Beyond Personae: Honest Failures in Projecting Social Meaning

Christopher Davis[1(✉)] and Grégoire Winterstein[2(✉)]

[1] Department of Global Languages and Cultures, University of the Ryukyus, Nishihara, Okinawa, Japan
cmdavis@cs.u-ryukyu.ac.jp
[2] Département de linguistique, Université du Québec à Montréal, Montréal, QC, Canada
winterstein.gregoire@uqam.ca
https://cmdavis.org/,
https://professeurs.uqam.ca/professeur/winterstein.gregoire/

Abstract. We apply the concepts of honest signaling and honest failure to linguistic variation and social meaning. We argue for the importance of the acquisition cost of variants and uneven access to their normative deployment across a community of speakers. We connect our discussion to Third Wave theorizing about social meaning and argue for particular ways in which models of such meaning need to be expanded to account for honest signaling.

AQ1

Keywords: Variation · Social Meaning · Third Wave Sociolinguistics · Honest Signaling

1 Introduction

The notion of *honest signaling* is commonly used in certain domains (particularly within evolutionary biology) to discuss signals, called *honest signals*, that cannot be faked: the very fact that an agent can send an honest signal indicates that they possess the characteristic associated with the signal. As an example, Thomson gazelles (among other animals) will 'stot' in the presence of certain predators, i.e. they perform a specific form of jump, simultaneously lifting all of their feet off the ground. This behavior signals the fitness of the animal to its potential predators, and acts as a deterrent to be chosen as prey (FitzGibbon and Fanshawe, 1988), given that a predator is deterred from pursuing a fit animal. Stotting is an honest signal: one can only properly stot if one is physically fit. On the other hand, fiddler crabs use the size of one of their claws as a signal of strength (esp. when fighting for territory) and the possession of the resources needed to maintain a big claw. This signal is, however, potentially at least partially dishonest: after losing its big claw, a fiddler crab can grow another one, equally big, but which lacks the muscle tissue of the original one and is not effective in combat. Yet, the display of a big claw has the same effects on other crabs (rivals and partners),

D. Bekki et al. (Eds.): LENLS 2025, LNCS 16698, pp. 1–17, 2027.
https://doi.org/10.1007/978-3-032-31949-4_9

independently of whether it is an original or a regrown one (Lailvaux et al., 2009).

In this paper, we argue for the importance of honest signaling in natural language communication, focusing on how speakers convey social meaning via the use of alternative linguistic codes. We start with a brief introduction to social meaning and its formalization in Sect. 2, focusing for concreteness on variable pronunciations of the English suffix -ING. Section 3 introduces an example of variation, namely that between *who* and *whom*, in which not all speakers are able to deploy both variants in a normatively correct way. We propose that variation in these cases should be framed in terms of alternative linguistic codes, rather than just linguistic forms, and discuss how high acquisition cost and uneven access to a code leads to a situation in which its use and misuse give rise to honest social signals. Section 4 discusses the problems this poses for formal models of social meaning, and a potential fix, in which listener's reason about group membership on the basis of (non)normative deployment of socially meaningful variants. Section 5 discusses the relevance of honest signaling for a range of other linguistic phenomena with a connection to social meaning. Section 6 concludes.

2 Social Meaning

In the context of Third Wave Sociolinguistics (Eckert, 2018), a large body of work has focused on how speakers use language not just to share information about the world, but also to project a certain image of themselves, a class of meaning often referred to as *social meaning* (Podesva, 2011; Hall-Lew et al., 2021). Thus, parts of the linguistic code can be used by speakers as indices of certain properties, with the overarching goal of projecting a *persona*, understood as a maximal consistent set of such properties (Eckert, 2008; Burnett, 2023). We begin here with a brief example of the typical sort of linguistic phenomenon associated with social meaning, namely phonetic variation, and then outline a recent proposal for the formalization of the mechanisms that underlie social meaning.

2.1 Social Meaning and Variation

The idea of social meaning is tied to the study of linguistic variation. A well-known and much studied case of socially meaningful variation is the pronunciation of the -ING suffix in English, which has the variant pronunciations *-ing* ([ɪŋ]) and *-in* ([ɪn]) (Campbell-Kibler 2006, 2007, 2008; Eckert 2008; Labov 2012; a.o.). These two pronunciations have no difference in terms of their truth-conditional or grammatical properties: they both denote the gerund form of a verb. It has however been observed that the deployment of these alternative pronunciations has different effects on the listener's perception of the speaker's persona, effects that are complex and context-dependent. Generalizing, however, we can say that (among many other things) the use of *-ing* is perceived as more *educated* and

articulate than the use of *-in*. (Burnett, 2023). From a Third Wave Sociolinguistics perspective, both variants are part of the linguistic system of a community of speakers, and part of the repertoire of an individual speaker. It is the *choice* (conscious or not) of one variant or another that serves to index particular properties, and thus to signal certain aspects of a speaker's persona. Within the usual picture, such variants are taken to be freely available to a given speaker, who can then strategically deploy them in the construction of a desired persona.

