Linguistics and Philosophy (2019) 42:419–450
https://doi.org/10.1007/s10988-018-9254-y

Signalling games, sociolinguistic variation and the construction of style

Heather Burnett[1]

Published online: 14 March 2019

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 linguistic phenomena (Goffman in Encounters: Two studies in the sociology of interaction, Bobbs-Merrill, Indianapolis, 1961; in Interaction ritual: Essays on face-to-face interaction, Aldine, Oxford, 1967; in Strategic interaction, vol 1, University of Pennsylvania Press, Philadelphia, 1970; Bourdieu in Soc Sci Inf 16(6):645–668, 1977; Dror et al. in Lang Linguist Compass 7(11):561–579, 2013; in Lang Linguist Compass 8(6):230–242, 2014; Clark in Meaningful games: Exploring language with game theory, MIT Press, Cambridge, MA, 2014, among others), 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 in Ann Rev Anthropol 41:87–100, 2012, for an overview) using *signalling games* (Lewis in Convention, Harvard University Press, Cambridge, 1969) and a probabilistic approach to speaker/listener beliefs of the kind commonly used in the Bayesian game-theoretic pragmatics framework (see Goodman and Lassiter in Probabilistic semantics and pragmatics: Uncertainty in language and thought. Handbook of Contemporary Semantic Theory, Wiley, Hoboken, 2014; Franke and Jäger in Zeitschrift für Sprachwissenschaft, 35(1):3–44, 2016, for recent overviews).

✉ Heather Burnett
heather.susan.burnett@gmail.com

[1] Laboratoire de Linguistique Formelle, CNRS-Université de Paris 7-Diderot, 75013 Paris, France

Keywords Game-theoretic pragmatics · Social meaning · Sociolinguistic variation · Bayesian pragmatics · Identity construction

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 such as those in (1)–(2). 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*.

(1) (ING)

a. I'm work**ing** on my paper. [ɪŋ]
b. I'm work**in'** on my paper. [ɪn]

(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

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; Jäger 2011, for overviews), 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 linguistic phenomena (Goffman 1961, 1967, 1970; Bourdieu 1977; Dror et al. 2013, 2014; Clark 2014, among others), a general formal 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 an overview) using *signalling games* (Lewis 1969) and a probabilistic approach to speaker/listener beliefs of the kind commonly used in the Bayesian game-theoretic pragmatics framework (see Goodman and Lassiter 2014; Franke and Jäger 2016, for recent overviews).

[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, swear words and social networks (Van Rooy 2003; Mühlenbernd and Franke 2012; Mühlenbernd 2013; Quinley and Mühlenbernd 2012; McCready 2012; McCready et al. 2013, among others). Additionally, there is already some work aiming at integrating formal semantics/pragmatics and variationist sociolinguistics using non-game-theoretic methods (Lassiter 2008; Smith et al. 2010; Acton 2014, 2016; Acton and Potts 2014; 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).

The paper is laid out as follows: in Sect. 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, and I propose that social meaning of the kind studied in this paper should be viewed as an instance of pragmatic enrichment. As a consequence, a unified framework that can treat both social meaning and other kinds of meaning in context should be developed. In Sect. 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 analysed 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 analysed 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 on 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 my proposal to have any substantive content, I must be more precise concerning the definition of the games (the players, the architecture, utility functions etc.) and what their solution concepts are. In Sect. 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 patterns of sociolinguistic variation, on the one hand, and the options for and constraints on the construction of personal linguistic styles, on the other. 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 Social meaning as pragmatic enrichment

Suppose we are having a conversation and the person that we are talking to says (1a) [repeated as (3a)]. What do we understand from this utterance?

[2] The main proposals in this paper are limited to modelling the aspects of linguistic variation that are determined by what sociolinguists call *social*, *external* or *non-linguistic* factors. Patterns of linguistic variation are also determined by other factors which are not social/strategic in nature [for example general cognitive factors associated with linguistic production and comprehension, as well as grammatical factors (what Labov (1966) 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.

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

(3) a. I'm work**ing** on my paper. [ɪŋ]
b. I'm work**in'** on my paper. [ɪn]

From hearing (3a), 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 (3a), possibly something about the properties of the speaker, of the working event, or maybe even of both. Likewise, if our interlocutor says (1b) [repeated as (3b)], 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 (3a).

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 et al. 1960). 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, but 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. In other words, 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 a MGT study with 70 American participants (the majority in their early 20s) using stimuli formed from political speeches of 6 American politicians (Barack Obama, John Edwards, Nancy Pelosi, George W. Bush, Hillary Clinton, and Condoleezza Rice). As in Campbell-Kibler's study, the /t/ release study yielded a number of results concerning associations with released versus 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 wa[ɾ]er[4]). On the other hand, Nancy Pelosi was rated as significantly less friendly and less sincere when she used released /t/, and Barack 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.

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 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 (4a) or implicatures generated by expressions with expressive content (4b) (see also McConnell-Ginet 2011; Smith et al. 2010; Acton 2014; Beltrama 2016, for additional support for versions of this claim).

(4) 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. (2015)'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. Furthermore, participants in Podesva et al.'s study stated that they found Pelosi's use of released /t/ to be pretentious and fake, so using the flap makes her sound more sincere and creates a positive evaluation. In other words, social enrichment is dependent on

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

speaker identity, but also (more importantly) on listeners' **interpretations** of speakers' linguistic performances.

In this section, I proposed that social meaning should be viewed as an implicature that is triggered by the use of particular variants and should be integrated into a broader theory of formal pragmatics. 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 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,[5] 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 the *Third Wave* approach to variation. 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 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 White House 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 White House lawn. In this 'careful' context, his rate of *-in'* drops

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

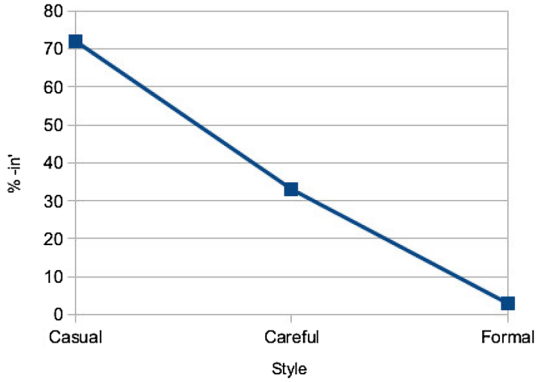

Fig. 1 President Obama's use of (ING) across three contexts

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 Fig. 1, reproduced from Labov (2012, 22).

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,

> 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**. (Labov 2012: 22)

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, and recent studies of style shifting of private citizens have shown that intra-speaker variation is widespread, with people significantly changing their use of variants across contexts (Cheshire 1982; Kiesling 1998; Podesva 2007; Gratton 2016; van Hofwegen 2017, among many others) and even across sections of discourse (see Kiesling 2009; Calder 2018, for example).

These studies 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

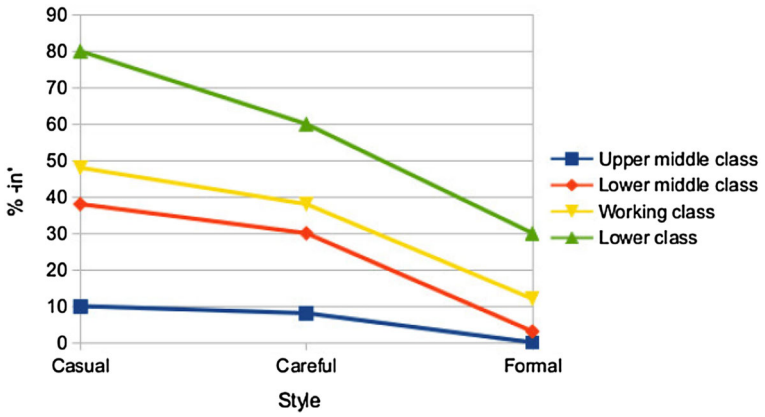

Fig. 2 Labov (1966): (ING) by social class and style (casual, careful, formal) in New York City

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* or *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: the most famous results from this field are 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 analysed as interactive language use, and thus the game-theoretic model that will be presented in Sect. 4 aims to model 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 Fig. 2, reproduced from Labov (1966)'s famous study of language use in New York City. As we saw with Obama in the previous section, the use of the *-in'* variant decreases with a more formal style,[6] 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).

[6] 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.

