grammar. To continue with the examples discussed already in this paper: the conditioning of the variable (ING) has been shown to be affected by both grammatical category (Labov, 1966, and many others) and the abstract morphological structure of the variant (Houston, 1985; Tamminga, 2014). Additionally, Podesva et al. (2015) shows that social enrichment for /t/ release is easier when the pertinent variant occupies a word-medial position (8-a) rather than final position (8-b).

(8) a. I want a glass of wa/**t**/er.
b. We should mee/**t**/.

Likewise for discourse *like*: Maddeaux and Dinkin (2015) show that while noun-phrase initial *like* (5-a) can trigger inferences of unintelligence, the use of discourse *like* in other syntactic positions, such as verb-phrase initial *like* (9), is not socially meaningful in the same way.

(9) I'm not sure if my eight year old **like** understands that.
(N/X/m/46), cited in (D'Arcy, 2005, 172)

A more striking example of grammatical dependence of social meaning comes from varia-

[8]It may be the case that some expressions like swear words (*bastard*) may do double duty, being speaker oriented in the sense that the listener must conclude that the speaker has a bad opinion of Kaplan, but upon hearing *bastard*, the listener may conclude that the speaker is tough or rebellious (or brutish or vulgar, depending on the interpretation). Furthermore, (as we will see in the next section) the speaker can exploit this likely inference. So the expressive content of *bastard* would be speaker-oriented, but its social meaning would be listener oriented. However, I leave the careful study of different 'layers' of pragmatic meaning to future work.

tion in the future tense in dialects of French. In some varieties of Canadian and European French, the semantic difference between the periphrastic future (10-a) and the simple future (10-b) has been neutralized, and these forms are in variation in the dialects in question. This variation is conditioned primarily by properties such as age and social class, with (depending on the study) older and/or more educated speakers favouring the more 'standard' simple future form (10-b) compared to younger and/or less educated speakers (see Deshaies and Laforge, 1981; Poplack and Turpin, 1999; Wagner and Sankoff, 2011; Sankoff et al., 2012, among many others).

(10) a. Je **vais** manger.
b. Je manger**ai**. 'I will eat.'

Interestingly, in quantitative studies of this variable, such as (Wagner and Sankoff, 2011; Sankoff et al., 2012, for Montréal French), social stratification is found **only** in affirmative sentences. If future tense is interpreted within the scope of a semantically decreasing operator, such as negation (11), then stratification disappears and all members of the community strongly prefer the simple future.

(11) Je vais **pas** manger. ≈ Je mangerai **pas**.

Although this phenomenon would need to be investigated in greater detail using, for example, experimental methodology, these preliminary results suggest that social enrichment of the future variants, which creates the stratificational patterns observed in the studies cited above, is blocked in decreasing contexts. It is well-known that pragmatic processes like scalar enrichment are grammatically dependent, and these inferences are particularly sensitive to decreasing operators (Ducrot, 1969; Récanati, 2003; Chierchia et al., 2008; Chemla and Spector, 2011; Potts et al., 2015, among many others). Thus, I tentatively suggest that social enrichment can have a very close relationship with the structure building operations of the grammar, possibly similar to that of scalar implicatures[9]; however, I leave the very complex (but empirically rich) project of fleshing out the grammatical dependencies that social enrichment is subject to to future research.

In summary, following previous research into sociolinguistic perception, I have argued that social meaning should be viewed as an implicature that is triggered by the use of particular variants. As such, I propose that social meaning should be integrated into a broader theory of formal pragmatics, while still accounting for its (possibly unique) *listener oriented* quality. Of course, there are very many pragmatic frameworks with very many different properties available in the literature that we might choose from for this integration. In the next section, I argue that there is one framework in particular that looks especially

[9]Expressives on the other hand have been argued not to have so many interactions with the grammar, i.e. inferences associated with expressive content tend to be preserved in embedded environments and remain within the scope of decreasing operators (Kaplan, 1999; Potts, 2007, among others).

promising: game-theoretic pragmatics.

3 Sociolinguistic Variation as Rational Language Use

This section argues, following previous work on sociolinguistic production, that speakers have implicit knowledge of the kinds of inferences that listeners draw based on their linguistic usage patterns, and that they exploit this knowledge in order to influence which properties their interlocutor will attribute to them. In other words, in this section, I will argue that linguistic variation is a social phenomenon that is both **interactive**, in the sense that speakers and listeners make hypotheses concerning their interlocutors' beliefs and interpretation strategies[10], and (approximately) **rational**, in the sense that speaker/listener behaviours are (loosely) optimized to some criteria (Anderson, 1991). I first demonstrate these proposals with reference to intra-speaker variation (*style shifting*) and then make similar observations with respect to inter-speaker variation based on the perspective developed in *Third Wave* variation theory. The general conclusion to be drawn from this section is that a formal model of social meaning and its relation to socially conditioned patterns of linguistic variation should be able to capture both the interactive and rational aspect of the phenomena under study. Since interactivity and rationality are built into the architectures of game-theoretic approaches to meaning in context, I suggest that game theoretic tools are particularly well-suited to modelling this kind of communication.

3.1 Style-shifting as rational language use

A particularly clear example of linguistic variation as rational language use comes from existence of contextually-based intra-speaker variation, i.e. *style shifting*. This is a robust, well-documented phenomenon, and we can give a first simple illustration of it from Labov (2012)'s study of President Obama's use of (ING). (Labov, 2012, 22) finds significant differences in Obama's use of (ING) across three different recordings taken in three different contexts: (what Labov calls) *casual, careful* and *formal*. The first recording that Labov studied was one of Obama barbecuing at a Father's Day barbecue on the Whitehouse lawn: a 'casual' context. Labov finds that Obama uses *-in'* 72% of the time in this context, i.e. he is doing a lot of *grillin'*, *eatin'* and *drinkin'* at the barbecue. Then the barbecue finishes, and Obama moves to answer political questions from reporters on the Whitehouse lawn. In this 'careful' context, his rate of *-in'* drops to 33%. Finally, Labov studied Obama's use of (ING) in a scripted acceptance speech at the Democratic National Convention (a 'formal' context). He finds that, in this recording, the President uses *-in'* only 3% of the time. Obama's use of (ING) across three different contexts is summarized in Figure 1, reproduced from (Labov, 2012, 22).

[10]This was already demonstrated for listeners in the previous section and within the works cited. So this section concerns speakers.

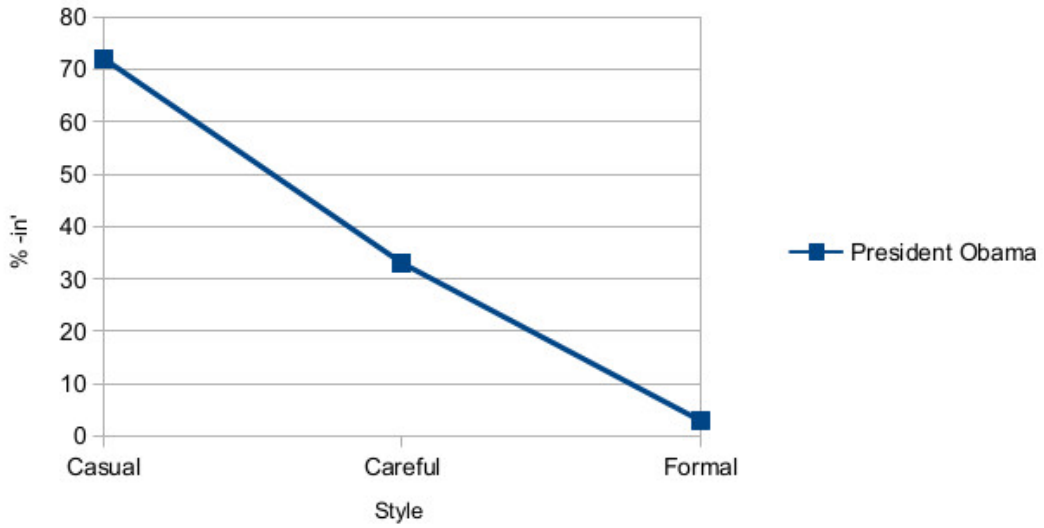

Figure 1 – President Obama's use of (ING) across three contexts

Why do we find this pattern? According to Labov, it arises because we have conventionally associated meanings with *-in'* and *-ing*, which allow us to communicate extra information to each other through phonetic variation. He describes what he calls our *hidden consensus* as follows (p. 22)[11]:

> This consensus is publicly available and in one sense, understood by all. In the classroom, or on the pulpit, people will attribute the use of the -in' form to laziness, ignorance, or just plain rascality. Yet the high value we put on the -in' norm in other contexts is not hidden from public view. When we see the large illuminated sign, DUNKIN' DONUTS, we recognize the claim that dunkin' doughnuts taste better than dunking doughnuts. . . A Philadelphia travel agency is named with an electric sign spelling out CRUSIN'. We understand this as an advertisement that we will have a better time cruisin' than we would cruising.

I have chosen to give a first illustration of style shifting using President Obama, and (for concreteness) we will continue to study this example throughout the paper. However, it is important to stress that style-shifting is not a phenomenon that is uniquely associated with public figures, although these are the kinds of individuals for whom we tend to have the most available data. For example, Podesva (2004) (cited in Eckert (2005)) finds significant differences in a medical student's use of /t/ release in a clinic setting and when he is at a barbecue. Indeed, recent studies of style shifting of private citizens have shown that intra-

[11]This example also highlights the conventionalization of the meanings associated with the variable. If the inferences associated with (ING) were directly and straightforwardly iconic (i.e. based solely on general reasoning processes associated with rate of speech or carefulness in articulation), we would not expect the difference between the written forms of *-in'* and *-ing* to be meaningful, which it clearly is.

speaker variation is widespread, with people significantly changing their use of variants across contexts (Cheshire, 1982; Podesva, 2007; Gratton, 2015, among many others) and even across sections of discourse (see Kiesling, 2009, for example).

It is also important to note that, while, at first glance, it may be tempting to reduce the use of variables like (ING) or /t/ release to a question of formality or 'register', with the unreduced more 'standard' variant being limited to formal contexts and the reduced 'vernacular' variant being limited to informal contexts, there are reasons to think that such a theory is too simplistic and that (ING) can be used to communicate properties that are not directly related to formality. One study that shows this particularly well is Gratton (2015)'s investigation of the link between (ING) and gender presentation. In this work, Gratton conducted group interviews with *non-binary* individuals, that is, individuals whose gender identity does not respect the male/female binary. She focussed on two people: Flynn, who was assigned female at birth, and Casey, who was assigned male at birth, and she conducted two sets of interviews with these consultants: the first one being in a queer-friendly environment (their home and a queer-café, respectively) and the second one taking place in a public café. As shown in Table 3, reproduced from Gratton (2015)'s Tables 1-2, neither Flynn nor Casey show a significant difference in their use of *-in'* vs *-ing* in the queer-friendly environments. However, in the public café, things are very different: Flynn, who was assigned female at birth, uses significantly more *-in'* (80%), while Casey, who was assigned male at birth, uses significantly more *-ing* (89%).