A famous illustration of the impact of the -ING suffix is its use by former president of the United States Barack Obama. Labov (2012) discusses how Obama varies in two ways in his use of the *-ing* and *-in* variants, which Labov refers to as *style shifting*. First, in a given situation, Obama uses both variants across multiple utterances, and second, the proportion of *-ing* vs. *-in* varies across different types of situations, with *-ing* being deployed more frequently in formal situations and *-in* more frequently casual situations. Thus at a casual event like the White House barbecue party, Obama uses *-in* rather frequently (about 72% of the time), a proportion which drops to 33% when answering questions to journalists after that event, and drops even more severely at more formal evens such as the Democratic National Convention, where *-ing* is used almost exclusively (about 97% of the time).

2.2 Formal Models

The way social meaning is deployed and interpreted has been modelled formally in a game-theoretic framework by Burnett (2019, 2023) with the use of *Social Meaning Games* (SMGs). A clear upshot of SMGs is that they give a formal way to model variation at all levels, both within and across speakers, using a single framework. As shown by Burnett (2023), by setting the parameters of the models at the appropriate value, the model's prediction can match observed levels of variation such as those displayed by Obama in various settings.

We will not in this paper go into the details of how such games are formalized (and refer the interested reader to the relevant sources). Instead, we focus on certain of their aspects, keeping in mind that Burnett discusses several versions of her model, varying in complexity depending on the phenomena she wishes to capture. Here we give a minimal introduction to show how the model applies to the case at hand.

On the speaker side, SMGs involve a value function over personae (noted μ) that represents the preference of the speaker regarding each persona π they can be assigned to, and that Burnett uses as a proxy for a speaker's agency. A toy example of such a function, borrowed from Burnett (2023), is given in table 1. There, only four persona, built on two dimensions (competent/incompetent, friendly/aloof), are considered. In this example, the speaker most prefers the *cool guy* persona, and strongly disprefers the *arrogant asshole* persona, with the other two personae sitting between.

One can map μ to a probability measure P_{pers} over personae π (e.g. using a SoftMax function as Burnett does). Having a probability measure over personae is a way to approach variation: if one only considered the valuation function μ, it

Table 1. An example of preferences over four personae

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

would be expected that rational speakers would always go for the persona they value the most, and would not vary, contrary to what is observed. P_{pers} is part of the computation of the probability that speaker S produces a particular variant v as in (1).

(1) $P_S(v) = \sum_\pi P_{pers}(\pi; \mu) \times P_S(v|\pi)$

The linguistic side of this reasoning is in the likelihood component of producing a certain variant v under the assumption that one has a certain persona π (i.e. $P_S(v|\pi)$). In the context of the -ING example, the two variants considered would be v_1 =*-ing*, and v_2=*-in*, and the production probabilities would be such that (for example) $P($*-in* $|\pi_i)$ is high for the personae that index friendliness and possibly incompetence, i.e. all the ones in Table 1 except the 'Stern leader' one.

The simplest way to model a (naive) listener is to consider that they update their belief that the speaker is projecting a persona π in typical Bayesian fashion, i.e. as in (2).

(2) $$P(\pi|v; \mu) = \frac{P_{pers}(\pi; \mu) \times P_S(v|\pi)}{\sum_{\pi'} P_{pers}(\pi'; \mu) \times P_S(v|\pi')}$$

In (2), we can assume that P_{pers} is shared by all discourse participants, i.e. that they share the same μ function from which a probability measure on personae is derived. It is nevertheless easy enough to use different functions for different agents if deemed relevant.

2.3 Summary

Here we want to highlight an important assumption that underlies the view of social meaning sketched above. As alluded to, the variants for -ING in English are assumed to be equally accessible to all speakers (at least those with a typical competence). Using the sartorial simile that some sociolinguists are fond of, the situation is one in which agents have an infinite wardrobe: they can pick any outfit they want from the wardrobe, in accordance with the persona they want to "bricole".

Authors like Burnett (2023) note that in certain linguistic environments, some variants might be more costly because some linguistic factors (e.g. phonological assimilation) make one variant more difficult to produce, and thus more costly, but this applies to all speakers equally, i.e. every speaker pays the same

cost to produce a given variant that they have in their repertoire. Particularly competent and savvy speakers like Obama are distinguished by their verbal dexterity and aptitude at style shifting, but ultimately everyone has access to all the variants in the system.

In this research, we want to distinguish between this kind of cost, which we will refer to as a *production cost*, from the cost of having access to the variant in the first place, which we refer to as the *acquisition cost*. The acquisition cost represents the background effort required to be able to correctly deploy the variant, with the idea that not all variants are equally accessible to all speakers, i.e. that some speakers have a more limited wardrobe than others. Thus, as the name suggests, the acquisition cost reflects the resources required to acquire the variant as part of one's linguistic system. Burnett (2023) explicitly conflates the two costs, but we believe there is value in keeping them distinct, as we aim to show.

3 Costly Signals and Honest Failures

In the rest of this paper, we discuss linguistic variants that are not freely available to all speakers, at least not in their normative use. We frame the discussion around the idea of acquisition cost, which is the cost a speaker pays in order to acquire the normatively correct usage of a particular linguistic variant. In the case of *-ing* and *-in*, we assume that the acquisition cost of both variants is zero, since both variants are in the phonetic inventory of every competent English speaker. This means that any given speaker is taken to be able to deploy either variant in accordance with background grammatical norms.[1] But this is not always the case: some otherwise analogous cases of alternation involve a non-trivial acquisition cost for the correct deployment of one of the alternatives.