Likewise, in developing his influential *Audience Design* theory of style shifting, Bell (1984) says that the relation between style shifting and social stratification is "more than an interrelation. It is a *derivation*;" in particular, he says that "variation on the style dimension within the speech of a single speaker derives from and echoes the variation which exists between speakers on the 'social' dimension." (Bell 1984, 151) Following this line of research, I argue 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/ (i.e. pronuncing [r jt] (centralized) vs [rajt] for *right* and [m ws] (centralized) vs [maws] for *mouse*) in Martha's Vineyard, an island south of Cape Cod in Massachusetts. 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) Labov observes that speakers' orientations towards or away from the island and the old way of life was the best predictor of which variant they prefer to use, with more locally oriented individuals showing much higher degrees of centralization (Labov 1963, 30), i.e. much higher use of the older, more local variant. Thus, a formal theory of social stratification should be able to capture how inter-speaker differences break down along lines associated with ideological stances and other social practices.

3.3 The Third Wave approach to variation

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 [see also Dror et al. (2013)]. 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, the Third Wave approach to variation pursues a unified analysis of style shifting and social stratification as rational language use,[7] and (in a nutshell) the account is as follows;[8] The use of a variant in context is related to (or in this discipline's

[7] The treatment of social stratification as rational language use distinguishes the "Third Wave" of variationist sociolinguistics from both the "First Wave", in which social stratification is analysed as a reflection of demographic social structures (i.e. gender, age and social class), and the "Second Wave", in which social stratification is analysed as a reflection of locally socially constructed structures (*jocks* vs *burnouts* etc.). See Eckert (2012) for a more detailed description of the three waves.

[8] 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 implications for a theory of language change. In the model that will be presented in the next section, variants come with associated sets of properties (indexical fields) and the model does not have any evolutionary or large-scale

terminology *indexes*) sets of properties, stances or other concepts/ideas that are to be attributed to the speaker (Ochs 1992, 1993; Silverstein 1979, 2003; Eckert 2008, 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.

TW clearly makes reference to notions like interactivity and rationality; however, it is not a mathematical theory. Thus, although its insights concerning social meaning, sociolinguistic variation and identity/persona construction are clear, by virtue of its form, such a theory cannot be directly incorporated into a broader formal theory of pragmatics. I argue that such an incorporation is desirable for a number of reasons. Firstly, I argued above (following others) that inferences triggered by sociolinguistic variants should be analysed as instances of pragmatic enrichment; thus, formal theories of language use and interpretation will be incomplete if they cannot be applied to this empirical domain. Secondly, their lack of formalization has isolated the insights of influential sociolinguistic theories such as TW inaccessible to have been isolated from work in cognitive science, computer science and artificial intelligence, to the detriment of these fields.[9] Finally, beyond the field of linguistic pragmatics, social meaning and persona/identity construction are fundamental theoretical notions in many disciplines of the humanities and social sciences. They play a key role in our understanding of linguistic and non-linguistic phenomena studied in anthropology, sociology, philosophy, psychology and gender studies (see Cerulo 1997; Hacking 1999, for overviews). Moreover, methods based on identity construction through language are widely employed outside academia, for example in education (e.g. Roberts 1991; Varelas 2012; Kelly 2014, among others), management (DeRue and Ashford 2010; Dutton et al. 2010, among others), social work (Miehls and Moffatt 2000, among others), digital communication (Zhao et al. 2008), and even social justice (Taylor 2000; Charmaz 2011, among others). It is therefore crucial that our understanding of the relationship between language, meaning, identity and variation be as detailed, as explicit and as well-founded as possible. Formalization is a powerful tool that we can use to carefully distinguish between different aspects of theoretical proposals made in sociolinguistics and for precisely identifying empirical predictions made by competing analyses.

Footnote 8 continued
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.

[9] See Cameron (2016) and Hardaker (2016) for recent discussion of the importance of incorporating more sophisticated information about social meaning and its relation to identity into online gender-based harassment prevention tools. See Nguyen et al. (2016) for a comprehensive overview of the difficulties of integrating informal insights from sociolinguistics into Natural Language Processing.

With these considerations in mind, in the second half of the paper I propose to formalize the Third Wave approach to variation using signalling games and a Bayesian approach to speaker/listener uncertainty. In doing so, I hope to bring social meaning into the domain of game-theoretic pragmatics and, more generally, identity construction through language into the domain of formal pragmatics.

4 Social meaning games

This section presents the *social meaning game* (SMG) framework. 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). In a nutshell, a *signalling game* is a game between two agents, *S* (the *speaker/sender*) and *L* (the listener/receiver). *S* has a piece of information that they wish to communicate to *L* (their *type*). S's action is to choose a message *m* to send L, and L's action is to assign an interpretation to *m*, and, 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 Lewisian signalling games, *S* and L's pay-offs are calculated based on coordination: (broadly speaking) both players win if *L* correctly interprets S's message, updating their beliefs accordingly, and they both lose if S's type and L's interpretation do not converge, and *L* 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 interaction between two agents: *S* (speaker/sender) and *L* (listener/receiver). *S* has a set of properties characterizing themselves they wish to communicate to *L* (their *type*). S's action is to choose a message *m* to send L[10], and L'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 L's pay-offs and, consequently, the **solution concept** (or rule that determines how the game is played) will be very similar to what is found in *Iterated Best Response* (IBR) or *Rational Speech Act* (RSA) models, which have been widely used in formal approaches to Gricean pragmatics (Franke 2009; Frank and Goodman 2012; Lassiter and Goodman 2013; Degen and Franke 2012; Franke and Jäger 2016; Degen and Tanenhaus 2015; Bergen et al. 2016, among others). However, contrary to classic IBR/RSA models, something else will play an important a role in calculating an agent's utility (and subsequent actions): *S* and L's personal preferences in the context, which we will call their *values*.

[10] Note that the framework does not at all assume that all or even most aspects of message/interpretation selection or utility calculation are conscious or intentional. We know from the psychological literature [for example, the literature on motion planning (Rosenbaum et al. 2007, 2012)] that we make enormously many subconscious choices and calculations when we engage in daily cognitive activities (see Dennett 1993; Graziano 2013, among others). Furthermore, work in the field of *evolutionary game theory* (see Gintis 2000) has shown that even agents not possessing consciousness (like single-celled organisms) appear to engage in the kinds of utility maximizing calculations that are presented in this paper. Thus, the present proposal has nothing to say about the role/limitations of consciousness in sociolinguistic variation and interpretation.

4.1 Basic setup

The definition of a SMG is laid out more formally in Definition 4.1. Some of the lines of Definition 4.1 will doubtlessly be opaque to the reader; however, they will be further elaborated in the rest of this section.

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

1. S and L 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 asymmetric.
3. M is a finite set of **messages**.
4. C is a function from M to $\mathbb{R}$ describing the **cost** of each message.
5. $[\cdot]$ is the **indexation** relation (to be described below).
6. Pr is a probability distribution over sets of properties describing L'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 relationships between properties, namely incompatibility; 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 (5), 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).

(5) $\mathbb{P}$ = {competent, incompetent, friendly, aloof}
   a. competent > incompetent
   b. friendly > aloof

In addition to the attribution of individual properties and the meaning of individual variables, TW 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; Zhang 2005, 2008; Eckert 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 the maximally consistent sets of properties,[11] as shown in Definition 4.2.

[11] For simplicity, we assume that the personae are maximally consistent sets; however, in future empirical work, it may prove to be interesting to consider 'sub personae', i.e. non maximal sets.

Definition 4.2 π is a possible **persona** ($\pi \in$ PERS) iff

1. $\pi \subseteq \mathbb{P}$ and there are no $p_1, p_2 \in \pi$ such that $p_1 > p_2$. Consistency
2. There is no $\pi' \in$ PERS such that $\pi \subset \pi'$. Maximality

In our simple example, then, by Definition 4.2, the possible personae in the universe in (5) are shown in (6): 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.

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

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

(7) **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.

In order to show how SMGs work, we will start by showing how to model Labov (2012)'s study of President Obama's use of (ING) across three contexts. Thus, in the game, we will have two messages (8).

(8)

MESSAGE	COST
-ing	0
-in'	0

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,[12] we will assume for our example that there are no differences in the cost of using *-in'* than in the cost of using *-ing*.

This being said, having articulated costs may become important when it comes time to integrate SMGs within a broader model of linguistic production and interpretation. 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

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

depends on a wide range of factors, only a subset of which depend on social meaning and persona construction. 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 may induce a bias in favour of one variant over another. For example, it has been shown that (ING) is conditioned by grammatical category and other abstract properties of morphological structure (Labov 1966; Houston 1985; Tamminga 2014), so it seems reasonable to capture the generalization that *-ing* is disfavoured in some grammatical environment compared to *-in'* by assigning an occurrence of *-ing* in that environment a higher cost than is assigned to *-in'* in the same context, possibly through the use of a *harmonic grammar* (Legendre et al. 1990; Smolensky and Legendre 2006).[13] Of course, adding internal/grammatical conditioning factors to the model would require much more elaborate message representations which would complicate the exposition here, so we will assign a cost of zero to both variants of (ING) in (8).