		Home/Queer café		Popular café	
	Variant	N	%	N	%
Flynn	-ing	18	44	13	20
	-in'	23	56	52	80
	Total N	41		65	$\chi^2 = 5.84, p = .0157$
Casey	-ing	42	58	86	89
	-in'	30	42	11	11
	Total N	72		97	$\chi^2 = 19.07, p < .0001$

Table 3 – (ING) in queer-friendly vs public settings, based on (Gratton, 2015, Tables 1-2)

A naive theory in which the use of *-in'* vs *-ing* was uniquely determined by formality or 'register' cannot account for this pattern, since it is not obvious that the two interview contexts differ in formality. Furthermore, even if we suppose (for example) that the queer-friendly café is less formal than the public café, our 'formality' analysis could only explain Casey's behaviour; Flynn's pattern, where they increase their use of *-in'* in the public café, is unexpected. Instead of formality, what Gratton suggests is going on in Table 3 (based on an ethnographic analysis of her interviews) is that "consultants vary their use of (ING) between situations based on their perceived level of security in presenting their gender identity". Indeed, in studies of demographically balanced corpora of naturalistic speech, it is common to find a stratificational pattern in which men favour *-in'* and women favour *-ing* (see Hazen, 2006, for a review). Thus, there appears to be some association between

-in' and masculinity[12] and/or *-ing* and femininity.

In summary, the studies discussed in this section show that speakers assess how their speech will be evaluated by their interlocutors in a particular discourse context, i.e. which properties that they think their interlocutors will attribute to them. In other words, sociolinguistic variation is an *interactive* phenomenon. Then, after this evaluation, speakers choose the form that (they think) will be the most successful to construct their desired persona. In other words, there is an aspect of *optimization/rationality* as well. Since interactivity and rationality form an important part of the architecture of game-theoretic frameworks, I propose that such approaches are particularly well adapted to modelling this kind of linguistic communication.

This being said, style-shifting is only one of the focusses of variationist sociolinguistics, with the majority of the most famous results from this field being associated with patterns of *social stratification*, i.e. inter-speaker differences in the use of variants. In the next section, (following others) I will argue that cases of social stratification should also be analyzed as rational language use, and thus the game-theoretic model that will be presented in section 4 has the potential to be applied to the full range of social aspects of linguistic variation.

3.2 Social stratification as rational language use

Since the beginning of the quantitative study of sociolinguistic variation, there has been an interest in developing a unified theory of style shifting and the kind of variation that has been the principal empirical focus of variationist sociolinguistics: social stratification. An obvious motivation for such a theory comes from the observation that the exact same linguistic variables are used in both the intra-speaker and inter-speaker dimensions of variation. For example, consider the graph in Figure 2, reproduced from Labov (1966)'s famous study of language use in New York City (Labov's Table 10.7). As we saw with Obama in the previous section, the use of the *-in'* variant decreases with a more formal style[13], but also with a rise in social class. In this way, (in the words of Labov) it becomes difficult to distinguish "a casual salesman from a careful pipefitter" (Labov, 1972, 240).

Likewise, in developing his influential *Audience Design* theory of style shift, (Bell, 1984, 151) says that the relation between style shifting and social stratification is "more than an interrelation. It is a *derivation* [Bell's emphasis], which can be expressed as an axiom of sociolinguistic structure." He calls this axiom the *Style Axiom*.

(12) **Bell's Style Axiom**
Variation on the style dimension within the speech of a single speaker derives from

[12]Eckert (2005) and Campbell-Kibler (2006) suggest that this relationship is mediated by *casualness*.

[13]Note that in the 1966 study, *casual, careful* and *formal* styles correspond to interview speech, reading passages and word lists, respectively. This is different from categories in the 2012 Obama study; however, the overall point remains the same here.

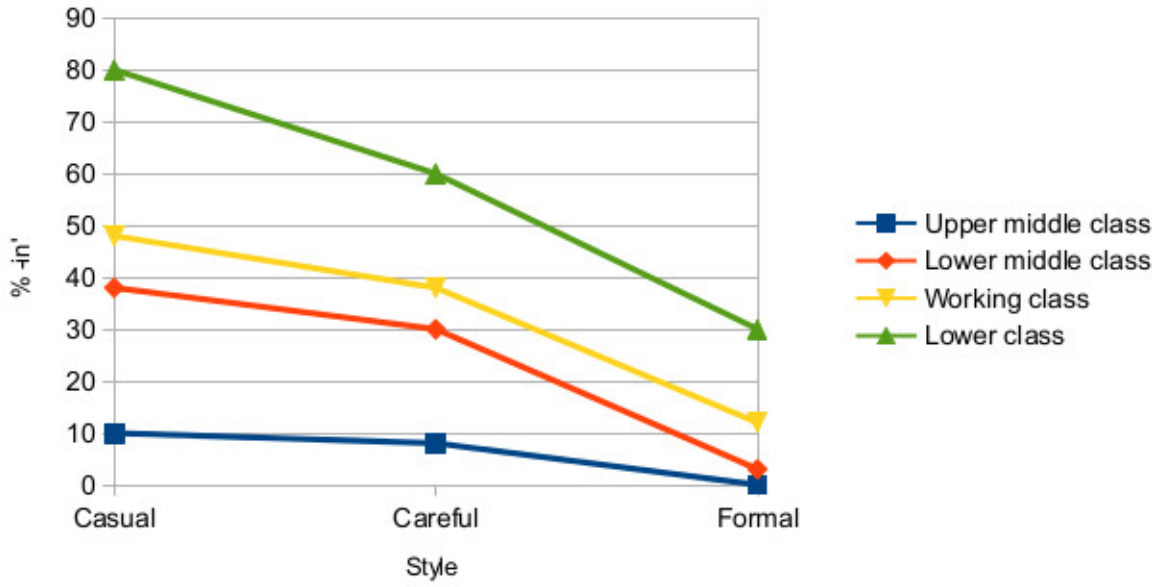

Figure 2 – Labov (1966): (ING) by social class and style (casual, careful, formal) in NYC

> *and echoes the variation which exists between speakers on the "social" dimension.*
> (Bell, 1984, 151)

In this section, I argue (following previous research, particularly within the *Third Wave*) that stratificational patterns are derived from the same basic principles as those that underly the creation of patterns of style shifting: the principles governing rational language use.

As Eckert (2008) notes, this point can already be made (for at least some cases of social stratification) based on Labov (1963)'s study of, among other variables, the centralization of diphthongs /ay/ and /aw/ in Martha's Vineyard, an island south of Cape Cod in Massachusetts. Based on data from 69 sociolinguistic interviews, Labov found that, rather than being conditioned by gender, age or other demographic categories, centralization on Martha's Vineyard was best predicted through looking at speakers' stances towards the changing demographics of the island. During the period of Labov's investigation, the main industries on the Vineyard were in the process of moving from whaling and fishing to tourism, creating significant hardships for islanders who had built their lives around the fishing industry. As such, the participants in Labov's study were divided with respect to how they viewed these changes, having reactions "varying from a fiercely defensive contempt for outsiders to enthusiastic plans for furthering the tourist economy" (Labov, 1963, 28). Furthermore, as Labov observes, speakers' orientations towards or away from the island and the old way of life make an enormous difference with respect to which variant

they prefer to use, with more locally oriented individuals showing much higher degrees of centralization (Labov, 1963, 30). This contrast is particularly clear among younger speakers, who must decide whether or not to leave the island for good when they finish high school. Labov points out (p. 32) that "comparatively few of the sons of the English-descent group will be earning their living on the Vineyard in the next 20 years" and he says that "a marked contrast was observed between those who plan to leave the island and those who do not. The latter show strong centralization, while the former show little, if any." He illustrates this point with the comparison between four 15 year old students of English descent, who differ in their future plans on the island (reproduced in Table 4[14]). Labov reports that "the two down-islanders who intend to leave for careers in business and finance show little or no centralization: the two up-islanders who hope to go to college and return to make their living on the island show considerable centralization. The indexes speak for themselves" (Labov, 1963, 32).

	Speaker 1	Speaker 2
Down-island, leaving	00-40	00-00
Up-island, staying	90-100	113-119

Table 4 – Centralization indexes for /ay/-/aw/ for four students (based on Labov, 1963, 32)

Another group of studies that highlight the agentive source of social stratification are Eckert (1989, 2000)'s studies of linguistic variation in Belten High, a high school in a middle class suburb of Detroit. In these works, Eckert studies the linguistic and social conditioning of a variety of sociophonetic variables and one syntactic variable (negative concord). Eckert observes that within the high school there was a high degree of inter-speaker variation, polarized largely around two social groups in the school: the *jocks* and the *burnouts*. As Eckert (1989, 2000) shows, members of these two groups distinguish themselves in very many ways: through their actions, their attitudes, their ways of dressing, and also their use of the variables under study. Although the jocks and the burnouts were the most 'extreme' users of the different variants, Eckert (2000) found that, in the whole population of Belten High, use of variants was best predicted by students' urban/suburban orientation and/or institutional engagement, where urban orientation was measured by activities like *cruising*[15], as shown in Table 5, and their institutional engagement was measured by participation in (school's) extra-curricular activities[16].

In summary, studies such as those conducted by Labov and Eckert show that (at least in the cases that they studied) inter-speaker differences break down along lines associated

[14]The values in Table 4 represent a measure of centralization, with higher values corresponding to more centralization.

[15]Cruising involves going joy-riding around the edge of the urban centre of Detroit.

[16]Note that (Eckert, 2000, 170) found that involvement in school organized extra-curricular activities was an even better predictor of linguistic variation than academic achievement, so her results show that students' use of variables is determined by factors other than simple mastery of the 'standard language' taught in class.

Variable		+Cruising	-Cruising	Input	Sig.
(∧) backing	Girls	.563	.458	.422	.000
	Boys	.530	.460	.447	.014
(e) backing	Girls	.544	.464	.331	.029
	Boys	.557	.437	.368	.001
(ay) raising	Girls	.765	.381	.011	.000
	Boys	.636	.295	.009	.004
(ay) monophongization	Girls	.634	.405	.036	.000
negative concord	Girls	.777	.294	.106	.000
	Boys	.637	.338	.241	.000
(oh) fronting	Girls	.583	.440	.259	.012
(aeh) raising	Girls	.569	.450	.447	.000
	Boys	.465	.538	.299	.033
(∧) fronting	Boys	.419	.606	.027	.009

Table 5 – Factor weights for cruising at Belten High (Eckert, 2000, 151)

with other social practices, such as moving, cruising or participating in extra-curricular activities. As Eckert (2000) observes, these correlations can be immediately explained if the use of one variant over another is itself also a social practice, saying (p.169),

> To the extent that we find linguistic differences correlating across the population in such a way that social category recedes into the background, one might ask whether the linguistic behaviour of the categories themselves is not simply derivative.

Thus, the formal theory that we are developing should also be able to capture the rationality underlying social stratification.

3.3 Third Wave variation theory

Above, I argued that, given its empirical properties, it is natural to attempt to model sociolinguistic variation within a game-theoretic framework. This being said, since game theoretic tools are so general, without saying anything else, we are still very far away from a full theory of social meaning and variation[17]. Fortunately, there is already a well-articulated and influential theory in the sociolinguistics literature that can provide the basis for our formal implementation. As mentioned above, Third Wave variation theory pursues a unified analysis of style shifting and social stratification as rational language use, and (in a nutshell) the account is as follows[18]: The use of a variant in context is related to, or *indexes*, sets of properties, stances or other concepts/ideas that are to be attributed

[17]See also the discussion in Dror et al. (2013).

[18]Properly speaking, what I present here is just a small portion of the full TW theory. In particular, my model does not cover the parts concerning how variants come to index particular properties and their

to the speaker (Silverstein, 1976; Ochs, 1992, and others)). Speakers use these linguistic resources to (attempt to) construct the persona that will be the most useful to them in their context-specific goals. Here we take the notion of *goals* to be very general, encompassing concrete aims such as getting a job in a flower shop, but also more abstract things such as making friends or even communicating one's 'true' self to their interlocutor. By virtue of what speakers' goals and desires are, and by virtue of what resources they have to use, different variants will be more useful to different speakers in different contexts. Thus, in the same way that the properties indexed by *-ing* are more useful to someone like Obama in a formal setting than in an informal setting, the properties indexed by *-ing* are more useful to upper middle class speakers (in the context of being interviewed by a researcher) than to working class speakers[19].