As an example, consider the use of *who* versus *whom* in English. In the spirit of Third Wave theorizing, we might posit that the use of *whom* indexes a certain set of social properties; something like *educated, pedantic, aloof*, etc., not unlike the properties indexed by *-ing*. But the deployment of *whom* versus *who* is normatively restricted to a subset of grammatical environments, namely, accusative case environments. In order to deploy the 'high register' variant *whom* in accordance with this set of associated norms, one must acquire not only the *form* but also this associated grammatical *system*. A speaker who has *not* acquired the associated grammatical system can still deploy the form *whom*, but they risk doing so in a way that violates the associated grammatical norms, e.g. producing sentences like "*Whom is coming to the party?*", where *whom* occurs in the normatively deviant nominative case position. The following examples (from the social media site Instagram) illustrate actual cases of such non-normative uses:

(3) a. I want to thank everyone **whom** came out to celebrate me & Da Legga birthday last night @bellagio223_chicago it was a stone cold groove.

[1] Though see Menner (1937) for a discussion of hypercorrect use of *-ing*.

https://www.instagram.com/p/DK0iOX3x5W0/
(consulted Dec. 1st, 2025)

b. To all **whom** came out to our shows this year THANK YOU were gonna be so back next year
https://www.instagram.com/p/DIq92Wdu9Ai/
(consulted Dec. 1st, 2025)

A speaker who deploys the high-register variant *whom* in this non-normative fashion will be understood as (attempting to) signal a certain persona, namely one associated with the properties indexed by *whom* (for example, *educated*). But, to a listener who has acquired the grammatical system associated with deployment of *whom*, the speaker will have 'outed' themselves as someone who has *not* acquired the system, since a competent user of the system would not have used *whom* in this environment. The speaker may be understood as attempting to index certain properties associated with *whom*, but they will also signal something about themselves that they do not intend; namely, that they are not competent at the grammatical system associated with deployment of *whom*.

To understand the situation here, we need to articulate a picture that involves more than the free deployment of alternative forms, characteristic of Third Wave theorizing. Instead, we have alternative *codes*. In the case at hand, the difference between the two codes is straightforward: Code 1 (the 'standard register' code) is one where the variant *who* is deployed across all grammatical environments. Code 2 (the 'high register' code) is one with two alternant forms, *who* and *whom*, whose distribution is determined by grammatical factors (nominative versus accusative case position). These codes are associated with different social groups, and have different acquisition costs. Code 1, we can say, is available across the board, to every speaker of English; it is what would most likely be described in a descriptive grammar of the language (see e.g. Huddleston and Pullum 2002, Chap. 5:16.2.3). Code 2, by contrast, is grammatically more complex (involving a distinction between two forms and an associated set of grammatical environments), and is moreover a code that is normally only acquired in certain circumstances, namely, in highly educated/academic/pedantic speech communities. In our terms, Code 2 has a high acquisition cost, since it is both grammatically complex and the environments in which it can be acquired are socially exclusive. The use of Code 2 is thus associated with membership in those socially exclusive communities where its acquisition is the norm, and signals membership in those communities, and, by extension, social properties associated with those communities. As the linguistic default, Code 1 may not have much indexical force, but insofar as it contrasts with Code 2, it might end up (at least for some listeners) with its own set of associations.

We now have two codes, at least one of which (Code 2) has the potential for indexing social meaning. For a listener who has acquired Code 2, there are two kinds of grammatical environment: those that license *whom*, and those that do not. There are thus two kinds of 'mismatch' that can exist, from the perspective of Code 2: environments E_1 where *whom* is expected but in which *who* is used, and environments E_2 where *who* is expected but in which *whom* is used. The

use of *who* in E_1 is a 'mistake', being non-normative according to the norms of Code 2, but is the norm for Code 1. Insofar as Code 1 is itself associated with non-academic/non-pedantic social groups, the use of *who* in such environments can serve to signal second-order properties associated with such groups, such as friendliness or casualness.[2] Basically, this is a kind of code-switching, with the norms (and social indices) of Code 1 in play. By contrast, the use of *whom* in E_2 is unlicensed by either code; its deployment can only be understood by someone who has acquired Code 2 as a mistaken attempt to deploy the resources of Code 2. The misuse of *whom* thus 'outs' the speaker as someone who is trying to use Code 2 without having paid the associated acquisition cost. They are thus not a member of the social groups associated with Code 2, but are acting as if they are. The listener concludes that the speaker is a *poser*.