As mentioned in Sect. 3, in 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 (9).

(9) **Indexation relation** ($[\cdot]$).
For all messages $m \in M$, $[m] \subseteq \mathbb{P}$.

Much current work within TW is devoted to studying the structure of indexical fields, investigating whether there are different orders of indexicality within a field (Silverstein 2003), whether there are meaningful relations between the properties that make up a variant's field (Eckert 2008), and, if so, whether there exists some kind of algorithm that can extract these relations automatically (Oushiro 2015). In this paper, I will keep things as simple as possible and not impose any structure on these sets, 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) and Campbell-Kibler (2009), we will assume that the variants of (ING) are associated with the sets shown in (10), which I will call *Eckert fields*.

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

The Eckert fields shown in (10) correspond to the standard representation within sociolinguistics (see the representations proposed by Campbell-Kibler 2008, 2009; Eckert 2008; Moore and Podesva 2009; Walker et al. 2014; Beaton and Washington 2015; Tyler 2015; Drager 2015; Oushiro 2015); however, as we will see below, it turns out that the objects in (10) do not give the right result when incorporated into an IBR/RSA-style model. Therefore, we will take advantage of Richard Montague's

[13] Indeed, there exist mathematical connections between game-theoretic syntax/semantics and optimality-based syntax-semantics (Dekker and Van Rooy 2000; Franke and Jäger 2012).

Table 1 Messages in Obama example

Variant	Eckert field	Eckert–Montague field
-ing	{competent, aloof}	**{comp., aloof}**, {comp., friend.}, {incomp., aloof}
-in'	{incompetent, friendly}	**{incomp., friend}**, {comp., friend}, {incomp., aloof}

Table 2 Potential voter's prior beliefs (that Obama is **aloof**)

Persona	Stern leader	Cool guy	Asshole	Doofus
π	{comp, aloof}	{comp, friend}	{incomp, aloof}	{incomp, friend}
$Pr(\pi)$	0.30	0.20	0.30	0.20

important observation that (formally speaking) we often have multiple ways of looking at an object: either we can look at it directly, or, equivalently, we can look at it as its set of characterizing properties. Thus, in the spirit of Montague (1973), we will look at Eckert indexical fields equivalently through the personae that they have the potential to construct;[14] in other words, the Eckert–Montague field on {competent, friendly} will consist of all the personae that are either competent or friendly. We will call these type-lifted fields *Eckert–Montague fields*, shown in Table 1. As shown in this table, there is some overlap in the Eckert–Montague fields of (ING) but crucially only *-ing* can be used to construct the {competent, aloof} (stern leader) persona; whereas, only *-in'* can be used to construct the {incompetent, friendly} persona (the doofus).

We are now ready to describe how the game is played; that is, how the moves of the speaker and the listener are chosen.

As in IBR/RSA models, the listener (L) has prior beliefs about the properties of the speaker before they speak. These beliefs can be specific (i.e. *George Bush is like X*) or general (i.e. *Americans are like X*). These beliefs are represented as a probability distribution (Pr) over personae. Returning to Labov's Obama example: consider the *casual context* in which Obama is at the barbecue. Suppose, in this context, Obama's interlocutor has the prior belief that he is aloof, since he is the president. Indeed, one of the main reasons that politicians do such 'meet and greet' events is to try to counteract this impression in their potential voters. The way that we encode this belief in the model is by putting more probability mass on the personae that are aloof (the stern leader and the asshole) than on the personae that are friendly (the cool guy and the doofus), as shown in Table 2.

As in IBR/RSA models, the social interpretation process proceeds in a couple of steps: when they hear a variant, the listener first focusses their attention on the personae in the variant's Eckert–Montague field, discarding the other personae and updating their beliefs accordingly. More technically, L conditions their beliefs on the meaning

[14] In the terminology of *Generalized Quantifier Theory* (Barwise and Cooper 1981; Keenan and Stavi 1986), the Eckert–Montague field on {competent, friendly} is derived through taking the *Montagovian Individual* on *competent or friendly*. The Montagovian Individual on an object a is the characteristic function of the set of properties that include a, see Montague (1973, 260) and Peters and Westerståhl (2006). Montagovian Individuals are useful in formal semantics because they allow for proper names to be treated as Generalized Quantifiers (functions from properties to truth values), just like more clearly quantified noun phrases like *every student*. Therefore, we can have a single semantic type for the syntactic category of *determiner phrase*.

Table 3 L's beliefs immediately after hearing m at the barbecue ($Pr(\pi|m)$)

m	Stern leader {comp, aloof}	Cool guy {comp, friend}	Asshole {incomp, aloof}	Doofus {incomp, friend}
-ing	0.375	0.25	0.375	0
-in'	0	0.286	0.428	0.286

of the message (11): they intersect each persona with the variant's indexical field and then normalize the measure. Observe that the conditionalization operation would not give the right results applied to simple Eckert fields, since the only persona that would remain under consideration would be the persona corresponding to the field itself.

(11) $Pr(\pi|m) = \frac{Pr(\{\pi\} \cap [m])}{Pr([m])}$ **Conditionalization**

The results of conditionalization for both variants of (ING) on listener priors at the barbecue are shown in Table 3. If the listener hears *-ing*, they are certain that their interlocutor is not a doofus, but they remain uncertain about the remaining possibilities. Likewise, if they hear *-in'*, they are certain that S is not the stern leader (because a stern leader would not say *-in'*); so they assign this persona a zero probability and normalize over the cool guy, asshole and doofus.

IBR/RSA models aim to formalize certain aspects of Gricean pragmatics, most notably, the role that informativity plays in successful communication. According to this framework, coordination, and therefore communication, occurs because speakers try to make the most informative statement possible, and listeners know this. Thus, informativity serves as an external convergence point for both speaker and listener. As such, message informativity is encoded as part of S's utility function (U_S): the speaker's utility of a message, given that they wish to construct a particular persona, is measured as the informativity of the message (given the desired persona), minus whatever costs are associated with the message (12). Following Frank and Goodman (2012), who follow Shannon (1948), we measure the informativity of a message m for a persona P as the natural logarithm (ln) of the prior probability of π conditioned on the meaning of m.

(12) $U_S(\pi, m) = ln(Pr(\pi|m)) - C(m)$ **RSA-style utility function**

In SMGs, we will also adopt the proposal that persona/identity construction through language is driven (in part) by informativity:[15] the speaker is trying to give the listener the most information possible about their desired persona, and the listener assumes that the speaker is giving them (intentionally or not) the most information about the kind of person that they are. Note that the joint assumption of informativity does not require that the listener is positively disposed or actively trying to coordinate with their interlocutor; rather, they are simply trying to extract the most information possible out of their interlocutor's linguistic offering.

Plugging the values from Table 3 into the utility function in (12) assigns the following utilities to pairs of personae and variants in Table 4.

[15] See Burnett (2017) for arguments that informativity also plays a role in sociolinguistic production.

Table 4 Obama's utility for trying to construct π with m at the barbecue ($U_S(\pi, m)$)

m	Stern leader {comp, aloof}	Cool guy {comp, friend}	Asshole {incomp, aloof}	Doofus {incomp, friend}
-ing	− 0.9808293	− 1.386294	− 0.9808293	$-\infty$
-in'	$-\infty$	− 1.251763	− 0.8486321	− 1.251763

4.2 Predicting speaker behaviour

A major assumption underlying game-theoretic treatments of language use and understanding is that speakers and listeners are (at least) *approximately rational*. We assume that they are *rational* in the sense that they are trying to maximize their utility; however, we assume that they are only approximately so, meaning that they may **not** in fact always pick the optimal action. It is well known that mental computation can be impeded by a variety of things (tiredness, attention deficits etc.). Therefore, in order to account for possible variability in action selection, we assume that, rather than just picking the variant with the highest utility, S chooses the best option given a noise-perturbed assessment of 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 Bayesian game-theoretic approaches to a variety of pragmatic phenomena (Frank and Goodman 2012; Degen et al. 2013; Lassiter and Goodman 2015; Franke and Jäger 2016; Bergen et al. 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 in (13), where $P_S(m|\pi)$ notates the probability of S using m, given that they're trying to construct π. The constant α in (13) represents how much indeterminacy the model allows. Setting α to ∞ recovers deterministic choice; whereas, setting it to a low value allows more variation.

$$P_S(m|\pi) = \frac{exp(\alpha \times U_S(\pi, m))}{\sum_{m' \in M} exp(\alpha \times U_S(\pi, m'))} \quad \textbf{Soft max choice rule} \tag{13}$$

Suppose, at the barbecue, Obama wishes to construct the *cool guy* persona ({competent, friendly}). Plugging the values in Table 4 into the choice rule in (13), and with α set at 6 ($\alpha = 6$), we predict that Obama will use *-in'* around **69%** of the time (and *-ing* around 31% of the time), which is close to what Labov found.