As described above, TW clearly makes reference to notions like interactivity and rationality; however, it is not a mathematical theory. Thus, although its insights concerning social meaning and sociolinguistic variation are clear, by virtue of its form, such a theory cannot be directly incorporated into a broader theory of formal pragmatics. What I propose to do in the second half of the paper is to formalize Third Wave variation theory using signalling games and a probabilistic approach to speaker/listener uncertainty. In doing so, I hope to bring social meaning into the domain of game-theoretic and probabilistic pragmatics.

4 Social Meaning Games

This section presents the *social meaning game* (SMG) framework. In what follows, I will define the games in detail and illustrate the different aspects of the definitions with a series of examples. For convenience, the full framework is more succinctly laid out in the appendix.

As mentioned above, the framework combines a modification of Lewis (1969)'s signalling games with a probabilistic/Bayesian approach to speaker/listener beliefs and uncertainty (see Tenenbaum et al., 2011, for an overview)[20]. In a nutshell, a *signalling game* is a game of coordination between two agents, S (the *speaker/sender*) and R (the *listener/receiver*). S has a piece of information that they wish to communicate to R (their *type*). S's action is to choose a message m to send R, and R's action is to assign an interpretation to m, and,

implications for a theory of language change. In the model that will be presented in the next section, it is stipulated which variants are associated with which sets of properties (indexical fields), and the model does not have any evolutionary or large-scale diachronic component. However, I believe that it would be of great interest to extend my proposal to capture these other aspects of TW in the future.

[19]See Bourdieu and Passeron (1970); Bourdieu (1979, 1980) for detailed studies of the complex relationship between social class and values, goals and desires.

[20]There are certainly other possibilities for this aspect of the framework (see Halpern, 2003, for other options), but we stick with a Bayesian approach here so that this work can be comparable to the work done on other kinds of pragmatic phenomena within the probabilistic pragmatics framework (see Goodman and Lassiter, 2014; Franke and Jäger, 2016).

in doing so, update their prior beliefs about the world using the information communicated by m (Stalnaker, 1978; Lewis, 1979; Heim, 1982, among others). In signalling games, S and R's payoffs are calculated based on coordination: (broadly speaking) both players win if R correctly interprets S's message, updating their beliefs accordingly, and they both lose if S's type and R's interpretation do not converge, and R comes to believe something different about the world than that which S intended.

Social meaning games will have a similar structure: they are games of coordination between two agents: S (speaker/sender) and R (listener/receiver). S has a set of properties characterizing themself that they wish to communicate to R (their *type*). S's action is to choose a message m to send R, and R's action is to attribute a set of properties to S based on m and their prior beliefs about S, and, in doing so, update their beliefs. S and R's payoffs are calculated in part based on coordination; however, contrary to classic signalling games, something else will play an important a role in calculating agents' utility: S and R's personal preferences in the context, which we will call their *values*.

4.1 Basic Setup

The definition of a SMG is laid out more formally in Def. 4.1. Def. 4.1 give a general overview of the 'moving parts' of the model. The various lines of this definition will be further elaborated in the rest of this section.

Definition 4.1. *A **Social Meaning Game** is a tuple* $\langle \{S, R\}, \langle \mathbb{P}, > \rangle, M, C, [\cdot], \mu_S, \mu_R, \rho, \sigma, Pr \rangle$ *where:*

1. *S and R are the players.*
2. *$\langle \mathbb{P}, > \rangle$ is the **universe** (a relational structure), where*
   - *$\mathbb{P} = \{p_1, \ldots, p_n\}$ is a finite set of properties.*
   - *$>$ is a relation on $\mathbb{P}$ that is irreflexive and transitive.*
3. *$M = \{m_1, m_2 \ldots m_n\}$ is a finite set of **messages**.*
4. *C is a measure function on M describing the **cost** of each message.*
5. *$[\cdot]$ is the **indexation** relation (to be described below).*
6. *μ_S is a measure function on $\mathbb{P}$ describing S's **values** in the context.*
7. *μ_R is a measure function on $\mathbb{P}$ describing R's **values** in the context.*
8. *V_S is a function from pairs of sets of properties to real numbers describing how **useful** S finds R's interpretation.*
9. *V_R is a function from pairs of sets of properties to real numbers describing how **useful** R finds S's type.*

10. *ρ is a receiver/listener strategy, representing **S' beliefs** about how R will interpret their message.*

11. *σ is a sender/speaker strategy, representing **R's beliefs** about why S would choose their message.*

12. *Pr is a probability distribution over sets of properties describing R's **prior beliefs** about S.*

As shown above, the basic domain of interpretation is $\mathbb{P}$, a set of properties. In this paper, we will have the relation $>$ encode incompatibility between properties; that is, $P_1 > P_2$ just in case P_1 and P_2 are contraries: they cannot both be true of an individual at the same time. This will be the extent of the structure that we will impose on the universe; however, in future extensions of the model it may be desirable to enrich the universe with scales, antonymy relations or other more complicated structures.

As a concrete example, let us consider a universe specified as shown in (13), where it is impossible to be both competent and incompetent at the same time, and it is impossible to be both friendly (where we should understand *friendliness* as also regrouping properties such solidarity and authenticity) and aloof (where we should understand *aloofness* as regrouping properties such as pretension, exclusion and snobbishness).

(13) $\mathbb{P} = \{\text{competent, incompetent, friendly, aloof}\}$

a. competent $>$ incompetent

b. friendly $>$ aloof

In addition to the attribution of individual properties and the meaning of individual variables, Third Wave variation theory also focuses on how those variables combine together into *styles*, which are both related to and construct particular social types called *personae* (see Podesva, 2004; Eckert, 2008; Zhang, 2008, among many others). In this paper, we take personae to be particular collections of properties that 'go together'. Thus, the set of possible personae are all the subsets of the universe that do not contain incompatible properties, as shown in Def. 4.2.

Definition 4.2. *P is possible **persona** ($P \in$ PERS) iff*

(14) $P \subseteq \mathbb{P}$ *and there are no* $p_1, p_2 \in P$ *such that* $p_1 > p_2$.

In our simple example, then, by Def. 4.2, the possible personae in the universe in (13) are shown in (15): we have the set {competent, friendly}, (what we might think of as) the 'cool guy' type; {competent, aloof}, the 'stern leader' type; {incompetent, friendly}, the 'doofus' type; and {incompetent, aloof}, the 'arrogant asshole' type. And we also have all the singleton personae, whose role in the rest of the paper will be limited.

(15) PERS = {{competent, friendly}, {competent, aloof}, {incompetent, friendly}, {incompetent, aloof}, {competent}, {friendly}, {incompetent}, {aloof}}

As mentioned above, S's type (in the game-theoretic sense), i.e. what they are trying to communicate to R, is a set of properties: a persona $P \in$ PERS. In classic signalling games, S's type is generally conceived of as a context-independent piece of information about the world that they wish to transmit to R. In SMGs, however, things are a bit different: S's type will depend on what their values and goals are at the time, i.e. which persona they think will be most useful to them in the context. Likewise, R's action will be to assign a persona to S based on their use of a message m, and, in doing so, update their beliefs concerning the properties that characterize the speaker.

As in classic signalling games, we have a set of messages which come with a set of costs.

(16) **Messages and Costs.**

a. $M = \{m_1, \ldots, m_n\}$ is the set of messages (i.e. variants) that S can pick from.

b. C is a function from M to the real numbers that assigns a cost to each message.

How should we interpret the costs associated with variants? One idea might be to identify the cost of a message with the comfort or ease (or lack thereof) that a speaker has with manipulating it. For example, if m is a prestige or standard form which requires a certain amount of exposure/engagement with educational institutions in order to manipulate properly, then, for speakers who have not had such exposure, m would be more costly to use than a more vernacular message m' (see Bourdieu and Boltanski, 1975; Bourdieu, 1980). Parallely, if m is a highly vernacular form that the speaker is not familiar with or does not form part of the speaker's 'native' dialect (as in cases of *language crossing* (Rampton, 1995; Bucholtz, 1999, 2010, among others), the same principle may apply. Since both variants of (ING) are used by members of all educational levels (Hazen, 2006) and the use of *-in'* is not particularly stigmatized[21], we will assume for our example that there are no differences in the cost of using *in'* than in the cost of using *-ing*. A game in which there are no non-trivial costs associated with messages is standardly called a *cheap talk* game.

As mentioned in section 3, in the Third Wave, individual variants have meaning that goes beyond their truth conditional meaning. More precisely, variants are proposed to index sets of properties, called their *indexical field* (Eckert, 2008). In SMGs, messages are proposed to be related to their field via the *indexation* relation, as shown in (17).

(17) Indexation relation ($[\cdot]$) and indexical fields ($\mathcal{F}$)

a. For all messages $m \in M$, $[m] \subseteq \mathbb{P}$.

b. $\mathcal{F}_{m_i}$ notates the indexical field associated with m_i, i.e. $[\mathrm{m}_i] = \mathcal{F}_{m_i}$.

[21] That is, (ING) is a *marker* rather than a *stereotype* in the sense of Labov (1966).

Much current work within TW is devoted to studying the structure of indexical fields, investigating whether there are meaningful subclasses of properties within a field (Silverstein, 2003) or whether there are meaningful relations between the properties that make up a variant's field (Eckert, 2008). In this paper, I will keep things as simple as possible and not impose any structure on the $\mathcal{F}_{m_i}$s, but, again, the structure of the fields could easily be enriched, should we find empirical arguments in favour of doing so.

In today's example, following (simplified) Eckert (2008), we will assume that the variants of (ING) are associated with the fields shown in (18).

(18) Indexical fields associated with (ING)
   a. [-ing] = {competent, aloof}
   b. [-in'] = {incompetent, friendly}

One important difference between a classical signalling game and an SMG is the precise way in which utility is calculated. In particular, as shown in Def. 4.1, in each game/context, each player comes with their set of values, expressed by the $\mu_{S/R}$ functions; that is, we propose that $\mu_{S/R}$ is a measure function that assigns to individual properties a real number describing how desirable S or R find them in the context. More specifically, μ_S describes how useful S thinks having those properties attributed to them will be to achieving their goals (whatever they may be). μ_R describes how useful interacting with an interlocutor that possesses these properties will be to R in achieving their goals.

To give a concrete illustrative example, consider the case discussed in the previous section of Obama at the barbecue. Although this is not necessary, suppose for the example that Obama and his interlocutors at the barbecue have the same values, namely they like people who are competent and friendly[22]. So, in this situation, a possible μ that reflects these values is shown in (19).

(19)
$$\begin{array}{ll} \mu_{S/R}(\text{competent}) = 2 & \mu_{S/R}(\text{friendly}) = 2 \\ \mu_{S/R}(\text{incompetent}) = 1 & \mu_{S/R}(\text{aloof}) = 1 \end{array}$$

Since we are interested not only in agents' values for particular properties, but also in their values for particular personae, we extend μ_S and μ_R to measure functions on personae in the natural way. Namely, we say that the value of a particular persona P is the average of the values of the individual properties making up P, as shown in (20).