Obviously, a speaker will not normally intend or desire to trigger the inference that they are a poser. So why would a Code 1 speaker ever use a form associated with Code 2, i.e. *whom*? We speculate that among those who have not acquired Code 2, the form *whom* might itself come to index the social properties that Code 2 does. That is, for non-Code 2 speakers, the deployment of *whom* might itself come to signal certain social properties (e.g. *educated, formal*, etc.) Such Code 1 speakers might then use the form associated with Code 2 to index those properties and a corresponding persona. This is presumably what is happening in the examples in (3). For a listener who, like the speaker, is not familiar with the norms of Code 2, the 'normal' kind of social meaning intended by the speaker in their deployment of *whom* can go through, full stop. But for a Code 2 listener, the notion of honest signaling, and more specifically of *honest failure*, comes into play. Such a listener will perceive the deployment of *whom* as an error, and this error will trigger associated inferences: that the speaker is a phony, that they are uneducated, and so on. For a Code 2 listener, the usage in examples like (3) is thus potentially associated with two layers of social meaning. The first layer, which the speaker in such examples presumably intends, would be the indexing of e.g. formality (in these examples, the context of thanking fans is naturally understood as one deserving of elevated or formal speech). The speaker might also be intending to index *educated*, but this signal fails – since *whom* is misused here, the speaker honestly signals (despite and regardless of their intentions) their lack of fluency in using Code 2, and thus their lack of having paid the associated acquisition cost, leading to inferences like *poser*, *uneducated*, etc.

4 SMG and Acquisition Costs

The formalization of SMG introduced in Sect. 2.2 does not offer any direct way to integrate the observations covered in the previous section. The naive listener

[2] As already mentioned, since Code 1 is the linguistic default in contemporary usage, there may not be any strong indexical associations with its use, but by *not* using Code 2, the speaker will *not* be indexing the properties associated with that code (i.e. educated, formal, etc.), and thus can be understood as avoiding signaling such properties. Thank you to an anonymous reviewer for raising this point.

model of (2) is potentially a good model for Code 1 users who consider *who* and *whom* freely interchangeable variables differing only in their social meaning (in a way that parallels variants of the -ING suffix). But the model does not work for Code 2 users who consider the grammatical environment in which a variant is used. Specifically, it is of no use in deriving the 'poser' inference that was discussed in the previous section. The equations in Sect. 2.2 also do not make any reference to acquisition cost, which as discussed we wish to keep distinct from the production cost that Burnett (2023) uses in some versions of her model.

We are interested here in modeling the inferences made by a listener who is themselves an in-group user of Code 2. As discussed, the form *whom* itself is marked as belonging to Code 2, whose normative use involves a non-trivial acquisition cost, and which is thus associated with an 'in-group' of normative Code 2 users (those who have paid the acquisition cost, and who thus typically are members of social groups where Code 2 is typically acquired). The listener uses two pieces of information in interpreting the social meaning of a deployment of *whom*. First, they have background beliefs about the speaker's group membership (i.e. about whether they are an in-group member who has paid the acquisition cost). They then consider the grammatical environment in which *whom* was deployed, and determine whether its deployment in this environment is correct or not, according to what they take to be the in-group norms of Code 2. In environments where *whom* is normatively deployed, nothing goes wrong: if the listener takes the speaker to be an out-group member, then they interpret them as someone who is trying, maybe successfully, to deploy the code associated with the in-group. The problem arises when *whom* is deployed in a non-normative way (i.e. when it is used ungrammatically, according to the norms of Code 2). Since the grammar of a fluent Code 2 user does not license *whom* in such environments, and since *whom* is not a part of Code 1, it is 'impossible' that a fluent Code 2 speaker would deploy *whom* here.

One way to adapt SMG to model the kind of reasoning sketched above is given in the listener's model shown in (4). In this model, γ is a variable that reflects the speaker's being part of a group that has acquired a certain code, and E is the linguistic environment in which the variant v appears. In the -ING example we have been pursuing, we would consider two groups: a group γ_1 of speakers who only acquired Code 1, and a group γ_2 of people who acquired Code 2 in addition to Code 1.[3]

$$(4) \quad P(\gamma|v,E) = \frac{P(\gamma) \times P_L(v|\gamma,E)}{\sum_{\gamma'} P(\gamma') \times P_L(v|\gamma',E)}$$

In (4), P_L represents the beliefs of the listener L about linguistic productions, be it about their own productions or that of others. When L identifies as a member of the group γ under consideration, one way of approximating these beliefs is to use L's own internal production grammar (i.e. equivalent to asking themselves "*Would I produce such a form?*" or producing a naturalness judgment

[3] Conceivably, we could imagine a group of (socially secluded) people who only acquired Code 2, but we leave them out of the picture for simplicity's sake.

about the form). When L does not identify as a member of the relevant code users, e.g. as when a Code 1 user interprets a use of *whom*, we have $P_L(v|\gamma, E) = P_L(v|\gamma)$, i.e. the beliefs about using variant v are independent of the environment in which the variant is used (which corresponds to the naive listener model discussed in Sect. 2.2).

The 'poser' inference is then triggered whenever the posterior belief in γ is not much higher, or even lower than the prior belief in it. Borrowing from the lexicon of argumentation, we can see the use of v in environment E as a piece of evidence for the conclusion that the speaker is part of the relevant group of code users: deploying *whom* in an environment in which it is licensed by Code 2 is evidence that one knows about Code 2. Thus, in the ideal case v in E acts as a strong piece of evidence for γ, i.e. it brings $P(\gamma|v, E)$ closer to certainty, or at least to a certain confidence threshold, and is typically such that $\frac{P(\gamma|v,E)}{P(\gamma)} > 1$. When that previous constraint does not obtain, the observation of v in E acts as a counterargument: it raises the probability that the speaker has not mastered the code, which we assume is what triggers the 'poser' inference.