After the barbecue finishes, Obama moves to take questions from reporters on the White House lawn. In this situation, we might imagine that although Obama is still in a relatively informal context (and so wants to construct the cool guy persona), his

Table 5 Journalist's prior beliefs (that Obama is **incompetent**)

Persona π	Stern leader {comp, aloof}	Cool guy {comp, friend}	Asshole {incomp, aloof}	Doofus {incomp, friend}
$Pr(\pi)$	0.20	0.20	0.30	0.30

interlocutors are antagonistic journalists who may think he's incompetent. Again, we represent this belief as Obama's interlocutor's priors: in Table 5, personae that are incompetent are weighted higher than competent personae.

This change (from Tables 2, 3, 4, 5) has an important effect on Obama's predicted linguistic choices since changing prior beliefs will cause the informativity of the messages to change, which, in turn, will cause the speaker utility of a message to change, resulting in the predicted use of *-in'* to drop to around **31%**, which again is similar to Labov's observation.

Finally, when Obama makes a speech at the Democratic National Committee, he is in a very formal situation. In such contexts, it is generally not very useful to appear particularly friendly; rather, what is valued in very formal contexts is the aloofness of the stern leader. So we might think that he constructs the *stern leader* ({competent, aloof}) in this context. Since neither *competent* nor *aloof* are in the indexical field of *-in'*, this variant cannot be used to construct Obama's chosen persona; therefore, we predict that he should avoid this variant in this context. Thus, the complex patterns observed in Labov's sociolinguistic study of style shifting can be captured in the SMG framework through the interplay between speakers' (context-sensitive) persona selection and how they reason about their interlocutor's beliefs.

4.3 Persona selection games

The Obama example in the previous section highlights the role that speakers' persona selection plays in sociolinguistic variation, and this also raises questions concerning how similar social meaning games are/ought to be to classical signalling games. In particular, in the kind of signalling games that are most commonly used in formal linguistics, S's type is determined by 'Nature', as described by Franke:

> Game theorists like to think of the states of a signaling game as initial chance moves by a third player, called **Nature**, who selects any state $t \in T$ with probability $Pr(t)$, without any strategic concern of her own (cf. Harsanyi 1967, 1968a, b). In a signaling game, Nature reveals her choice to only the sender, but not the receiver. (Franke 2009, 129)

The metaphor that S's type is chosen by impartial Nature seems perfectly appropriate for communication of most kinds of truth conditional meaning: S observes a fact about the world and then tries to report it to L. However, it is clear that, at least sometimes, the speaker has a hand in choosing which identity to construct and they are consciously aware of this: we reflect on how we want to present ourselves when we think about getting a new haircut, watch how we dress/talk on job interviews and

other important occasions. So we need to incorporate speaker agency into persona selection: S's type should be chosen not by *Nature*, but by *Human Nature*. I therefore propose to incorporate persona selection into SMGs by supposing that the speaker has a preferential ordering over personae, which they may or may not be conscious of, and which contributes to determining when/how often they choose a particular persona to communicate to L. Of course, in most utterances, our attention is focused on whatever it is we are trying to do and/or the propositional information we are trying to communicate, not on identity construction, so persona selection for many of our utterances is unconscious.[16]

More formally, I propose to extend the structure of SMGs with a pay-off function μ mapping personae to real numbers, which I call S's *values*.

Definition 4.3 An SMG with persona selection is a tuple $\langle \{S, L\}, \langle \mathbb{P}, > \rangle, M, C, [\cdot], Pr, \mu \rangle$, where $\langle \{S, L\}, \langle \mathbb{P}, > \rangle, M, C, [\cdot], Pr \rangle$ is an SMG (as in Definition 4.1) and μ is a function from PERS to $\mathbb{R}$ mapping each π to the value that S assigns to constructing it in the context.

It seems reasonable to think that this μ should be taken from a larger, non-linguistic game that captures S and L's interaction situation. For example, if we return to the barbecue example: social conventions at barbecues are such that, if someone is friendly to you, you should be friendly back. Likewise, if someone is not friendly to you, then why would you be friendly to them? Thus, coordination on (un)friendliness is optimal for most private citizens in this social context. Additionally, in a friendly interaction, it is better to be viewed as competent than incompetent, so we can suppose that both players prefer to be and interact with competent people. Of course, the goals of a politician at a barbecue are often different from those of private citizens (i.e., securing votes vs making friends), and friendliness/likeability is a property that is very highly valued in American presidents (Teven 2008). Furthermore, intelligence and competence are properties that have traditionally been valued in American presidents, and this is particularly the case with Obama (Alim and Smitherman 2012). So we might suppose that Obama wishes to appear competent and (above all) friendly, regardless of who he is interacting with.[17] An example of a value function satisfying Obama's preferences is shown in Table 6.

Note that S's preferences on persona selection (μ) are generally constrained by aspects of S's experiences. For example, our experiences created by external social structures (based on gender, class, ethnicity etc.) influence our internal dispositions, which go on to affect which identities we desire to construct (Bourdieu 1972) and ultimately which sociolinguistic variants we use. Likewise, properties of our bodies put enormous constraint our persona selection: as Butler (1993, x) says, it is not the case that "one [wakes] in the morning, peruse[s] the closet [...] for the gender of choice,

[16] Note that the fact that the persona selection process is often not consciously available to the speaker does not undermind the idea that they play a role in it: we have known since Antiquity (Finger 2001) that much (if not most) of our reasoning and planning is unconscious, and many areas of cognition have been argued to involve unconscious decision making (see Dennett 1993; Graziano 2013; Dehaene 2014, for overviews). In fact, there are even some philosophers (Carruthers and Veillet 2011; Carruthers 2017) for whom none of our thoughts are actually conscious.

[17] Establishing corresponding values for Trump requires intense, difficult ethnographic work...

Table 6 Example of a μ function for Obama at the barbecue

Persona	π	$\mu(\pi)$
Cool guy	{competent, friendly}	2
Stern leader	{competent, aloof}	1
Doofus	{incompetent, friendly}	1
Arrogant asshole	{incompetent, aloof}	0

don[s] that gender for the day, and then restor[s] the garment to its place at night." However, if our bodies change, then properties of μ can change too, and this can affect our language. For example, Calder (2018) performed a study of /s/ fronting based on sociolinguistic interviews with SoMA (San Francisco) drag queens while they were transforming from male bodied individuals to queens. Calder (2018, 45) reports that "as the queens transform into feminine drag, their pronunciation of /s/ gets both backer and louder … lower centers of gravity and higher amplitude are being used to index a stronger type of femininity when the queens are in drag." Thus, as their bodies change, new personae open up to the queens, which is then reflected in their speech.

The utility function for Obama for persona selection (μ in Table 6) can very naturally be mapped onto a probability distribution over personae (P_{PERS}) by means of the non-deterministic Soft max choice rule (14), which again involves a parameter α'.

(14) $$P_{\text{PERS}}(\pi; \mu) = \frac{exp(\alpha' \times \mu(\pi))}{\sum_{\pi' \in \text{PERS}} exp(\alpha' \times \mu(\pi'))}$$ **Soft max choice rule**

Taking into account possible variation in persona selection, we can then recover the predicted probability for a speaker using a particular variant m ($\mathcal{P}_S(m)$) as the sum of the probabilities that S will pick some persona and use m to try to construct it, as shown in (15). If, following the Obama barbecue example in the previous section, we set both α and α' to 6, we make roughly the same prediction as above, since Obama most highly values the {competent, friendly} persona in all cases: $\mathcal{P}_{Obama}(\textit{-in'}) \approx \mathbf{0.689}$ and $\mathcal{P}_{Obama}(\textit{-ing}) \approx \mathbf{0.311}$.

(15) $$\mathcal{P}_S(m) = \sum_{\pi} P_{\text{PERS}}(\pi; \mu) \times P_S(m|\pi)$$

Although incorporating persona selection into the game does not change the main results of the previous section, taking into account variation in personae provides a way to formally realize the idea inherent in TW and in many other sociolinguistic theories (discussed in Sect. 3) that both intra-speaker (style shifting) and inter-speaker (stratification) variation can be derived from the same basic principles. Consider again Labov (1966)'s result concerning class-based stratification of (ING) in New York City, particularly what we find in the sociolinguistic interview portion of the study reproduced in Table 7. How can a pattern such as this arise from the combination of the social meanings of the variants, speaker values and conjectures about listener prior beliefs?