(20) For all $P \in PERS$, $\mu^*_{S/R}(P) = \frac{\sum_{p \in P} \mu_{S/R}(p)}{|P|}$

So, in our 'Obama at the barbecue' example, the value that S and R assign to the 'cool guy'

[22]For example, in this context, Obama wants to show that he is a friendly, down to earth guy; however, he still wants people to think that he is competent. He is the president, after all...

persona ({competent, friendly}) is 2; whereas, the value that they assign to the 'arrogant asshole' is 1. With these measure functions, the 'doofus' persona is assigned the value 1.5. Note also that, since we have kept the singleton personae in the model, {competent} is also assigned 2, as shown in (21).

(21) a. $\mu^*_{S/R}(\{\text{competent, friendly}\}) = (2+2)/2 = 2$
b. $\mu^*_{S/R}(\{\text{incompetent, aloof}\}) = (1+1)/2 = 1$
c. $\mu^*_{S/R}(\{\text{incompetent, friendly}\}) = (1+2)/2 = 1.5$
d. $\mu^*_{S/R}(\{\text{competent}\}) = 2/1 = 2$

Although nothing in the framework forces this, we will make the natural assumption in the examples in this paper that S is always trying to perform the persona that lies within the set of personae that they most highly value. Therefore, let's suppose for the Obama barbecue example that Obama's type is {competent, friendly}.

S's utility is based on how desirable S finds the persona that R attributes to them, but, because sociolinguistic variation is a kind of communication, we propose that S's utility is also based on coordination with R, as in a classic signalling game. In order to do this, we will have S's utility determined by their values, but instances in which S's type and R's interpretation converge get a bonus.

Definition 4.3. ***Speaker Reward (V_S).*** *Let V_S be a reward function from possible types and actions to real numbers.*

(22) *For all $P, P' \in PERS$,* $V_S(P, P') = \begin{cases} n \times \mu^*_S(P') & \text{if } P = P' \\ n' \times \mu^*_S(P') & \text{otherwise} \end{cases}$ *where $n > n'$.*

The value for S of an event in which S wants to communicate a persona P and R interprets P' is determined by how happy S is about being assigned P', multiplied by a number n (if R's interpretation is what S wanted) or a smaller number n' (if R fails to 'get' what S is trying to tell them).

As in classic signalling games, the utility of a message is its rewards minus its costs.

Definition 4.4. ***Speaker Utility (U_S).*** *For $m \in M$, $U_S(P, m, P') = V_S(P, P') - C(m)$.*

Similarly to speaker reward and utility, listener utility is based on R's values and coordination with S, as should in Def. 4.5.

Definition 4.5. ***Receiver Reward (V_R) and Utility (U_R).***

1. *Let V_R be a reward function such that:*

(23) *For all* $P, P' \in PERS$, $V_R(P, P') = \begin{cases} n \times \mu^*_R(P) & \text{if } P = P' \\ n' \times \mu^*_R(P) & \text{otherwise} \end{cases}$ *where* $n > n'$.

2. *For* $m \in M$, $U_R(P, m, P') = V_R(P, P')$.

In our Obama barbecue example, suppose that the values for n and n' in the definitions of speaker utility above are 1.25 and 1, respectively. In this example then, we have the payoff structure shown in Table 6.

S/R	{com fr}	{com}	{fr}	{in fr}	{com alo}	{in alo}	{in}	{alo}
{com, fr}	2.5,2.5	2,2	2,2	1.5,2	1.5,2	1,2	1,2	1,2
{com}	2,2	2.5,2.5	2,2	1.5,2	1.5,2	1,2	1,2	1,2
{fr}	2,2	2,2	2.5,2.5	1.5,2	1.5,2	1,2	1,2	1,2
{in fr}	2,1.5	2,1.5	2,1.5	1.875,1.875	1.5,1.5	1,1.5	1,1.5	1,1.5
{com alo}	2,1.5	2,1.5	2,1.5	1.5,1.5	1.875,1.875	1,1.5	1,1.5	1,1.5
{in alo}	2,1	2,1	2,1	1.5,1	1.5,1	1.25,1.25	1,1	1,1
{in}	2,1	2, 1	2,1	1.5,1	1.5,1	1,1	1.25,1.25	1,1
{alo}	2,1	2, 1	2,1	1.5,1	1.5,1	1,1	1,1	1.25,1.25

Table 6 – Payoff structure of 'Obama at the barbecue' SMG (values aligned)

As shown in Table 6, the classic Nash equilibria are associated with coordination.

Note that in SMGs, equilibria may change if either (or both) of the agents' reward or value functions change. For example, suppose that instead of S and R being aligned in their values, S assigns the highest value to the persona {competent, friendly} (and its sub-personae {competent} and {friendly}); whereas, R assigns the highest value to the {competent, aloof} persona and its pertinent sub-persona[23]. As shown in Table 7, there is no Nash equilibrium involving {incomptent, friendly} because this persona is so devalued by R that not even coordination can save it. This being said, I leave a full study of the solution concepts (classic equilibria or otherwise) for these games and their implications for stability and change at the level of social meaning to future research.

Speaker strategies will be described using functions from personae (possible types) to probability distributions over messages being used to signal those types. Listener/receiver strategies will be described using functions from messages to probability distributions over personae. One of the main contributions of this work is to give an explanation of how the use of particular linguistic variants can help change the listener's beliefs about the speaker. As such, in the model, there has to be some way of having the speaker's choice

[23]Unlike in the United States or Canada, where it is often important for a leader to appear friendly and down to earth in informal contexts, this is not the case in some other cultures. For example, French president François Hollande has frequently been criticized for acting too goofy when he is outside l'Elysée, and this lack of gravitas has been viewed by the press and others as unbecoming for a French president. Thus, we might think of the example in Table 7 as possibly depicting a game between an American speaker and a French listener.

S/R	{com fr}	{com}	{fr}	{in fr}	{com alo}	{in alo}	{in}	{alo}
{com, fr}	2.5, 1.875	2,1.5	2,1.5	1.5,1.5	1.5,1.5	1,1.5	1,1.5	1,1.5
{com}	2, 2	2.5,2.5	1.5,2	1.5,2	1.5,2	1,2	1,2	1,2
{fr}	2, 1	2,1	2.5,1.25	1.5,1	1.5,1	1,1	1,1	1,1
{in fr}	2,1	2,2	2,1	1.875,1.25	1.5,2	1,1.5	1,1	1,2
{com alo}	2,1.5	2,1.5	2,1.5	1.5,1.5	1.875,2.5	1,1.5	1,1.5	1,1.5
{in alo}	2,1	2,1	2,1	1.5,1	1.5,1	1.25,1.25	1,1	1,1
{in}	2,1	2, 1	2,1	1.5,1	1.5,1	1,1	1.25,1.25	1,1
{alo}	2,2	2, 2	2,2	1.5,2	1.5,2	1,2	1,2	1.25,2.5

Table 7 – Payoff structure of 'Obama at the barbecue' SMG (values at odds)

of message and indexical field (in our example, *-in'* vs *-ing*) have at least some effect on listener beliefs[24]. In order to do this, we will impose a simple constraint on speaker/listener strategies: that they be *existentially anchored* to indexical fields. This is shown in Defs. 4.6 and 4.7.

Definition 4.6. ***Speaker strategies (σs)*** *are functions from the set of possible personae (*PERS*) to probability distributions over M satisfying the following condition:*

(24) ***Indexical Anchoring (S):***
For all $P \in PERS$ and $m \in M$, if $\sigma(P, m) \neq 0$, then there is some $p \in P$ such that $p \in \mathcal{F}_m$.

In English, what (45) says is the following: Suppose that S's type is a persona P. Then, if they have a non-zero probability of selecting a message m, this is because there is at least one property p in the persona that they are trying to communicate that is indexed by m. In other words, we stipulate that S can't just completely randomly pick any message and expect it to do the work of communicating their persona.

Listener strategies will be similarly anchored to indexical fields, as shown in Def. 4.7.

Definition 4.7. ***Listener Strategies (ρs)*** *are functions from M to probability distributions over PERS satisfying the following condition:*

(25) ***Indexical Anchoring (R):***
For all $m \in M$ and $P \in PERS$, if $\rho(m, P) \neq 0$, then there is some $p \in P$ such that $p \in \mathcal{F}_m$.

What (44) says is that if R hears a message m and they have a non-zero probability of interpreting m as a persona P, then this is because there is at least one property p in this

[24]Note that this issue is familiar from the use of signalling games with meaningful messages to analyze other pragmatic phenomena such as scalar implicatures. See (Franke, 2009, section 1.2.4.) for a discussion.

persona that is indexed by m.

4.2 Speaker Beliefs and Expected Utility

As stated in Def. 4.1, we will use speaker and listener strategies to represent the beliefs of S and R, as is common in Bayesian game-theoretic approaches. Thus, we will represent S's beliefs concerning how R will interpret their use of variants as an indexically anchored listener strategy ρ, i.e. a probability distribution over PERS for each $m \in M$.

To continue with our Obama example: suppose that at the barbecue Obama holds the beliefs depicted in Table 8 concerning how he will be interpreted if he uses *-in'* or *-ing*.

	ρ(ing)	ρ(in')
{competent, friendly}	0.2	0.6
{competent, aloof}	0.4	0
{incompetent, friendly}	0	0.2
{incompetent, aloof}	0.4	0.2
Other P	0	0

Table 8 – Listener strategy ρ representing Obama's beliefs at the barbecue

The way to read the probability distribution in Table 8 is that Obama thinks that there is decent chance that he will be considered competent no matter what he says. He also suspects that his interlocutors will interpret his use of *ing* as aloof/pretentious, but *in'* as friendly. By virtue of indexical anchoring, he thinks that it is impossible that his interlocutor interprets *-ing* as signalling the 'doofus' persona ({incompetent, friendly}), since neither *incompetent* nor *friendly* lie in *-ing*'s indexical field. Likewise, he thinks that the probability of R interpreting *-in'* as the 'stern leader' type ({competent, aloof}) is zero, since neither competence nor aloofness/pretension are indexed by *-in'*.

We define speaker expected utility in the standard way (see Osborne and Rubinstein, 1994): S's expected utility for a variant m, given that their type is P and that they have the beliefs ρ, is the sum all S's utilities for all the possible interpretations of m multiplied by how likely they think that R will make those interpretations.

Definition 4.8. ***Speaker Expected Utility*** *(EU_S). Let m be a message, $P \in$ PERS and let ρ be an indexically anchored listener strategy. The,*

$$EU_S(m, P, \rho) = \sum_{P' \in \mathbb{A}} \rho(m, P') \times U_S(P, m, P')$$

In the 'Obama at the barbecue' example, his expected utility for *-ing* is 1.5; whereas, his expected utility for *-in'* is 2, as shown in (26). Since Obama's expected utility for *-in'* is

higher than *-ing* in this context, (under the hypothesis that he is rational), we predict that he will use *-in'*, which indeed was his highly favoured variant in this context in Labov's study.

(26) Obama's expected utility for (ING) at the barbecue
  a. EU_S(-ing, {comp, friend}, ρ) = (0.2*2.5)+(0.4*1.5)+(0.4*1) = 1.5.
  b. EU_S(-in', {comp, friend}, ρ) = (0.6*2.5)+(0.2*1.5)+(0.2*1) = 2.

4.2.1 Predictions of the model for speaker behaviour

How does the model laid out so far explain the patterns of intra-speaker behaviour described in section 3?

The first thing to observe is that the system as defined up to this point makes categorical predictions for linguistic variation. That is, by virtue of sharing most of the properties of classic signalling systems, the system picks only one winner out of set of messages: the variant with the highest expected utility. Of course, sociolinguistic production data is gradient, with Obama's use of *-in'* being at 72% at its highest observed rate, dropping to 33% and then to 3%. How should we understand the relationship between the expected utilities derived above and the observation of gradience in language use within a single context?