Finally, we propose to integrate acquisition cost as part of the computation of the prior belief $P(\gamma)$ that the speaker is part of the group of relevant code users. The intuition at work here is that the higher the acquisition cost, the less likely it is that any given speaker has paid that cost, i.e. acquired the code in question. We can thus propose a simple calculation as in (5) where $AC(\gamma)$ refers to the acquisition cost of γ.

$$(5) \quad P(\gamma) = \frac{1}{1 + AC(\gamma)}$$

Note that the equation in (5) applies to the calculation of prior beliefs about a speaker who is unknown to the listener, i.e. is meant as a baseline that supposes that no other information is known about S. In some (many) cases, the identity of the speaker is already known to the listener. In particular, there are cases where it is known that the speaker belongs to the community associated with the code γ. In such instances, we can assume that $P(\gamma) = 1$ (or close to it), and that the listener is not ready to modify their beliefs about the group membership of the speaker. In those cases, we propose that the listener instead modifies their beliefs about the likelihood of producing v in E by members who know γ, meaning that listeners will revise $P_L(v|\gamma, E)$, which amounts to dropping rigidity conditions in the model (Jeffrey, 1983). We discuss an example of such a situation in Sect. 5.4.

As a final note in this section, we want to underline that the above considerations are not dealing directly with personae anymore, unlike what Burnett proposes. The inference that the listener draws in (4) is one about being part of a group of users of a certain code. Certain codes are associated with particular communities, and these communities are associated with certain properties that in turn can be used to form personae. So we see using a certain code as an indirect signal of being part of a certain community, which in turn can (indirectly) serve as an index of certain persona-forming properties.

5 Beyond *who/whom*

The *who/whom* alternation examined in this paper illustrates in a maximally simple way a situation in which two alternate codes coexist in a speech community, and in which one can, according to the norms of the prestige code (whose use comes with associated social meaning), make one of two distinct kinds of mistake: a 'hypocorrection' mistake, in which the resources of the low-prestige code have been used instead of the prestige code, and a 'hypercorrection' mistake, in which the resources of the prestige code have been misused or overapplied.

We think the kind of situation illustrated by *who/whom* is widespread. As a general rule, in cases like these the resources of the prestige code will not be accessible to every speaker in the community. If they were, then the prestige code would no longer function as a means of enforcing honest signaling; it is only in virtue of having a high acquisition cost and a limited set of environments in which acquisition takes place that the code serves as a kind of social gatekeeping device. In this section, we survey some other kinds of linguistic alternation where we think the logic of honest signaling and honest failure is relevant.

5.1 Speech Levels in Javanese

Javanese is, like other languages of Western Indonesia, characterized by a system of so-called 'speech levels'. The discussion below is based on the use of speech levels in Central Javanese, in particular as reflected in the use of aristocratic speakers in the court centers of Yogyakarta and Surakarta. Oversimplifying things a bit, the speech level system works via sets of synonymous suppletive alternative forms across the lexicon, with alternate forms signaling *Ngoko* ('low') or *Krama* ('high') speech level. The system is illustrated by the following example from Clynes (1989), where suppletive alternatives mark the resulting sentence as Krama or Ngoko:

(6) Bu Siti sampun nedha ingkang menika. **(Krama speech level)**
Bu Siti wis mangan sing kuwi. **(Ngoko speech level)**
Ms. Siti already eat REL that
"Bu Siti has already eaten that one."

In this example, there are two distinct alternates for the word meaning 'eat': *nedha*, signaling Krama speech level, and *mangan*, signaling Ngoko speech level. Similar alternations are found throughout the sentence. A mixture of Krama and Ngoko forms is also possible, giving rise to the so-called *Madya* ('mid') speech level. Gradations are also possible within these three levels; the fine-grained details of the system are left aside to simplify the discussion.[4] For the purposes of this paper, we can understand the system as consisting of two distinct codes: the Ngoko code and the Krama code, with Madya resulting from an admixture of the codes within a single utterance.

[4] See Davis 2021 for an analysis of the combinatorics of and gradations within the speech level system, building on Clynes 1989 and Wolff and Poedjosoedarmo 1982.

The choice between Ngoko, Madya, and Krama speech levels is made on the basis of social factors holding between the speaker and addressee: relative age, intimacy, and social status, with Krama being canonically used with older, non-intimate addressees of higher social status (Wolff and Poedjosoedarmo, 1982). Social status in Java is traditionally broadly divided into *priyayi* (aristocratic) and non-*priyayi* classes. Facility with the two codes is itself not evenly distributed across social classes. The Ngoko code is widely acquired and used across different social strata, but "proper" use of Krama is strongly associated with an aristocratic (*priyayi*) background. That is to say, non-*priyayi* do not generally have full command of the variety of Krama that is normative amongst the *priyayi* (Wolff and Poedjosoedarmo, 1982; Errington, 1985). However, non-*priyayi* speaking to *priyayi* are, traditionally, socially compelled to use the Krama speech level, as a marker of the distinction in their social status. The result is that, in Javanese, non-*priyayi* are socially required to use a code in which they are not generally proficient. As Wolff and Poedjosoedarmo (1982) put it, non-*priyayi* in such situations "use the closest to [Krama] that they can muster". A non-*priyayi* in such a situation ends up signaling their non-*priyayi* status in two ways; conventionally, via their attempt to use Krama with their *priyayi* interlocutor, and secondarily via their failure to use that code in conformance with the norms of *priyayi* users of the code. This constitutes an honest failure, signaling the speaker's non-*priyayi* status. In this case, since the social norms surrounding speech levels demand use of Krama in this circumstance, there is no 'poser inference', but simply an inference about non-*priyayi* group membership. The situation is one in which the non-*priyayi* speaker is compelled to out themselves as such via their linguistic behavior.