Following the line of explanation in Eckert (2000, 2012), I suggest that the key to a social meaning-based account of stratificational patterns lies in the idea that speakers of different social classes differ with respect to their preferences over personae, i.e.

Table 7 (ING) by social class (casual style) in New York City

Social class	Approx. % *-in'*
Upper middle class	10
Lower middle class	40
Working class	50
Lower class	80

Table 8 Class-differentiated value functions (inspired by Bourdieu and Lamont)

Persona	μ_{WC}	μ_{LMC}	μ_{UMC}
{competent, friendly}	2	1	2
{competent, aloof}	1	1	2
{incompetent, friendly}	2	1	1
{incompetent, aloof}	1	1	1

their value functions (μ). An abundance of work in sociology (such as Bourdieu and Passeron 1970; Bourdieu 1979; Gans 1974; Lamont 1992, 2009, among others) has detailed how individuals of different education levels and occupations value different kinds of properties in themselves and in others. For example, in a study of symbolic boundaries in working and lower middle class culture in the United States and in France, Lamont (2009) shows that, in semi-structured interviews, working class participants expressed admiration of properties such as *interpersonal altruism, generosity* and *collective solidarity*, much more frequently than the lower middle class participants (see Lamont 2009, Table on p. 21). Furthermore, building on Bourdieu (1979)'s pioneering work on the relationship between social class and taste, Lamont (1992) shows that upper middle class participants (especially those with an elite university education) are more likely to value properties such as *intelligence* and *sophistication*[18] than individuals without a university education.

With these observations in mind, we can distinguish between three classes of speakers with three different value functions: a *working class* value function (μ_{WC}), which values friendly personae over non-friendly ones; an *upper middle class* value function (μ_{UMC}), which values competence/educated personae over uneducated personae; and, a *lower middle class* value function (μ_{LMC}), which has no great preference for friendliness or competence/education. These value functions are shown in Table 8.

One of the properties of the classic Labovian sociolinguistic interview is that the interviewer is not typically close with the person being interviewed. In fact, in many studies in this tradition, participants are interviewed by complete strangers (Tagliamonte 2006). Thus, when modelling speaker behaviour in a stratified sociolinguistic corpus, we might assume that speakers hypothesize that their interlocutors have no particularly strong prior beliefs about them and treat the function Pr as uniform over personae.[19]

[18] Although there are important differences between how exactly the French and American elites cash out these terms in Lamont (1992).

[19] This is undoubtedly an idealization, since studies have also shown that listeners bring their ideological baggage even to the interpretation of strangers' linguistic performances (see Campbell-Kibler 2010;

Table 9 SMG model prediction (uniform Pr; values from Table 8)

Social class	Approx. % *-in'*
Upper middle class	33
Lower middle class	50
Working class	66

Since we have both the functions Pr and μ for the different categories of speakers, we can now calculate the predictions that the SMG models make for the distribution of (ING). These are found in Table 9, and they reproduce the main lines of the pattern found in Labov (1966).

I therefore conclude that SMG models can provide a formal unified analysis of style shifting and social stratification, which is a main goal of Third Wave variationist sociolinguistics.

4.3.1 The relationship between persona and variant selection

Adding persona selection to the model raises questions about whether there is some dependency between persona selection and variant selection. In the model presented above, the choice of persona and the choice of variant are **sequential**: first the speaker picks the persona π according to their μ and then they calculate the approximately optimal variant to construct π: $P_S(m|\pi)$. However, it is possible that our persona selection is influenced not only by our dispositions, but also by how likely we think we are to be successful in constructing certain personae.[20] For example, Obama is known for being particularly versatile in his persona construction and he attributes his ability to variably construct a black or white identity to his mastery of multiple dialects, which is a product of his background. He says,

> You go to the cafeteria … and the black kids are sitting here, white kids are sitting there, and you've got to make some choices. For me, basically I could run with anybody. Luckily for me, largely because of growing up in Hawai'i, there wasn't that sense of sharp divisions. Now, by the time I was negotiating environments where there were those kinds of sharp divisions, I was already confident enough to make my own decisions. It became a matter of being able to speak different dialects. That's not unique to me. Any black person in America who's successful has to be able to speak several different forms of the same language … It's not unlike a person shifting between Spanish and English. (Alim and Smitherman 2012, 1)

In contrast to Obama, other speakers are comfortable with only one set of variants may avoid selecting personae that can only be constructed using a different set, if they they think their construction is unlikely to be successful.

Footnote 19 continued
Levon 2007, 2014, among many others); however, the context of the Labovian sociolinguistic interview is probably one in which listeners' priors would be as close to uniform as we would ever find in real communication.

[20] I thank Michael Franke for very helpful discussion of this point.

To allow for interactions between variant selection and persona selection, we need to have the speaker jointly select a persona and a variant, as shown in (16). This speaker function makes different predictions from (13), particularly with costly messages; however, I leave their detailed exploration to future work.

$$(16) \qquad P_S(m, \pi; \mu) = \frac{exp(\alpha \times U_S(\pi, m)) \times exp(\alpha' \times \mu(\pi))}{\sum_{m'} \sum_{\pi'} exp(\alpha \times U_S(\pi', m')) \times exp(\alpha' \times \mu(\pi'))}$$

4.4 Predicting listener behaviour

Given the model of the speaker's behaviour that we have developed above in (15) and (14), what should the listener do when they hear a variant?[21] Since persona selection is determined by S's values (μ), if the listener knows μ, then, to infer the speaker's persona, they would simply combine S's likelihood of picking a persona (P_{PERS}) with S's likelihood of constructing a particular persona given the message that they heard (P_S) (13).

$$(17) \qquad P_L(\pi | m; \mu) = \frac{P_{\text{PERS}}(\pi; \mu) \times P_S(m|\pi)}{\sum_{\pi'} P_{\text{PERS}}(\pi'; \mu) \times P_S(m|\pi')}$$

Of course, it is not really realistic to think that, when we interact with someone, we know the details of their value function. So how do we, as listeners, go about interpreting the linguistic performances of our interlocutors?

In answering this question, I will build on aspects of Erving Goffman's *Expression Games* model of self-presentation and deception (Goffman 1970), which is highly influenced by work in game theory.[22] Goffman observes that when we are on the receiving end of a performance, we will have different ways of interpreting it depending on how much strategy we think has been involved in creating it. The first interpretative move is what Goffman calls a *naive* move, and it involves the listener assuming that the speaker's strategy in the context does not enter into play into how they present themselves.

In terms of the model developed here, this would involve the listener simply combining the likelihood of *S* using a variant to construct a persona with their prior beliefs about S's persona, and not reasoning about S's values. This Bayesian-style interpretation rule is shown in (18).

$$(18) \qquad P_L(\pi | m) = \frac{Pr(\pi) \times P_S(m|\pi)}{\sum_{\pi'} Pr(\pi') \times P_S(m|\pi')} \qquad \textbf{Naive listener}$$

Although this should be investigated empirically, my impression is that this is the most frequent mode of social interpretation. In order to exemplify how this model works, we will model a subset of Podesva et al. (2015)'s results on the interpretation of (un)released /t/ by American politicians. Following (simplified) Eckert (2008),

[21] I am grateful to Michael Franke for detailed discussion of the different listener models at play in this framework.

[22] *Expression games* was written after Goffman spent a sabbatical year at Harvard working with the Nobel Prize winning game theorist Thomas Schelling (Manning 1992). I think it would be worthwhile to further study Goffman's model in light of the mathematical model presented here, and in light of IBR signalling games more generally. However, since it deals with the dynamics of deception rather than true identity construction, many of the aspects of Expression Games are not relevant here.

Table 10 Uniform prior beliefs about Rice

π	{artic., aloof}	{artic., friend}	{inartic., aloof}	{inartic., friend}
Pr(π)	0.25	0.25	0.25	0.25

Table 11 $P_{Rice}(m|\pi)$ (all values for α)

m	{artic., aloof}	{artic., friend}	{inartic., aloof}	{inartic., friend}
t^h	1	0.5	0.5	0
ɾ	0	0.5	0.5	1

we assume an Eckert field for released and flapped/unreleased /t/ as shown in (19) which is similar to the fields of (ING), only that we replace *(in)competence* with *(in)articulateness*.[23]

(19) Eckert fields associated with /t/
   a. [t^h] = {articulate, aloof}
   b. [ɾ/t] = {inarticulate, friendly}

For the illustration, we will consider interpretations related to articulateness/authoritativeness and friendliness assigned to three different politicians: Condoleeza Rice, Nancy Pelosi, and George W. Bush.