As mentioned in the introduction, SMGs aim only to model the **social** or **strategic** aspect of linguistic variation. When we go to speak, which form we end up picking depends on a wide range of factors, only a subset of which depend on social meaning and persona construction. For example, in addition to social (or what Labov calls *external*) factors, physiological or psycholinguistic factors such as ease of articulation, frequency, priming or other processing factors may play a role in favouring the use of one variant over another. Likewise, grammatical factors of the type discussed in section 2, which Labov calls *internal* factors, may induce a bias in favour of one variant over another. This being said, as shown in most of the studies discussed in sections 2 and 3, even if we control for the contribution of psycholinguistic and grammatical factors, it may be the case that there is still *inherent variability*; that is, speakers still alternate between the different options available to them.

To account for possible inherent variability, we can weaken the rule that we use to pick the form that S will use such that, rather than just picking the variant with the highest expected utility, S chooses the best option given a noise-perturbed assessment of expected utilities. One such weaker choice rule, called the *Soft-Max Choice Rule* (Luce, 1959; Sutton and Barto, 1998), is widely used in both reinforcement learning and in Bayesian approaches to a variety of pragmatic phenomena (Frank and Goodman, 2012; Degen et al., 2013; Lassiter and Goodman, 2015; Franke and Jäger, 2016, among others). For example, in their accounts of both vague adjectives and scalar implicatures, (Lassiter and Goodman, 2015, 9) "employ a relaxed version of this model according to which agents choose stochastically, i.e.,

that speakers sample actions with the probability of making a choice increasing monotonically with its utility... Apparently sub-optimal choice rules of this type have considerable psychological motivation. They can also be rationalized in terms of optimal behavior for an agent whose computational abilities are bounded by time and resource constraints, but who can efficiently approximate optimal choices by sampling from a probability distribution".

Set in the SMG framework, the Soft-Max choice rule looks as shown in (27), where $p_S(m_i|P,\rho)$ notates the probability of S using m_1, given that their type is P and they have the beliefs ρ about how R will interpret their linguistic offering. The constant β is called the *temperature*, and represents how much indeterminacy the model allows.

(27) $p(m_1|P,\rho) = \frac{\exp(\beta \times EU_S(m_1,P,\rho))}{\sum_i \exp(\beta \times EU_S(m_i,P,\rho))}$

In the Obama example, if we set the temperature to 2, then we predict (roughly) the observed distribution, as shown in (28).

(28) a. p_S(-in'|{competent,friendly},ρ)=exp(2*2)/(exp(2*2) + exp(2*1.5)) = 0.731
b. p_S(-ing|{competent,friendly},ρ)=exp(2*1.5)/(exp(2*2) + exp(2*1.5)) = 0.268

In what follows, I will review the empirical studies discussed section 3 and provide a sketch of how they could be modelled within the SMG framework. Obviously each study is complex and in order to do the modelling, in the absence of concrete information concerning speaker/listener beliefs, we must make conjectures that are as reasonable as possible about agents' beliefs and interpretation strategies. Given that we are working from empirical studies that have not explicitly investigated questions related to speaker/listener beliefs, this task is extremely difficult, if not impossible. As such, the aim in this section is to give a broad illustration of kinds of qualitative and quantitative predictions for language use made by the SMG framework using examples from the variationist sociolinguistics literature.

The previous subsection showed how we could calculate expected utilites for variants based on speaker beliefs and utilities; however, the model also predicts that, should speakers' beliefs about how their interlocutors will interpret their message change, their expected utilities, and therefore their linguistic use, should change as well. For example, suppose that after the barbecue, when Obama moves to take political questions from (often hostile) reporters, he is no longer as confident that he will be attributed a competent persona no matter which variant he uses. Rather, he is worried that his use of *-in'* may be interpreted as a sign of incompetence, as shown in Table 9, where the bulk of the probability mass associated with the use of *-in'* is now assigned to incompetent personae.

In this context, Obama's expected utilities for (ING) will change, with *-ing* now having a higher expected utility than *-in'* (29). Using the soft-max choice rule (with the temperature still set at 2), we predict that Obama should use *-ing* aroung 64.6 % of the time, which is very close to what Labov observed.

	ρ'(ing)	ρ'(in')
{competent, friendly}	0.4	0.2
{competent, aloof}	0.4	0
{incompetent, friendly}	0	0.4
{incompetent, aloof}	0.2	0.4
Other P	0	0

Table 9 – Listener strategy ρ' representing Obama's beliefs after the barbecue

(29) Obama's expected utility for (ING) after the barbecue

a. EU_S(-ing, {comp, friend}, ρ') = (0.4*2.5)+(0.4*1.5)+(0.2*1) = 1.8.

b. EU_S(-in', {comp, friend}, ρ') = (0.2*2.5)+(0.4*1.5)+(0.4*1) = 1.5.

Using these results associated with changes in speaker beliefs, we can also capture the pattern of Gratton (2015)'s non-binary participants in the home/queer cafe versus public cafe setting (section 3). Following Gratton's remarks, we might suppose that, in the queer-friendly, familiar environments, Flynn and Casey are confident that they will be attributed a persona consistent with their gender identity; however, as Gratton proposes, in the public environments, her consultants are worried about being mis-gendered by the strangers surrounding them, with Flynn concerned that they will be attributed a feminine persona and Casey concerned that they will be attributed a masculine persona. For this example, we need to extend our analysis (still along the lines of proposals by Eckert and others) of the indexical fields associated with the variants of (ING) as shown in (30).

(30) Indexical fields associated with (ING)

a. [-ing] = {competent, aloof, feminine}

b. [-in'] = {incompetent, friendly, masculine}

Flynn identifies as non-binary; that is, neither male nor female. Therefore, we might suppose that they would assign the lowest values to both of the properties *masculine* and *feminine*; for example, suppose that μ_{Flynn}(masculine) = μ_{Flynn}(feminine) = 1. Thus, Flynn's type (in this extended model) must not contain either *masculine* nor *feminine*. Obviously, a fair amount of ethnographic work is needed to really understand which persona Flynn was trying to perform when they were recorded[25]; however, in this example we can suppose for continuity and without loss of generality that Flynn wishes to communicate {competent, friendly}[26]. When they are interviewed in their home, Flynn is not worried about being misgendered[27]. We can represent this belief through the listener strategy ρ_H in which all the

[25]The same remark applies to Obama in and around the barbecue.

[26]I say 'without loss of generality' because we predict the same qualitative pattern in this game if other properties (other than masculine and feminine) are highly valued. What is important for the example is that feminine is highly devalued.

[27]Gratton concludes from her interviews that "in queer situations: no conscious presenting is needed."

two-property personae containing neither feminine nor masculine are weighed higher than any persona containing feminine/masculine, as shown in Table 10. Note that the 'special' belief that we attribute to Flynn here concerns primarily how they think that their *-ing* will be interpreted, i.e that it will not be interpreted in a 'feminine' way at home, because they know their family/friends will not view them as feminine. Obviously, since *-in'* does not index the property *feminine*, there is no particular risk of *-in'* being interpreted as feminine.

In the public cafe, on the other hand, Flynn is worried that they will be attributed a feminine persona, particularly if they use *-ing* which indexes femininity[28]. This belief is represented by the strategy ρ_C, where, unlike in ρ_H, the probability of attributing a feminine persona after hearing *-ing* is higher than a non-feminine persona.

	ρ_H(ing)	ρ_H(in')	ρ_C(ing)	ρ_C(in')
{competent, friendly}	0.2	0.2	0.053	0.2
{competent, aloof}	0.2	0.2	0.053	0.2
{incompetent, friendly}	0	0.2	0	0.2
{incompetent, aloof}	0.2	0.2	0.053	0.2
{competent, friendly, feminine}	0.1	0.067	0.21	0.067
{competent, aloof, feminine}	0.1	0	0.21	0
{incompetent, friendly, feminine}	0.1	0.067	0.21	0.067
{incompetent, aloof, feminine}	0.1	0.067	0.21	0.067
Other P	0	0	0	0

Table 10 – Strategies ρ_H, ρ_C representing Flynn's beliefs at home vs in the cafe, respectively.

Our model predicts an important difference in Flynn's use of (ING) between the home and the cafe: as shown in (31), at home we predict that Flynn's expected utilities for *-in'* and *-ing* should be almost identical, and that they should use *-in'* 52% of the time, which is close to what Gratton observed.

(31) a. EU_S(-ing, {comp, friend}, ρ_H) =
$= (2.5*0.2)+(1.5*0.2)+(1*0.2)+(1.67*0.1)+(1.33*0.1)+(1.33*0.1)+(1*0.1)$
$= 1.533$.

b. EU_S(-in', {comp, friend}, ρ_H) =
$= (2.5*0.2)+(1.5*0.2)+(1.5*0.2)+(1*0.2)+(1.67*0.067)+(1.33*0.067)+(1*0.067)$
$= 1.568$.

In the cafe, where Flynn is worried that *-ing* will influence their interlocutors into attributing them a feminine persona, *-ing*'s expected utility is much lower (32), and we predict

[28]Gratton says that "individuals whom they encounter will presuppose a binary gender based mainly on physiological characteristics, so speakers must distance themselves from the gender which is presumed - their gender-assigned-at-birth - by utilizing resources which resist that gender's norms.

that *in'* should be used 59% of the time, which is much higher than what we predict for the queer-friendly environment.

(32) EU_S(-ing, {comp, friend}, ρ_C) =
= (2.5*0.053)+(1.5*0.053)+(1*0.053)+(1.67*0.21)+(1.33*0.21)+(1.33*0.21)+(1*0.21)
= 1.384.

Finally, of course, speakers' values can change across contexts, namely, they can wish to perform different personae in different situations. For example, suppose that, when he is giving a speech at the Democratic National Convention, Obama does not wish to come off as a 'cool guy'; rather, he wishes to adopt the 'stern leader' persona and be perceived as competent and aloof. Thus, in this game/context, we might suppose that Obama's type is {competent, aloof} rather than {competent, friendly}. Furthermore, as mentioned in the Labov quotation in section 3, in such a context, there is a heightened risk of listeners interpreting *-in'* as a sign of incompetence. Therefore, we might suppose that Obama would have the beliefs more along the lines of ρ' than ρ. In this context, because *-in'* doesn't index any properties in the persona that Obama is trying to perform, we predict that this variant is useless for him in this context. Indeed, Labov found that only 3% of cases in Obama's speech were realized as *-in'*.

Theorem 4.1. *EU_S(-in', {competent, aloof}, ρ') = 0*[29].

Following proposals in TW, we also account for inter-speaker variation by proposing that, even in a single context such as a sociolinguistic interview, different speakers can have different types, i.e. different personae that they are trying to perform. Given this, we would of course like to know how and why different speakers come to have different ideas about which personae will be most useful to them in achieving their goals; however, such an explanation lies at the level of social theory[30], not linguistic theory, and therefore is out of the scope of this work. Once the values and types for different speakers are given, the account of social stratification in SMGs is straightforward.

4.3 Listener Beliefs and Expected Utility

As is common in probabilistic pragmatics, listener beliefs will be represented in three parts:

1. **Pr** $\in \Delta(\mathcal{P}(\mathbb{P}))$ is a probability distribution representing R's (prior) uncertainty about which properties hold of S, i.e. their hypotheses concerning what kind of person S is **before** they hear S say the message.