The situation here may seem bizarre, but consider, for example, a working-class young person at their first interview for a white-collar job. The interview context requires of the interviewee certain kinds of sartorial and verbal comportment, the standards of which the interviewee's background might not have provided them with the necessary verbal skill set (or wardrobe). More generally, we think a parallel conundrum applies to all situations in which a prestige code is expected, but in which facility in this code is unevenly distributed across the population. Failure to use the code normatively 'outs' the speaker as non-proficient, while successful use signals proficiency, and is thus evidence of having paid the associated cost. Both success and failure result in a variety of honest signal.

Returning to Javanese, the pitfalls of Krama have led to Javanese speakers abandoning it in favor of Indonesian (Vander Klok, 2019); in short, many contemporary speakers of Javanese use Ngoko Javanese exclusively, switching to Indonesian when social pressures demand a non-Ngoko form of the language. One can understand this as a strategy aimed at several ends, one of which is the avoidance of honestly signaling a lack of proficiency in the normatively correct (according to *priyayi* standards) Krama code.

5.2 Habitual *be* in AAE

The discussion so far has been about prestige language codes, with associated prescriptive norms; what Bourdieu (1991) calls the 'legitimate language'. But the phenomenon of honest signaling in social meaning does not only occur here. Consider the case of so-called habitual or invariant *be*, a grammatical feature characteristic of African American English (AAE; Zanuttini and Martin 2017). Its use is illustrated by the following examples from Jackson and Green (2005):

(7) a. Dee be in the house.
'Dee is usually in the house.'
b. Dee be running.
'Dee is usually running.' or 'Dee usually runs.'

As indicated in the paraphrases, habitual *be* is used to indicate a habitual or recurring action or state of affairs. Formally, the construction is marked by use of the uninflected verb *be*, whence the label 'invariant *be*'. Like other features of AAE, invariant *be* is familiar to speakers of other varieties of English, in particular to speakers of Standard American English (SAE). However, it seems that those who are not themselves fluent in AAE tend to misunderstand the semantics of sentences containing habitual/invariant *be*, treating *be* as a variant of the standard copula (8a) or progressive auxiliary (8b), so that the sentences in (7) are misunderstood as superficial variants of the SAE sentences in (8):

(8) a. Dee is in the house.
b. Dee is running.

Such misinterpretation by SAE-speaking children is demonstrated experimentally by Jackson and Green (2005), and presumably continues into adulthood for non-AAE speakers.

While thus misunderstood by the broader speech community, the linguistic resources of AAE are often borrowed/appropriated by non-AAE speakers. These AAE resources can then alternate with non-AAE variants in the construction of style and social meaning (see Bucholtz 1999 for one case study). In the case of habitual/invariant *be*, we can thus find examples of non-AAE speakers misusing the construction in an attempt to convey a particular style or social meaning associated with the deployment of AAE. This is illustrated by the following example (courtesy of John McCarthy), in which the white rapper Vanilla Ice is (mis)using the habitual/invariant *be* construction.

(9) Picture you and me on the beach and you **be climbing** up this tall coconut tree. (Vanilla Ice, *The Horny Song*)

The deployment here is syntactically correct, but is semantically/pragmatically odd in this context; the lyricist is asking us to picture a single instant, an instant in which the listener is climbing up a coconut tree. The context here is one that is compatible with a progressive interpretation, not a

habitual one. Thus, from the perspective of an AAE speaker, this deployment is ill-formed, and outs the singer as a poser, i.e. as someone attempting to deploy the linguistic resources of AAE without having paid the associated acquisition costs. The situation here is parallel to that of misuse of *whom*, with the difference that, in this case, it is a semantic error, rather than a simple grammatical error, that triggers the poser inference.[5]

5.3 Honest Signals, Accents, and Intentionality

The listener inferences associated with non-normative use of a code (e.g. that the speaker has not paid the associated acquisition costs for deployment of the code) are not normally what the speaker intends by their deployment of that code. Returning to the example around which we framed our initial discussion, when a Code 1 speaker deploys the Code 2 form *whom* in their speech, we take it they are intending to index certain properties associated with (the speakers of) Code 2, things like *educated* or *formal*. But the misuse of *whom* will trigger, for a Code 2 listener, a set of potentially contradictory and undesirable inferences, such as *uneducated* and *posering*. We take it that few if any speakers would intend their utterance to trigger such inferences.

The inferences generated here are a species of what Grice (1957) calls 'natural' meaning, which he illustrates with the example "*those spots mean measles*". The misuse of *whom* 'means' that a speaker is uneducated/a poser/etc. in the same way that the presence of spots 'means' that someone has the measles. Following Grice, we wouldn't say that "the speaker means by their misuse of *'whom'* that they are uneducated/a poser/etc." in these cases, because the speaker had no *intention* that such an inference be made from their use of *whom*.