I propose that the different interpretations of (non)released /t/ assigned to different politicians arise as the result of L having different kinds of prior beliefs about each politician. Condoleeza Rice is not a particularly (in)famous politician, so we might assume that listeners' in Podesva et al.'s study do not have particularly strong beliefs about her articulateness or friendliness. We will represent this belief as a uniform prior distribution over personae (Table 10). Thus, before hearing Rice say anything, the listener has a 0.5 belief that she is articulate and a 0.5 belief that she is friendly.

We obtain the likelihood of Rice trying to construct a persona π with a variant m in the way described above, and the results are shown in Table 11.

And using the weights in Table 10, the weights in Table 11, and our interpretation rule (18), we obtain a measure of the listener's beliefs concerning Rice's persona after hearing a variant. As shown in Table 12, after hearing a released /t/, L's posterior belief that Rice instantiates an articulate persona is 0.75; whereas, it is only 0.25 if she flaps.[24]

[23] Indeed, Eckert (following others) proposes that the indexical fields of [t^h] and *-ing* share some properties, but also differ on other properties (that are not relevant in this paper). The same goes for the fields of [ɾ/t] and *-in'*.

[24] Note that given the simple indexical fields and personae presented here, I do not have an immediate account for why Podesva et al. (2015) find the *opposite* pattern with Obama than with Rice and Edwards. A promising line of analysis would be to go deeper into the different classes of personae that Obama, Edwards and Rice are likely to instantiate (see Alim and Smitherman 2012) and to complicate the meanings of /t/; however, I leave working this out in detail to future work.

Table 12 L's predicted interpretation of π given m for Rice

m	{artic., aloof}	{artic., friend}	{inartic., aloof}	{inartic., friend}
t^h	0.5	0.25	0.25	0
ɾ	0	0.25	0.25	0.5

Table 13 L thinks that Pelosi is not articulate

π	{artic., aloof}	{artic., friend}	{inartic., aloof}	{inartic., friend}
$Pr(\pi)$	0.05	0.05	0.45	0.45

Table 14 L's predicted interpretation of π given m for Pelosi

m	{artic., aloof}	{artic., friend}	{inartic., aloof}	{inartic., friend}
t^h	0.09	0.09	0.82	0
ɾ	0	0	0.03	0.96

On the other hand, the SMG framework predicts very different results if listeners have different prior beliefs. Consider the case of Nancy Pelosi: in Podesva et al.'s study, "listeners appear to be less likely to associate released /t/ with competence-based meanings in Pelosi's speech—such as articulateness, intelligence, or authoritativeness." (Podesva et al. 2015, 79) According to Podesva et al., this is because there is a 'clash' between speakers' beliefs about her (that she is not genuinely authoritative/articulate/intelligent) and what they perceive her as trying to do: sound authoritative/articulate/intelligent. In the SMG framework, we can represent this situation as one in which L's prior beliefs about Pelosi (Table 13) affect how her used of released /t/ is interpreted.

The predictions for listener interpretations of Pelosi's linguistic performances, given Pr in Table 13 are shown in Table 14. There is no predicted difference in articulateness between hearing [t^h] and [ɾ] (both variants trigger almost exclusively inarticulate personae); however, there is predicted to be a difference in terms of friendliness: [t^h] results in an almost certain attribution of the {inarticulate, aloof} persona; whereas, [ɾ] results in an almost certain attribution of the {inarticulate, friendly} persona.

Finally, we can consider the cases of George W. Bush and Hilary Clinton. For these politicians, Podesva et al. say,

> The speech of two remaining politicians—George W. Bush and Hillary Clinton—was not judged to sound different in the released and unreleased guises for any of the nine attributes, either word-medially or word-finally. 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. (Podesva et al. 2015, 80)

Table 15 L is almost sure that Bush is an asshole

π	{artic., aloof}	{artic., friend}	{inartic., aloof}	{inartic., friend}
$Pr(\pi)$	0.01	0.01	0.97	0.01

Table 16 L's predicted interpretation of π given m for Bush

m	{artic., aloof}	{artic., friend}	{inartic., aloof}	{inartic., friend}
t^h	0.02	0.01	0.97	0
ɾ	0	0.01	0.97	0.02

The 'bulletproofing' effect that strong prior beliefs can have on sociolinguistic interpretation is predicted by the model. In line with the Podesva et al. quotation above, suppose, that L has an extreme belief that Bush is an *arrogant asshole*-type: {inarticulate, aloof}, shown in Table 15.

As shown in Table 16, it is predicted that using different variants will have no effect on L's beliefs about Bush. In other words, Bush can try as much as he likes to use his linguistic resources to construct a different identity, he will be 'stuck' being viewed as {inarticulate, aloof}.

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.[25]

The other move that the listener can make in Goffman's *Expression Games* is what he calls an *uncovering move*. This is when the listener takes into account their hypotheses about the speaker's strategy in persona construction in the context (Goffman 1970, 17). Set in this framework, the listener makes a hypothesis about the speaker's values in the form of a probability distribution over value functions, $P(\mu)$, and performs joint inference over S's persona and μ. This is shown in (20).[26]

$$P_L(\pi, \mu|m) = \frac{P(\mu) \times P_{\text{PERS}}(\pi;\mu) \times P_S(m|\pi)}{\sum_{\pi'} \sum_{\mu} P_{\text{PERS}}(\pi';\mu) \times P_S(m|\pi')} \quad \textbf{Uncovering listener} \tag{20}$$

The naive and uncovering listener functions do not make exactly the same predictions in all contexts;[27] however, teasing them apart empirically would require detailed experimental study, which I leave to future work.

[25] Furthermore, note that the framework also has a straightforward account of the phenomenon of language crossing (Rampton 1995) mentioned in Sect. 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 a (usually desirable) subset of the (un)desirable properties usually attributed to members of the second group.

[26] Of course, for (20) to be realistic, there must be a certain relatively small set of value functions that are salient in the community; however, as I suggested above in the discussion of macro-level patterns of social stratification, the social structure of our communities and our place within them contributes to shaping our values, so it is reasonable to think that salient value functions might emerge from social structure. I believe this is a fruitful area of future collaboration between formal linguists, sociologists and anthropologists.

[27] For example, if the listener has much uncertainty about the μ function.

5 Conclusion

In this paper I presented a new formal model of the social/strategic aspects of sociolinguistic variation, one that analyses 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 construct identities and linguistic styles, I introduced the social meaning games (SMG) framework, which is a formalization of the Third Wave approach to variation as a kind of signalling game with a Bayesian-style approach to interpretation. 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 sociolinguistic interpretation based on listeners' prior beliefs about the speaker.

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 persona-based inferences 'live' in relation to other kinds of non-at-issue content such as presuppositions, scalar implicatures and expressive content. As such, the natural next step would be to integrate the SMG approach into a broader model of grammar and language processing. Additionally, in this paper, I have only treated individual variants (e.g. (ING) and /t/ release), so a next related step would be to investigate 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 make-up, 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, identity and stylistic performance.

Acknowledgements This research has been partially supported by the program "Investissements d'Avenir" overseen by the French National Research Agency, ANR-10-LABX-0083 (Labex EFL), and a fellowship from the Center for the Study of Language and Information at Stanford University. I thank Eric Acton, Leon Bergen, Judith Degen, Chantal Gratton, Erez Levon, Ellin McCready, Devyani Sharma, Sali Tagliamonte, audiences at UCL, Institut Jean Nicod, LLF Paris-Diderot, Stanford, UCLA, UCSC and *NWAV45*, and especially Penny Eckert, Dan Lassiter and Michael Franke for very helpful comments and discussions. All errors are my own.

References

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

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

Acton, E. K., & Potts, C. (2014). That straight talk: Sarah Palin and the sociolinguistics of demonstratives. *Journal of Sociolinguistics*, *18*(1), 3–31.

Alim, H. S., & Smitherman, G. (2012). *Articulate while Black: Barack Obama, language, and race in the US*. Oxford: Oxford University Press.

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

Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. In J. Kulas, J. H. Fetzer, & T. L. Rankin (Eds.), *Philosophy, language, and artificial intelligence* (pp. 241–301). Dordrecht: Springer.

Beaton, M. E., & Washington, H. B. (2015). Slurs and the indexical field: The pejoration and reclaiming of *favelado* 'slum-dweller'. *Language Sciences*, *52*, 12–21.

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*. Ph.D. thesis, University of Chicago.