[29] Proof: By indexical anchoring, ρ'({competent, aloof}, -in') = 0, so EU_S(-in', {competent, aloof}, ρ') = 0. □

[30] See for example Bourdieu (1980) for such a theory.

2. σ is R's hypothesis concerning the (indexically anchored) strategy that S is using, i.e. R's beliefs concerning what kind of person would have used m.

3. π is a receiver strategy modelling R's posterior beliefs, i.e. their updated beliefs about S after they hear the message.

R's posterior beliefs (π) about which persona corresponds to S are **derived** from Pr and σ via *Bayesian update/conditionalization*. In other words, R's perception of S, given that they have heard a message m, depends on 1) the persona(e) that R already thinks S is, and 2) R's beliefs about what kind of persona(e) would use m. The update rule is given more formally in (33).

(33) Let $P \in \text{PERS}$ and let $m \in M$,

$$\pi(P|m) = \frac{Pr(P) \times \sigma(m|P)}{\sum_{P' \in PERS} Pr(P') \times \sigma(m|P')}$$

Observe that the property of *indexical anchoring* is preserved under Bayesian update, as shown in Theorem 4.2.

Theorem 4.2. ***Indexical anchoring of posterior beliefs***[31]. *Let σ be a speaker strategy that is indexically anchored to a set of messages M with associated indexical fields. Let Pr be prior beliefs, and let π be the receiver strategy derived via Bayesian conditionalization from σ and Pr. Then,*

(34) π is also indexically anchored to M.

For example, going back to the Obama at the barbecue example, consider the interpretation process of one of Obama's interlocutors: R. Suppose R has the same values as Obama in the example (they value competence and friendliness in this context), and furthermore suppose R has the beliefs about Obama shown in (35) before hearing him talk. In English, we can describe the beliefs in (35) as the state of R thinking that Obama is most likely a 'stern leader' type, but they are also entertaining the idea that he may be a 'cool guy' or an 'arrogant asshole', and R thinks that it is possible, but highly unlikely, that Obama is a 'doofus'.

(35)
1. Pr({competent, aloof}) = 0.4 (probably)
2. Pr({competent, friendly}) = 0.3 (maybe)
3. Pr({incompetent, aloof}) = 0.2 (maybe not)

[31] PROOF: Let $P \in \text{PERS}$ and $m \in M$ such that $\pi(P|m) \neq 0$ to show that there is some $p \in P$ such that $p \in [m]$. Suppose for a contradiction that there is no such $p \in P$. Then, by indexical anchoring (45), $\sigma(m|P) = 0$ and so, by the definition of Baysian conditionalization (33), $\pi(P|m) = 0$. $\perp$ So there is some $p \in P$ such that $p \in [m]$. $\square$

4.Pr({incompetent, friendly}) = 0.1 (probably not)
5.For all other $P \in$ PERS, Pr(P) = 0. (no way)

Furthermore, in this context, suppose that R has the beliefs σ, shown in Table 11, about what kind of people use (ING) at a barbecue. In particular, by indexical anchoring, R thinks it's impossible that the stern leader would use *-in'*, since none of this persona's properties are indexed by *-in'*, and they also think that it's impossible that a doofus would use *-ing*, since none of this persona's properties are indexed by this variant. However, R thinks that a cool guy is slightly more likely to use *-in'* than *-ing*; whereas, an arrogant asshole is slightly more likely to use *-ing* than *-in'*. For space reasons, the rest of the strategy σ is not depicted in Table 11, but the appropriate probabilities for the singleton personae can be straightforwardly derived from the requirement that σ map a persona to a probability distribution over M and indexical anchoring.

	σ({com, alo})	σ({com, fr})	σ({in, alo})	σ({in, fr})	other
ing	1	0.4	0.6	0	...
in'	0	0.6	0.4	1	...

Table 11 – Listener beliefs about Obama's strategy σ at the barbecue

Based on (35) and Table 11, we can calculate R's posterior beliefs π, i.e. what they believe about Obama **after** they hear what variant he uses. This is shown in Table 12.

P	π(P|-in')	π(P|-ing)
{competent, friendly}	0.5	0.188
{competent, aloof}	0	0.625
{incompetent, friendly}	0.277	0
{incompetent, aloof}	0.222	0.188
Other	0	0

Table 12 – Listener posterior beliefs at the barbecue

As shown in Table 12, hearing one variant versus another can have an important effect on R's beliefs about Obama. While R's prior belief was that Obama had a 30% chance of being {competent, friendly} and a 40% chance of being {competent, aloof} (35), after they hear *-in'*, they no longer think that {competent, aloof} is a possibility, and now they think that there is a much higher chance (50%) that Obama is {competent, friendly}. Likewise, if they hear Obama say *-ing*, R will update their beliefs such that the probability that he is {competent, aloof} will rise to 0.625 and the probability that R is {competent, friendly} lowers to 0.188.

Listener expected utility is calculated in a parallel manner to speaker expected utility: the listener expected utility (EU_R) of a message m' given that S's type is P and that R's posterior beliefs are ϕ is the sum of the utility for each interpretation of the message times

the interpretation's probability according to π. Thus, R's expected utility for the different ways of interpreting the variants of (ING) in the Obama barbecue example are shown in Table 13.

	EU_R(-in', π)	EU_R(-ing, π)
{competent, friendly}	1.86	1.596
{competent, aloof}	1.61	1.736
{incompetent, friendly}	1.693	1.501
{incompetent, aloof}	1.679	1.548
{competent}	1.61	1.501
{incompetent}	1.61	1.501
{friendly}	1.61	1.501
{aloof}	1.61	1.501

Table 13 – Listener expected utilities for (ING) at the barbecue

Since R's expected utility for interpreting *-in'* as {competent, friendly} is the highest, if they are rational they will assign this persona to Obama in this context[32]. Likewise, since the expected utility for {competent, aloof} is the highest of any interpretation after hearing *-ing*, we predict enrichment to this persona.

Note that in this way, social enrichment is different from other kinds of enrichment such as scalar implicatures in which an utterance's literal meaning is a proposition that then gets enriched into another, stronger, proposition (36).

(36) Scalar enrichment takes propositions to propositions.

a. **Utterance:** Mary ate some of the cookies.
b. **Literal/non-enriched meaning:** Mary ate at least some cookies.
c. **Enriched meaning:** Mary ate some but not all of the cookies.

In SMGs, the 'literal'/non-enriched/context-independent meaning of variant is an indexical field, and then the field gets enriched and pruned into a persona through the process described above (37).

(37) Social enrichment takes indexical fields to personae.

a. **Utterance:** *-in'*
b. **Literal/non-enriched meaning:** {incompetent, friendly}
c. **Enriched meaning:** {competent, friendly} (for example)

[32]Using the soft-max choice rule (with temperature 2), this model predicts that *-in'* should be interpreted as {competent, friendly} 18% of the time, with interpretations of {incompetent, friendly} and {incompetent, aloof} 13% of the time for each persona. It is hard to know whether these are correct quantitative predictions, since we don't have the appropriate data to test this on. However, in principle, the quantitative predictions of the SMG models for interpretation could be investigated experimentally in the future.

4.3.1 Predictions of the model for listener behaviour

How can we use the SMG model to understand the interpretative patterns discussed in section 2 and, in particular, the *listener orientation* property that I argued characterized social meaning?

One observation from section 2 is that the same variant can receive different interpretations in the mouths of different speakers. This was the case in the Podesva et al. (2015) study of /t/ release where John Edwards and Condoleeza Rice were perceived as more articulate with released /t/, but flapping did nothing to increase their friendliness. On the other hand, Nancy Pelosi was perceived as friendlier when she did not release her /t/, but, as discussed in section 2, she came off not as articulate but as pretentious when she release /t/.

In order to show how such patterns could be captured, let us consider an SMG such as in (38), where the universe is composed of four properties: {articulate, inarticulate, pretentious, friendly}, and the indexical fields for /t/ release, based on (simplified) proposals by Podesva (2007); Eckert (2008), are shown in (38-c) and (38-d).

(38) SMG for Podesva et al. (2015)

a. $\mathbb{P} = \{\text{articulate, inarticulate, pretentious, friendly}\}$
b. articulate > inarticulate; pretentious > friendly
c. $\mathcal{F}_{t^h} = \{\text{articulate, pretentious}\}$
d. $\mathcal{F}_{\text{ɾ}} = \{\text{inarticulate, friendly}\}$

In accordance with previous work, let's assume that a listener in Podesva et al.'s study might have the beliefs concerning the various politicians' language use σ shown in Table 14, where (by indexical anchoring) articulate and pretentious individuals never use the flap and inarticulate friendly people never release their /t/s. Suppose for simplicity that they have no great opinions about articulate/friendly people or inarticulate/pretentious people.

	$\sigma(\{\text{art, pret}\})$	$\sigma(\{\text{art, fr}\})$	$\sigma(\{\text{in, pret}\})$	$\sigma(\{\text{in, fr}\})$	other
t^h	1	0.5	0.5	0	...
ɾ	0	0.5	0.5	1	...

Table 14 – Listener beliefs about politicians strategies σ

Although there are probably multiple possible analyses for these data, one way of accounting for the difference between listeners' interpretations of Edwards/Rice and Pelosi lie in differences in their prior beliefs about these individuals. Podesva et al.'s participants had more polarized opinions about Edwards than Rice's properties (Podesva et al., 2015, 78), so I will limit this illustration to comparing differing interpretations between Edwards and Pelosi. For example, according to Podesva's speakers, Edwards is judged (in isolation) to be one of the least articulate and intelligent politicians, yet one of the most friendly speakers.

Pelosi, on the other hand, is judged as one of the most articulate and intelligent speakers, and she is not judged as friendly as Edwards. Thus, suppose that listeners' prior beliefs concerning Edwards and Pelosi are modelled as in (39) and (40).

(39) Prior beliefs about Edwards
1.Pr({articulate, pretentious}) = 0.1
2.Pr({articulate, friendly}) = 0.4
3.Pr({inarticulate, pretentious}) = 0.1
4.Pr({inarticulate, friendly}) = 0.4
5.For all other $P \in$ PERS, Pr(P) = 0.

(40) Prior beliefs about Pelosi
1.Pr({articulate, pretentious}) = 0. 4
2.Pr({articulate, friendly}) = 0.4
3.Pr({inarticulate, pretentious}) = 0.1
4.Pr({inarticulate, friendly}) = 0.1
5.For all other $P \in$ PERS, Pr(P) = 0.

In English, (39) can be read as saying that the listener is pretty sure that Edwards is not pretentious, but they are hesitating about whether to consider him articulate or not. Likewise, (40) can be read as saying that the listener is pretty sure that Pelosi is articulate, but they are hesitating about how friendly she is.

P	π_E(P|ɾ)	π_E(P|t^h)	π_P(P|ɾ)	π_P(P|t^h)
{articulate, friendly}	0.307	0.571	0.571	0.307
{articulate, pretentious}	0	0.286	0	0.615
{inarticulate, friendly}	0.615	0	0.286	0
{inarticulate, pretentious}	0.077	0.143	0.143	0.077
Other	0	0	0	0

Table 15 – Listener posterior beliefs after hearing Edwards (π_E)/Pelosi (π_P).

As shown in (39), before hearing Edwards say anything, R believes that the probability that he is articulate is 0.5 and the probability that he is friendly is 0.8. As shown in Table 15, if R hears a released /t/, then R updates to then take the probability that Edwards is articulate to be 0.857, and the probability that he is friendly to be 0.571. In other words, because of R's prior beliefs about Edwards, his linguistic use causes a larger change related to articulateness than to friendliness. Pelosi, on the other hand, shows exactly the mirror image pattern: R's prior beliefs gave her an 80% chance of being articulate and a 50% chance of being friendly. Using a flap moves her friendliness up to an 0.857, a larger change than her articulateness, which decreases to 0.571.