The role of (non)intentionality and its connection to the natural/non-natural meaning divide is explored in detail by Acton (2022). Following Campbell-Kibler (2009), according to whom social meaning is 'social content tied in the minds of a given speaker/hearer to a particular piece of linguistic behavior', Acton (2022, p.473) claims that 'a linguistic behavior may therefore be socially meaningful in this sense simply by suggesting something about the social world, irrespective of speakers' intentions.' The kind of inferences we have focused on in this paper are thus a species of non-intentional social meaning. Acton mentions as another example of non-intentional social meaning the case of a native speaker of American English making an utterance in their native speech code. Such a speaker can, by virtue of the form (grammatical and phonetic) of their utterance, trigger an inference in an informed listener about their geographical origins; i.e. that they are from the United States. Such a speaker does not of course *intend* to trigger such an inference. From our perspective, the fluent use of the American vernacular is itself an honest signal of the speaker having paid the acquisition cost for the normatively correct deployment of that vernacular. Given that this acquisition

[5] Note that syntactic errors in the deployment of habitual/invariant *be* are also highly plausible, given the syntactic complexity of the construction; see Jackson and Green (2005) for a description.

cost is normally only paid by someone raised in an American English speaking environment, the (correct) use of the vernacular can then trigger the inference that the speaker grew up in such an environment, and that they thus likely grew up in the United States. As an additional layer, a listener might then use their beliefs or stereotypes regarding Americans (e.g. *friendly*, *informal*, *arrogant*) to derive additional inferences of a social nature.

Honest signaling of social meaning is a feature of accents in general, and perhaps of foreign accents more particularly. Take a Latin American immigrant to the United States whose first language is Spanish. Assuming they learned English after a certain age (the 'critical period'), they are likely to have a 'Spanish accent' when speaking English. That is, they are likely to employ phonetic features characteristic of (their particular dialect of) Spanish in their articulation of English. This serves as an honest signal of their linguistic background (native speaker of Spanish, non-native speaker of English) and potentially their geographic background (Latin America, Mexico, etc.). As before, the speaker here is in no sense 'trying' to signal anything about themselves; the signal is an honest one, and is not connected with intentions.[6]

However, the phonetic/phonological and grammatical properties of such accents can themselves serve as resources in the development of distinct dialects of English, with their own logics, patterns of variation, and ways of signaling social meaning. In the case at hand, many Chicanos or Mexican-Americans employ a dialect of English known as Chicano English. A characteristic feature of this dialect is the so-called *sh*∼*ch* alternation (Flores-Bayer, 2017), whereby words like 'chair' are pronounced like 'share', or words like 'much' are pronounced like 'mush' (i.e. phonemic /sh/ is pronounced [ch] and phonemic /ch/ is pronounced [sh]). As described by Flores-Bayer (2017), this feature, which may originate in part from a non-native Spanish/Mexican accent, is now a regular part of the phonological grammar of native Chicano English speakers. Flores-Bayer (2017) shows experimentally that the use of this alternation is perceived by fellow community members[7] as signaling certain social properties (*relaxed*, *sincere*, *trustworthy*, etc.), but crucially *not* signaling 'foreign-born' or 'non-native English speaker'. This is a case where, plausibly, a feature that was originally

[6] Some readers will no doubt be familiar with the following passage from Judges in the Hebrew Bible (KJV, Judges 12:5–6):

> And the Gileadites took the passages of Jordan before the Ephraimites: and it was so, that when those Ephraimites which were escaped said, Let me go over; that the men of Gilead said unto him, Art thou an Ephraimite? If he said, Nay;
> Then said they unto him, Say now Shibboleth: and he said Sibboleth: for he could not frame to pronounce it right. Then they took him, and slew him at the passages of Jordan: and there fell at that time of the Ephraimites forty and two thousand.

The Gileadites in this case surely had no intention to signal their tribal affiliation via their accent, consequences of honest signaling in this case being quite dire.

[7] Particularly, members of the East Austin 'El Barrio' community.

an honest signal of a non-native English speaker of a particular background has now been grammaticalized or pragmaticalized into a feature that, within the relevant community, has acquired a new set of social meanings. This in turn allows speakers of Chicano English to deploy this variable strategically (i.e. with intention) in the construction of a particular social persona.

5.4 Legitimate Hypercorrection on Martha's Vineyard

Up to this point, we have seen cases where misuse of a variant serves as an honest failure; the use of *whom* where it doesn't belong marks the speaker as a phony, someone who hasn't properly acquired the in-group code. We speculate that this is generally true for hypercorrection, *when produced by out-group members*. But what about hypercorrections produced by members of the in-group? Do such things happen, and if so what inferences are drawn?