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

Bergen, L., Levy, R., & Goodman, N. D. (2016). Pragmatic reasoning through semantic inference. *Semantics & Pragmatics*, *9*, 20.

Bourdieu, P. (1972). *Esquisse d'une théorie de la pratique*. Geneva: Librairie Droz.

Bourdieu, P. (1977). The economics of linguistic exchanges. *Social Science Information*, *16*(6), 645–668.

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

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

Bourdieu, P., & Boltanski, L. (1975). Le fétichisme de la langue. *Actes de la Recherche en Sciences Sociales*, *1*, 2–32.

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

Bucholtz, M. (1996). Geek the girl: Language, femininity, and female nerds. In N. Warner et al. (Eds.), *Gender and belief systems: Proceedings of the fourth Berkeley Women and Language Conference* (pp. 119–131). Berkeley: Berkeley Women and Language Group.

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: 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.

Burnett, H. (2017). Sociolinguistic interaction and identity construction: The view from game-theoretic pragmatics. *Journal of Sociolinguistics*, *21*, 238–271.

Butler, J. (1993). *Bodies that matter: On the discursive limitations of sex*. New York: Routledge.

Calder, J. (2018). The fierceness of fronted /s/: Linguistic rhematization through visual transformation. *Language in Society*. https://doi.org/10.1017/S004740451800115X.

Cameron, D. (2016). Misogyny by the numbers. In *Language: A feminist guide*. https://debuk.wordpress.com/2016/05/29/.

Campbell-Kibler, K. (2006). *Listener perceptions of sociolinguistic variables: The case of (ING)*. Ph.D. 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.

Campbell-Kibler, K. (2009). The nature of sociolinguistic perception. *Language Variation and Change*, *21*(1), 135–156.

Campbell-Kibler, K. (2010). Sociolinguistics and perception. *Language and Linguistics Compass*, *4*(6), 377–389.

Carruthers, P. (2017). *The illusion of conscious thought*. London: Bloomsbury Press.

Carruthers, P., & Veillet, B. (2011). The case against cognitive phenomenology. In T. Bayne & M. Montague (Eds.), *Cognitive phenomenology* (pp. 35–56). Oxford: Oxford University Press.
Cerulo, K. A. (1997). Identity construction: New issues, new directions. *Annual Review of Sociology*, *23*(1), 385–409.
Charmaz, K. (2011). Grounded theory methods in social justice research. *The SAGE handbook of qualitative research*, *4*, 359–380.
Cheshire, J. (1982). *Variation in an English dialect: A sociolinguistic study*. Cambridge: Cambridge University Press.
Clark, R. L. (2014). *Meaningful games: Exploring language with game theory*. Cambridge, MA: MIT Press.
Degen, J. & Franke, M. (2012). Optimal reasoning about referential expressions. In *Proceedings of SemDIAL 2012*, (pp. 2–11).
Degen, J., Franke, M., & Jäger, G. (2013). Cost-based pragmatic inference about referential expressions. In *Proceedings of the 35th annual conference of the cognitive science society* (pp. 376–381).
Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive Science*, *39*(4), 667–710.
Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. London: Penguin.
Dekker, P., & Van Rooy, R. (2000). Bi-directional optimality theory: An application of game theory. *Journal of Semantics*, *17*(3), 217–242.
Dennett, D. C. (1993). *Consciousness explained*. London: Penguin.
DeRue, D. S., & Ashford, S. J. (2010). Who will lead and who will follow? A social process of leadership identity construction in organizations. *Academy of Management Review*, *35*(4), 627–647.
Drager, K. K. (2015). *Linguistic variation, identity construction and cognition*. Berlin: Language Science Press.
Dror, M., Granot, D., & Yaeger-Dror, M. (2013). Speech variation, utility, and game theory. *Language and Linguistics Compass*, *7*(11), 561–579.
Dror, M., Granot, D., & Yaeger-Dror, M. (2014). Teaching & learning guide for speech variation, utility, and game theory. *Language and Linguistics Compass*, *8*(6), 230–242.
Dutton, J. E., Roberts, L. M., & Bednar, J. (2010). Pathways for positive identity construction at work: Four types of positive identity and the building of social resources. *Academy of Management Review*, *35*(2), 265–293.
Eckert, P. (2000). *Language variation as social practice: The linguistic construction of identity in Belten High*. Hoboken, NJ: Wiley-Blackwell.
Eckert, P. (2005). Variation, convention, and social meaning. In In *Annual meeting of the Linguistic Society of America, Oakland, CA* (Vol. 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.
Finger, S. (2001). *Origins of neuroscience: A history of explorations into brain function*. New York, NY: Oxford University Press.
Frank, M. C., & 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*. Ph.D. thesis, Universiteit van Amsterdam.
Franke, M., & Jäger, G. (2012). Bidirectional optimization from reasoning and learning in games. *Journal of Logic, Language and Information*, *21*(1), 117–139.
Franke, M., & Jäger, G. (2016). Probabilistic pragmatics, or why Bayes' rule is probably important for pragmatics. *Zeitschrift für Sprachwissenschaft*, *35*(1), 3–44.
Gans, H. (1974). *High culture and popular culture: An analysis and evaluation of taste*. Nova York: Basic Books.
Gintis, H. (2000). *Game theory evolving: A problem-centered introduction to modeling strategic behavior*. Princeton, NJ: Princeton University Press.
Goffman, E. (1961). *Encounters: Two studies in the sociology of interaction*. Indianapolis, IN: Bobbs-Merrill.
Goffman, E. (1967). *Interaction ritual: Essays on face-to-face interaction*. Oxford: Aldine.
Goffman, E. (1970). *Strategic interaction* (Vol. 1). Philadelphia, PA: University of Pennsylvania Press.
Goodman, N. D., & Lassiter, D. (2014). *Probabilistic semantics and pragmatics: Uncertainty in language and thought. Handbook of Contemporary Semantic Theory*. Hoboken, NJ: Wiley-Blackwell.

Gratton, C. (2016). Resisting the gender binary: The use of (ING) in the construction of non-binary transgender identities. *University of Pennsylvania Working Papers in Linguistics*, *22*, 7.
Graziano, M. S. (2013). *Consciousness and the social brain*. Oxford: Oxford University Press.
Hacking, I. (1999). *The social construction of what?*. Cambridge, MA: Harvard University Press.
Hardaker, C. (2016). Misogyny machines, and the media: Or how science should not be reported. https://wp.lancs.ac.uk/drclaireh/2016/05/27/misogyny-machines-and-the-media-or-how-science-shouldnot-be-reported/.
Harsanyi, J. C. (1967). Games with incomplete information played by 'Bayesian' players, I- III, Part I. The basic model. *Management Science*, *14*(3), 159–182.
Harsanyi, J. C. (1968a). Games with incomplete information played by 'Bayesian' players, Part II. Bayesian equilibrium points. *Management Science*, *14*(5), 320–334.
Harsanyi, J. C. (1968b). Games with incomplete information played by 'Bayesian' players, Part III. The basic probability distribution of the game. *Management Science*, *14*(7), 486–502.
Hazen, K. (2006). In/ing variable. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (pp. 580–581). Amsterdam: Elsevier.
Heim, I. (1982). *The semantics of definite and indefinite noun phrases*. Ph.D. thesis, University of Massachussets, Amherst.
Houston, A. (1985). *Continuity and change in English morphology: The variable (ING)*. Ph.D. thesis, University of Pennsylvania.
Jäger, G. (2011). Game-theoretical pragmatics. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of logic and language* (pp. 467–491). Amsterdam: Elsevier.
Kaplan, D. (1999). *The meaning of 'ouch' and 'oops': Explorations in the theory of meaning as use*. Los Angeles, CA: UCLA.
Keenan, E. L., & Stavi, J. (1986). A semantic characterization of natural language determiners. *Linguistics and Philosophy*, *9*(3), 253–326.
Kelly, G. J. (2014). Discourse practices in science learning and teaching. *Handbook of Research on Science Education*, *2*, 321–336.
Kiesling, S. F. (1998). Men's identities and sociolinguistic variation: The case of fraternity men. *Journal of Sociolinguistics*, *2*(1), 69–99.
Kiesling, S. (2009). Style as stance: Can stance be the primary explanation for patterns of sociolinguistic variation? In A. Jaffe (Ed.), *Sociolinguistic perspectives on stance* (pp. 171–194). Oxford: Oxford University Press.
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*. Cambridge: Cambridge University Press.
Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.
Labov, W. (2012). *Dialect diversity in America: The politics of language change*. Charlottesville, VA: University of Virginia Press.
Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *The Journal of Abnormal and Social Psychology*, *60*(1), 44.
Lamont, M. (1992). *Money, morals, and manners: The culture of the French and the American upper-middle class*. Chicago, IL: University of Chicago Press.
Lamont, M. (2009). *The dignity of working men: Morality and the boundaries of race, class, and immigration*. Cambridge, MA: Harvard University Press.
Lassiter, D. (2008). Semantic externalism, language variation, and sociolinguistic accommodation. *Mind & Language*, *23*(5), 607–633.
Lassiter, D., & Goodman, N. D. (2013). Context, scale structure, and statistics in the interpretation of positive-form adjectives. *Semantics and Linguistic Theory*, *23*, 587–610.
Lassiter, D., & Goodman, N. D. (2015). Adjectival vagueness in a Bayesian model of interpretation. *Synthese*, *194*, 1–36.
Legendre, G., Miyata, Y., & Smolensky, P. (1990). *Harmonic grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations*. (Technical report 92–16.) Boulder: Institute of Cognitive Science, University of Colorado.
Levon, E. (2007). Sexuality in context: Variation and the sociolinguistic perception of identity. *Language in Society*, *36*(04), 533–554.
Levon, E. (2014). Categories, stereotypes, and the linguistic perception of sexuality. *Language in Society*, *43*(05), 539–566.