Furthermore, if we assume, similarly to the Obama example, that R values articulateness and friendliness in their American politicians, and that the utility functions in this game

are the same as in our (ING) games above, we predict that R should interpret Edwards' /t/ release differently than Pelosi's /t/ release. In particular, as shown in Table 16, R's highest expected utility for Edwards' flap is {inarticulate, friendly} and their highest expected utility for Edwards' released /t/ is {articulate, friendly}. Thus, we predict that interpretative variation between flap and released /t/ for Edwards should be at the level of articulateness; he is assigned a friendly persona using both variants. The highest expected utility for Pelosi's flap is {articulate, friendly} and R's highest EU for her released /t/ is {articulate, pretentious}. Therefore, we predict that the listener should find no difference in articulateness between Pelosi's flap and released /t/; however, we also correctly predict that flapping should make her sound more friendly, which was not the case for Edwards.

	$EU_R(ɾ, \pi_E)$	$EU_R(t^h, \pi_E)$	$EU_R(ɾ, \pi_P)$	$EU_R(t^h, \pi_P)$
{articulate, friendly}	1.767	1.999	1.999	1.767
{articulate, pretentious}	1.613	1.821	1.714	1.844
{inarticulate, friendly}	1.844	1.714	1.821	1.613
{inarticulate, pretentious}	1.633	1.749	1.749	1.633
{articulate}	1.613	1.714	1.714	1.613
{inarticulate}	1.613	1.714	1.714	1.613
{friendly}	1.613	1.714	1.714	1.613
{pretentious}	1.613	1.714	1.714	1.613

Table 16 – Listener expected utilities for /t/ for Edwards and Pelosi

We can extend this analysis to results associated with some of the other politicians in the study. In particular, although the use of released or flapped /t/ affected how some politicians in the study were judged, George W. Bush and Hilary Clinton were not judged to sound different in any of their guises. (Podesva et al., 2015, 80) attribute this result to particularly strong prior beliefs held by their participants concerning these two politicians, saying:

> We argue that listeners hold particularly strong views for these politicians, to the point that slight modifications in their speaking styles produce no effect on listener ratings. [...] Bush was rated as the least articulate, intelligent, authoritative, and sincere, often by a wide margin, and he was also rated as the second-to-least passionate and spontaneous. Focus group commentary was equally unfavorable. In a similar vein, Clinton's speech was evaluated by focus group participants as sounding, above else, clear-irrespective of the realization of /t/. In other words, extreme ideological stances taken toward speakers may generate floor or ceiling effects in the evaluation of their speech.

Such a pattern, where extreme prior beliefs about a speaker render the act of linguistic variation useless, is directly predicted by the SMG model. Suppose R has an extreme belief about George W. Bush, namely that he has the properties {inarticulate, pretentious}.

(41) Prior beliefs about W. Bush
1.Pr({articulate, pretentious}) = 0.05
2.Pr({articulate, friendly}) = 0.05
3.Pr({inarticulate, pretentious}) = 0.85
4.Pr({inarticulate, friendly}) = 0.05
5.For all other $P \in$ PERS, Pr(P) = 0.

R's posterior beliefs about W. Bush and the resulting expected utilities for interpretations of variants are shown in Table 17. Because R has such a strong belief in Bush's inarticulateness and pretension, hearing the flap and/or the released /t/ makes no serious impact on R's beliefs, and, furthermore, we predict exactly the same interpretation for flap and [t^h]: {inarticulate, pretentious}. Thus, Bush can try as much as he likes to perform a different persona, by virtue of his interlocutor's prior perception of him, he will be 'stuck' being viewed as {inarticulate, pretentious}.

	π_W(P|ɾ)	π_W(P|t^h)	EU_R(ɾ, π_P)	EU_R(t^h, π_P)
{articulate, friendly}	0.05	0.05	1.125	1.125
{articulate, pretentious}	0	0.1	1.1	1.138
{inarticulate, friendly}	0.1	0	1.138	1.1
{inarticulate, pretentious}	0.85	0.85	1.312	1.312
{articulate}	0	0	1.1	1.1
{inarticulate}	0	0	1.1	1.1
{friendly}	0	0	1.1	1.1
{pretentious}	0	0	1.1	1.1

Table 17 – Listener posterior beliefs and EU for /t/ for W. Bush

In sum, although the context-independent meanings of variants are fixed and shared across speakers, in the SMG framework, different prior listener beliefs can create different interpretations of those variants with different speakers in context; thus, this accounts for the listener orientation property discussed in section 2. An account similar to the one provided above for Podesva et al.'s /t/ release study could be given for Campbell-Kibler (2008)'s results concerning different speakers' different interpretations of Elizabeth's use of (ING): those participants who think that she is a 'natural' *-ing* user, i.e. that she instantiates personae that are constituted predominantly of properties (like aloofness) that are indexed by *-ing* rather than *-in'*, will use these beliefs to interpret Elizabeth's *-in'* as a condescending persona[33]; however, those whose prior beliefs are highly weighted on friendly personae will interpret her *-in'* as a sign of a compassionate persona. Furthermore, note that the framework also has a straightforward account of the phenomenon of *crossing* (Rampton, 1995) mentioned in section 4.1, in which members of one dialect group use a variant commonly associated with another dialect group, and this use is interpreted as only indicating

[33]Note that to do this example properly, we would need to extend the universe and the indexical fields appropriately with properties such as *condescending*. But the general line of analysis is clear.

a (usually desirable) subset of the (un)desirable properties usually attributed to members of the second group[34].

Given my account of listener orientation (as the result of the interaction between listener prior beliefs and the optimization process inherent to the SMG architecture), one might wonder how different my analysis of social enrichment actually is from other analyses of other kinds of pragmatic enrichment within the probabilistic pragmatics framework. Priors that are heavily weighted on certain objects and not others have been proposed to give rise to particular and sometimes strange kinds of interpretations associated with other kinds of pragmatic phenomena, such as vagueness (Lassiter and Goodman, 2013, 2015) and even scalar implicatures (Degen and Tanenhaus, 2015). Thus, it would be worth investigating in the future whether or not the difference between listener oriented social enrichment and 'objective' scalar enrichment can be reduced to the degree to which listeners' prior beliefs impact the pragmatic computation.

5 Conclusion

In this paper I presented a new formal model of the social/strategic aspects of sociolinguistic variation, one that analyzes social meaning as a kind of pragmatic enrichment. In order to give an analysis of the fine-grained differences in social meaning between variants and how speakers can use them to perform personae and construct linguistic styles, I introduced the *social meaning games* (SMG) framework, which is a formalization of Third Wave variation theory as a kind of signalling game with a probabilistic approach to agents' beliefs. This system thus represents the coming together of influential movements in quantitative sociolinguistics and formal pragmatics. I showed how the SMG framework can be used to predict patterns of language use based on speakers' values, i.e. the personae they are trying to construct in the context and their beliefs about how their interlocutors will interpret their linguistic offerings. I also showed how the system can be used to predict patterns of linguistic interpretation based on listener's values, their prior beliefs about the speaker, and their beliefs about the source of speakers' linguistic usage.

As stated, the SMG framework opens up a number of clear paths for future research into social meaning and language variation and change, the most pressing ones having to do with the relationship between social meaning and the grammar. I have said very little in this article about how social meaning is grammatically constructed and where the indexical fields and personae that are the meanings associated with variants 'live' in relation to other kinds of non-at-issue content such as presuppositions, scalar implicatures and expressive content. This was principally because I argued in section 2 that there are reasons to think that at least some kinds of social enrichment interact in non-trivial ways with grammatical structure. As such, the natural next step would be to integrate the SMG approach into

[34]See, for example, (Bucholtz, 1999, 2010) for the properties of crossing and the role of white American boys' use of African American variants in the construction of white masculinity in the US.

a broader model of grammar and language processing. Once this is done, we will be able to make more precise predictions for quantitative studies of linguistic variation, and we will be able to test more definitively the predictions of the model that I outlined in the previous sections. Additionally, in this paper, I have only treated individual variants (eg. (ING) and /t/ release), so a next related step would be to investigate what we might call the *compositional semantics of style*, extending the current static system that treats single messages (variants) to a dynamic system treating sequences of messages (styles). Finally, since this paper has dealt only with individual variants in isolation, in reality there is nothing in what I have said that limits the proposals in this paper to the meaning sociophonetic or even morphosyntactic variation. Since the game-theoretic tools used here are so general, the system extends directly to the meaning of non-linguistic variation such as systems of makeup, dress or other kinds of social signalling systems. Indeed, in the SMG approach, the differences between linguistic variation and sartorial variation would simply boil down to the inventory of messages, the kinds of meanings that can be associated with them, and the particular 'grammatical' rules that can be used to combine messages together in a more or less compositional way. In conclusion, then, I suggest that the new formal tools developed in this paper have rich applications within linguistics and semiotics more generally, and that they have potential to yield new theoretical and empirical insights into the relationship between form, meaning and stylistic performance.

Appendix: The SMG Framework

A **Social Meaning Game** is a tuple $\langle \{S, R\}, \langle \mathbb{P}, > \rangle, M, C, [\cdot], \mu_S, \mu_R, \rho, \sigma, Pr \rangle$ where:

1. S and R are the players.
2. $\langle \mathbb{P}, > \rangle$ is the **universe** (a relational structure), where
   - $\mathbb{P} = \{p_1, \ldots, p_n\}$ is a finite set of properties.
   - $>$ is a relation on $\mathbb{P}$ that is irreflexive and transitive.
   - P is possible **persona** ($P \in$ PERS) iff $P \subseteq \mathbb{P}$ and there are no $p_1, p_2 \in P$ such that $p_1 > p_2$.
3. $M = \{m_1, m_2 \ldots m_n\}$ is a finite set of **messages**.
4. C is a measure function on M describing the **cost** of each message.
5. $[\cdot]$ is the **indexation** relation.
   - For all messages $m \in M$, $[m] \subseteq \mathbb{P}$.
   - $\mathcal{F}_{m_i}$ notates the indexical field associated with m_i, i.e. $[\mathrm{m}_i] = \mathcal{F}_{m_i}$.
6. μ_S is a measure function on $\mathbb{P}$ describing S's **values** in the context.

7. μ_R is a measure function on $\mathbb{P}$ describing R's **values** in the context.
   - For all $P \in PERS$, $\mu^*_{S/R}(P) = \frac{\sum_{p \in P} \mu_{S/R}(p)}{|P|}$.
8. V_S is a function from pairs of sets of properties to real numbers describing how **useful** S finds R's interpretation.
   - Let V_S be a reward function from possible types and actions to real numbers.

     (42) For all $P, P' \in$ PERS, $V_S(P, P') = \begin{cases} n \times \mu^*_S(P') & \text{if } P = P' \\ n' \times \mu^*_S(P') & \text{otherwise} \end{cases}$ where $n > n'$.

   - For $m \in M$, $U_S(P, m, P') = V_S(P, P') - C(m)$.
9. V_R is a function from pairs of sets of properties to real numbers describing how **useful** R finds S's type.
   - Let V_R be a reward function such that:

     (43) For all $P, P' \in$ PERS, $V_R(P, P') = \begin{cases} n \times \mu^*_R(P) & \text{if } P = P' \\ n' \times \mu^*_R(P) & \text{otherwise} \end{cases}$ where $n > n'$.

   - For $m \in M$, $U_R(P, m, P') = V_R(P, P')$.
10. ρ is a receiver strategy, representing **S' beliefs** about how R will interpret their message.
    - **Listener Strategies (ρs)** are functions from M to probability distributions over PERS satisfying the following condition:

      (44) **Indexical Anchoring (R):**
      For all $m \in M$ and $P \in$ PERS, if $\rho(m, P) \neq 0$, then there is some $p \in P$ such that $p \in \mathcal{F}_m$.

11. σ is a speaker strategy, representing **R's beliefs** about why S would choose their message.
    - **Speaker strategies (σs)** are functions from the set of possible personae (PERS) to probability distributions over M satisfying the following condition:

      (45) **Indexical Anchoring (S):**
      For all $P \in$ PERS and $m \in M$, if $\sigma(P, m) \neq 0$, then there is some $p \in P$

such that $p \in \mathcal{F}_m$.

12. Pr is a probability distribution over personae describing R's **prior beliefs** about S.

References

Acton, E. (2016). Beyond grice: A socio-pragmatic framework for non-entailed meaning. *Linguistic Society of America Annual Meeting*.

Acton, E. K. (2014). *Pragmatics and the social meaning of determiners*. PhD thesis, Stanford University.

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, 14(03):471–485.

Bell, A. (1984). Language style as audience design. *Language in society*, 13(02):145–204.

Beltrama, A. (2016). *Bridging the gap. Intensifiers between semantic and social meaning*. PhD thesis, University of Chicago.

Benz, A., Jäger, G., Van Rooij, R., and Van Rooij, R. (2004). *Game theory and pragmatics*. Springer.

Bourdieu, P. (1979). *La distinction: critique sociale du jugement*. Les éditions de minuit, Paris.

Bourdieu, P. (1980). *Le sens pratique*. Les éditions de minuit, Paris.

Bourdieu, P. and Boltanski, L. (1975). Le fétichisme de la langue. *Actes de la recherche en sciences sociales*, 1:2–32.

Bourdieu, P. and Passeron, J.-C. (1970). *La reproduction: Éléments pour une théorie du système d'enseignement*. Les éditions de minuit, Paris.

Bucholtz, M. (1996). Geek the girl: Language, femininity, and female nerds. *Gender and belief systems*, pages 119–132.

Bucholtz, M. (1999). You da man: Narrating the racial other in the production of white masculinity. *Journal of sociolinguistics*, 3(4):443–460.

Bucholtz, M. (2010). *White kids: Language, race, and styles of youth identity*. Cambridge University Press.

Bunin Benor, S. (2001). The learned /t/: phonological variation in orthodox jewish english. *University of Pennsylvania Working Papers in Linguistics*, 7(3):2.

Campbell-Kibler, K. (2006). *Listener perceptions of sociolinguistic variables: The case of (ING)*. PhD thesis, Stanford University.

Campbell-Kibler, K. (2007). Accent,(ing), and the social logic of listener perceptions. *American speech*, 82(1):32–64.

Campbell-Kibler, K. (2008). I'll be the judge of that: Diversity in social perceptions of (ing). *Language in Society*, 37(05):637–659.

Chemla, E. and Spector, B. (2011). Experimental evidence for embedded scalar implicatures. *Journal of semantics*, pages 359–400.

Cheshire, J. (1982). Variation in an english dialect: A sociolinguistic study. *Cambridge Studies in Linguistics London*, 37.

Chierchia, G., Fox, D., and Spector, B. (2008). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In *Semantics: An International Handbook of Natural Language Meaning*. Mouton de Gruyter, Berlin.

Clark, R. L. (2014). *Meaningful games: Exploring language with game theory*. MIT Press.

D'Arcy, A. F. (2005). *Like: Syntax and development*. PhD thesis, University of Toronto.

Degen, J., Franke, M., and Jäger, G. (2013). Cost-based pragmatic inference about referential expressions. In *Proceedings of the 35th annual conference of the cognitive science society*, pages 376–381.

Degen, J. and Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive science*, 39(4):667–710.

Deshaies, D. and Laforge, E. (1981). Le futur simple et le futur proche dans le français parlé dans la ville de québec. *Langues et Linguistique*, 7:21–37.

Dror, M., Granot, D., and Yaeger-Dror, M. (2013). Speech variation, utility, and game theory. *Language and Linguistics Compass*, 7(11):561–579.

Ducrot, O. (1969). Présupposés et sous-entendus. *Langue française*, 4:30–43.

Eckert, P. (1989). The whole woman: Sex and gender differences in variation. *Language variation and change*, 1(03):245–267.

Eckert, P. (2000). *Language variation as social practice: The linguistic construction of identity in Belten High*. Wiley-Blackwell.

Eckert, P. (2005). Variation, convention, and social meaning. In *Annual Meeting of the Linguistic Society of America. Oakland CA*, volume 7.

Eckert, P. (2008). Variation and the indexical field. *Journal of sociolinguistics*, 12(4):453–476.

Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual review of Anthropology*, 41:87–100.

Frank, M. C. and Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998.

Franke, M. (2009). *Signal to act: Game theory in pragmatics*. PhD thesis, Institute for Logic, Language and Computation.

Franke, M. and Jäger, G. (2016). Probabilistic pragmatics, or why bayes' rule is probably important for pragmatics. *Zeitschrift fr̈ Sprachwissenschaft*, 35:3–44.

Goodman, N. D. and Lassiter, D. (2014). Probabilistic semantics and pragmatics: Uncertainty in language and thought. *Handbook of Contemporary Semantic Theory. Wiley-Blackwell*.

Gratton, C. (2015). Indexin' gender: variable (ING) and the creation of non-binary trans identities. *New Ways of Analyzing Variation*, 44.

Grice, P. (1975). Logic and conversation. *Syntax and Semantics*, 3:41–58.

Halpern, J. Y. (2003). *Reasoning about uncertainty*, volume 21. MIT press Cambridge.

Hazen, K. (2006). In/ing variable. In Brown, K., editor, *Encyclopedia of language and linguistics*, pages 580–581. Elsevier.

Heim, I. (1982). *The semantics of definite and indefinite noun phrases*. PhD thesis, University of Massachussets, Amherst.

Houston, A. (1985). *Continuity and change in English morphology: The variable (ING)*. PhD thesis, University of Pennsylvania.

Kaplan, D. (1999). The meaning of ouch and oops: Explorations in the theory of meaning as use. *Manuscript, UCLA*.

Kiesling, S. F. (2009). Style as stance. *Stance: sociolinguistic perspectives*, 171.

Labov, W. (1963). The social motivation of a sound change. *Word*, 19(3):273–309.

Labov, W. (1966). *The social stratification of English in New York city*. Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

Labov, W. (2012). *Dialect diversity in America: The politics of language change*. University of Virginia Press.

Lambert, W. E. (1967). A social psychology of bilingualism. *Journal of social issues*, 23(2):91–109.

Lassiter, D. and Goodman, N. D. (2013). Context, scale structure, and statistics in the interpretation of positive-form adjectives. In *Semantics and Linguistic Theory*, volume 23, pages 587–610.

Lassiter, D. and Goodman, N. D. (2015). Adjectival vagueness in a bayesian model of interpretation. *Synthese*, pages 1–36.

Levon, E. (2014). Categories, stereotypes, and the linguistic perception of sexuality. *Language in Society*, 43(05):539–566.

Lewis, D. (1969). *Convention*. Harvard UP, Cambridge.

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

Luce, R. D. (1959). On the possible psychophysical laws. *Psychological review*, 66(2):81.

Maddeaux, R. and Dinkin, A. (2015). Is like like like? evaluating the same variant across multiple variables. *New Ways of Analyzing Variation*, 44.

McConnell-Ginet, S. (2011). *Gender, Sexuality and Meaning: Linguistic Practice and Politics*. Oxford University Press, Oxford.

Mühlenbernd, R. (2013). *Signals and the Structure of Societies*. PhD thesis, Universität Tübingen.

Mühlenbernd, R. and Franke, M. (2012). Signaling conventions: who learns what where and when in a social network. *Proceedings of EvoLang IX*, pages 242–249.

Ochs, E. (1992). Indexing gender. *Rethinking context: Language as an interactive phenomenon*, 11:335.

Osborne, M. J. and Rubinstein, A. (1994). *A course in game theory*. MIT press.

Podesva, R. (2004). On constructing social meaning with stop release bursts. In *Sociolinguistics Symposium*, volume 15.

Podesva, R. (2006). *Phonetic detail in sociolinguistic variation: Its linguistic significance and role in the construction of social meaning*. PhD thesis, Stanford University.

Podesva, R. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of sociolinguistics*, 11(4):478–504.

Podesva, R. J., Reynolds, J., Callier, P., and Baptiste, J. (2015). Constraints on the social meaning of released/t: A production and perception study of us politicians. *Language Variation and Change*, 27(01):59–87.

Poplack, S. and Turpin, D. (1999). Does the FUTUR have a future in (canadian) french? *Probus*, 11:133–64.

Potts, C. (2007). The expressive dimension. *Theoretical linguistics*, 33(2):165–198.

Potts, C., Lassiter, D., Levy, R., and Frank, M. C. (2015). Embedded implicatures as pragmatic inferences under compositional lexical uncertainty. *Journal of Semantics*, page ffv012.

Quinley, J. and Mühlenbernd, R. (2012). Conquest, contact, and convention: simulating the norman invasion?s impact on linguistic usage. *Proceedings of BRIMS 2012*, pages 113–118.

Rampton, B. (1995). Language crossing and the problematisation of ethnicity and socialisation. *Pragmatics*, 5(4).

Récanati, F. (2003). Embedded implicatures. *Philosophical perspectives*, 17(1):299–332.

Rickford, J. and Closs Traugott, E. (1985). Symbol of powerlessness and degeneracy, or symbol of solidarity and truth? paradoxical attitudes towards pidgins and creoles. In Greenbaum, S., editor, *The English Language Today*, pages 252–261. Pergamon, Oxford.

Sankoff, G., Wagner, S. E., and Jensen, L. (2012). The long tail of language change: Québécois french futures in real time. *University of Pennsylvania Working Papers in Linguistics*, 18(2):13.

Silverstein, M. (1976). Shifters, linguistic categories, and cultural description. *Meaning in anthropology*, 1:1–55.

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, 23(3):193–229.

Smith, E., Hall, K. C., and Munson, B. (2010). Bringing semantics to sociophonetics: Social variables and secondary entailments. *Laboratory Phonology*, 1(1):121–155.

Stalnaker, R. (1978). Assertion. In Cole, P., editor, *Syntax and semantics*, volume 9, pages 315–332. Academic Press, New York.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT press.

Tamminga, M. (2014). *Persistence in the production of linguistic variation*. PhD thesis, University of Pennsylvania.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *science*, 331(6022):1279–1285.

Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban british english of norwich. *Language in society*, 1(02):179–195.

Van Rooy, R. (2003). Being polite is a handicap: Towards a game theoretical analysis of polite linguistic behavior. In *Proceedings of the 9th conference on Theoretical aspects of rationality and knowledge*, pages 45–58. ACM.

Wagner, S. E. and Sankoff, G. (2011). Age grading in the montréal french inflected future. *Language Variation and Change*, 23(3):275.

Weinreich, U., Labov, W., and Herzog, M. (1968). Empirical foundations for a theory

of language change. In Lehmann, W. P., editor, *Directions for historical linguistics: A symposium*, pages 95–195. University of Texas Press, Austin.

Zhang, Q. (2008). Rhotacization and the 'beijing smooth operator': the social meaning of a linguistic variable. *Journal of Sociolinguistics*, 12(2):201–222.