Lewis, D. (1969). *Convention*. Cambridge, MA: Harvard University Press.
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.
Manning, P. (1992). *Erving Goffman and modern sociology*. Hoboken, NJ: Wiley.
McConnell-Ginet, S. (2011). *Gender, sexuality, and meaning: Linguistic practice and politics*. Oxford: Oxford University Press.
McCready, E. (2012). Emotive equilibria. *Linguistics and Philosophy*, *35*(3), 243–283.
McCready, E., Asher, N., & Paul, S. (2013). Winning strategies in politeness. In Y. Motomura, A. Butler & D. Bekki (Eds.), *New frontiers in artificial intelligence. JSAI-isAI 2012*. (pp. 87–95). Berlin: Springer.
Miehls, D., & Moffatt, K. (2000). Constructing social work identity based on the reflexive self. *British Journal of Social Work*, *30*(3), 339–348.
Montague, R. (1973). The proper treatment of quantification in ordinary English. In K. J. J. Hintikka, J. M. E. Moravcsik, & P. Suppes (Eds.), *Approaches to natural language* (pp. 221–242). Dordrecht: Reidel.
Moore, E., & Podesva, R. (2009). Style, indexicality, and the social meaning of tag questions. *Language in Society*, *38*(04), 447–485.
Mühlenbernd, R. (2013). *Signals and the Structure of Societies*. Ph.D. thesis, Universität Tübingen.
Mühlenbernd, R., & Franke, M. (2012). Signaling conventions: Who learns what where and when in a social network. In *Proceedings of EvoLang IX* (pp. 242–249).
Nguyen, D., Doğruöz, A. S., Rosé, C. P., & de Jong, F. (2016). Computational sociolinguistics: A survey. *Computational Linguistics*, *42*, 537–593.
Ochs, E. (1992). Indexing gender. In A. Duranti & C. Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 335–358). Cambridge: Cambridge University Press.
Ochs, E. (1993). Constructing social identity: A language socialization perspective. *Research on Language and Social Interaction*, *26*, 287–306.
Osborne, M. J., & Rubinstein, A. (1994). *A course in game theory*. Cambridge, MA: MIT Press.
Oushiro, L. (2015). Social meanings of (-r) in Sao Paulo: A computational approach for modeling the indexical field. Paper presented at *New Ways of Analyzing Variation 44*. University of Toronto.
Peters, S., & Westerståhl, D. (2006). *Quantifiers in language and logic*. Oxford: Oxford University Press.
Podesva, R. (2004). On constructing social meaning with stop release bursts. Paper presented at *Sociolinguistics Symposium* 15. Newcastle upon Tyne.
Podesva, R. (2006). *Phonetic detail in sociolinguistic variation: Its linguistic significance and role in the construction of social meaning*. Ph.D. 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., & 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.
Quinley, J., & Mühlenbernd, R. (2012). Conquest, contact, and convention: Simulating the norman invasion's impact on linguistic usage. *Proceedings of BRIMS*, *2012*, 113–118.
Rampton, B. (1995). Language crossing and the problematisation of ethnicity and socialisation. *Pragmatics*, *5*(4), 485–513.
Rickford, J., & Closs Traugott, E. (1985). Symbol of powerlessness and degeneracy, or symbol of solidarity and truth? Paradoxical attitudes towards pidgins and creoles. In S. Greenbaum (Ed.), *The English Language Today* (pp. 252–261). Oxford: Pergamon.
Roberts, B. (1991). Music teacher education as identity construction. *International Journal of Music Education*, *18*(1), 30–39.
Rosenbaum, D. A., Chapman, K. M., Weigelt, M., Weiss, D. J., & van der Wel, R. (2012). Cognition, action, and object manipulation. *Psychological Bulletin*, *138*(5), 924.
Rosenbaum, D. A., Cohen, R. G., Jax, S. A., Weiss, D. J., & Van Der Wel, R. (2007). The problem of serial order in behavior: Lashley's legacy. *Human Movement Science*, *26*(4), 525–554.
Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423.
Silverstein, M. (1979). Language structure and linguistic ideology. In P. R. Clyne, W. F. Hanks, & C. L. Hofbauer (Eds.), *The elements: A parasession on linguistic units and levels* (pp. 193–247). Chicago, IL: Chicago Linguistic Society.
Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, *23*(3), 193–229.

Smith, E., Hall, K. C., & Munson, B. (2010). Bringing semantics to sociophonetics: Social variables and secondary entailments. *Laboratory Phonology*, *1*(1), 121–155.
Smolensky, P., & Legendre, G. (2006). *The harmonic mind: From neural computation to optimality-theoretic grammar (Cognitive architecture)* (Vol. 1). Cambridge, MA: MIT Press.
Stalnaker, R. (1978). Assertion. In P. Cole (Ed.), *Syntax and semantics* (Vol. 9, pp. 315–332). New York: Academic Press.
Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.
Tagliamonte, S. A. (2006). *Analysing sociolinguistic variation*. Cambridge, MA: Cambridge University Press.
Tamminga, M. (2014). *Persistence in the production of linguistic variation*. Ph.D. thesis, University of Pennsylvania.
Taylor, D. E. (2000). The rise of the environmental justice paradigm: Injustice framing and the social construction of environmental discourses. *American Behavioral Scientist*, *43*(4), 508–580.
Tenenbaum, J . B., Kemp, C., Griffiths, T . L., & Goodman, N . D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.
Teven, J. J. (2008). An examination of perceived credibility of the 2008 presidential candidates: Relationships with believability, likeability, and deceptiveness. *Human Communication*, *11*(4), 391–408.
Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society*, *1*(02), 179–195.
Tyler, J. C. (2015). Expanding and mapping the indexical field: Rising pitch, the uptalk stereotype, and perceptual variation. *Journal of English Linguistics*, *43*(4), 284–310.
van Hofwegen, J. (2017). *Everyday styles: Investigating the full scope of variation in the life of an individual speaker*. Ph.D. thesis, Stanford University.
Van Rooy, R. (2003). Being polite is a handicap: Towards a game theoretical analysis of polite linguistic behavior. In M. Tennenholtz (Ed.), *Proceedings of the 9th conference on Theoretical Aspects of Rationality and Knowledge (TARK IX)* (pp. 45–58). Bloomington, IN: Indiana University.
Varelas, M. (2012). *Identity construction and science education research: Learning, teaching, and being in multiple contexts* (Vol. 35). Berlin: Springer.
Walker, A., García, C., Cortés, Y., & Campbell-Kibler, K. (2014). Comparing social meanings across listener and speaker groups: The indexical field of spanish/s. *Language Variation and Change*, *26*(02), 169–189.
Weinreich, U., Labov, W., & Herzog, M. (1968). Empirical foundations for a theory of language change. In W. P. Lehmann (Ed.), *Directions for historical linguistics: A symposium* (pp. 95–195). Austin, TX: University of Texas Press.
Zhang, Q. (2005). A Chinese yuppie in Beijing: Phonological variation and the construction of a new professional identity. *Language in Society*, *34*(3), 431–466.
Zhang, Q. (2008). Rhotacization and the 'Beijing smooth operator': The social meaning of a linguistic variable. *Journal of Sociolinguistics*, *12*(2), 201–222.
Zhao, S., Grasmuck, S., & Martin, J. (2008). Identity construction on facebook: Digital empowerment in anchored relationships. *Computers in Human Behavior*, *24*(5), 1816–1836.