Labov's famous study of vowel centralization on Martha's Vineyard (Labov, 1963) provides an illuminating discussion in this regard. Oversimplifying things, Labov found that among locals on Martha's Vineyard the degree of vowel centralization in the dipthongs /ai/ and /au/ was correlated with speaker identity and attitudes, with a more positive orientation toward Martha's Vineyard associated with a greater degree of centralization. According to Labov, the 'meaning of centralization' on Martha's Vineyard is '*positive orientation towards Martha's Vineyard*'. Crucial to Labov's explanation is the idea that an underlying regional accent feature (vowel centralization) underwent a form of hypercorrection *by in-group speakers* in its metamorphosis to a linguistic variable with a strategically deployable social meaning. That is, speakers of Martha's Vineyard English who have a strong positive orientation toward the island have exaggerated the degree of vowel centralization in their own productions as a means of indexing that orientation. Far from 'outing' the speakers as posers, hypercorrection in this case seems to have been a crucial ingredient in the development of a socially meaningful linguistic variable.

The logic here differs from that seen earlier in that speakers in this case are uncontroversially in-group members of the Martha's Vineyard community. Following Bourdieu (1991), we can say that the norms of the 'legitimate code' (in this case, the code of authentic Martha's Vineyarders) are *determined by* the linguistic productions of the legitimate speakers of that code. In this sense, it is in a way *impossible* for such a speaker to make an error. Or to put it another way, the 'error' of hypercorrection becomes, due to the legitimacy assigned to the hypercorrector, the new norm.

6 Conclusion

We introduced the notions of honest signaling and honest failure in relation to social meaning, arguing for the importance of considering acquisition cost of variants and uneven access to their normative deployment across a community of speakers. We connected our discussion to Third Wave theorizing about social

meaning, and in particular to the SMG model of Burnett (2023), suggesting one way that this model might be expanded to encompass the kind of honest social meaning described. One upshot of this discussion is that alternative linguistic codes with non-trivial acquisition costs can serve as honest signals of group membership. Correct (according to in-group norms) deployment of the code is a signal of group membership, while incorrect use (honest failure) is a signal of non-membership in the group. In the case of honest failures, speakers 'out' themselves, unintentionally, as non-members of the in-group. The inferences here are primarily about group (non-)membership, and only secondarily about associated social properties and personae. The honest signals in question are a type of social meaning, broadly construed, but one which goes beyond the kind of (intentional) persona construction characteristic of Third Wave theorizing.

Acknowledgments. This study was partially funded by NSERC (grant number RGPIN-2024-06718).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

Acton, E.K.: Sociophonetics, semantics, and intention. J. Linguist. **58**(3), 465–494 (2022)
Bourdieu, P.: Language and Symbolic Power. Harvard University Press, Cambridge (1991)
Bucholtz, M.: You da man: narrating the racial other in the production of white masculinity. J. Socioling. **3**(4), 443–460 (1999)
Burnett, H.: Signalling games, sociolinguistic variation and the construction of style. Linguist. Philos. **42**, 419–450 (2019)
Burnett, H.: Meaning, Identity, and Interaction. Cambridge University Press, Cambridge (2023)
Campbell-Kibler, K.: Listener perceptions of sociolinguistic variables: the case of (ING). Ph.D. thesis, Stanford University (2006)
Campbell-Kibler, K.: Accent, (ING), and the social logic of listener perceptions. Am. Speech **82**(1), 32–64 (2007)
Campbell-Kibler, K.: I'll be the judge of that: diversity in social perceptions of (ING). Lang. Soc. **37**(5), 637–659 (2008)
Campbell-Kibler, K.: The nature of sociolinguistic perception. Lang. Var. Chang. **21**(1), 135–156 (2009)
Clynes, A.: speech styles in javanese and balinese: a comparative study. Master's thesis, Australia National University (1989)
Davis, C.: Deriving categorical and continuous properties of Javanese speech levels. In: Clark, T., Dussere, J., Ting, C. (eds.) Proceedings of the Twenty-Eighth Meeting of the Austronesian Formal Linguistics Association (AFLA), pp. 11–28 (2021)
Eckert, P.: Variation and the indexical field. J. Socioling. **12**(4), 453–476 (2008)
Eckert, P.: Meaning and Linguistic Variation: The Third Wave in Sociolinguistics, 1 edn. Cambridge University Press (2018)

Errington, J.J.: On the nature of the sociolinguistic sign: describing the Javanese speech levels. In: Semiotic Mediation: Sociocultural and Psychological Perspectives, pp. 287–310. Elsevier (1985)

FitzGibbon, F., Fanshawe, J.: Stotting in thomson's gazelles: an honest signal of condition. Behav. Ecol. Sociobiol. **23**, 69–74 (1988)

Flores-Bayer, I.: Sociolinguistic variation in practice: an ethnographic study of stylistic variation and social meaning in the Chicano English of 'El Barrio'. Ph.D. thesis, Stanford University (2017)

Grice, H.P.: Meaning. Philos. Rev. **66**(3), 377–388 (1957)

Hall-Lew, L., Moore, E., Podesva, R.J. (eds.): Social Meaning and Linguistic Variation: Theorizing the Third Wave, 1 edn. Cambridge University Press (2021)

Huddleston, R., Pullum, G.K.: The Cambridge Grammar of the English Language. Cambridge University Press, Cambridge (2002)

Jackson, J.E., Green, L.: Tense and aspectual *be* in Child African American English. In: Frazier, L., et al. (eds.) Perspectives on Aspect, vol. 32, pp. 233–250. Springer, Dordrecht (2005)

Jeffrey, R.: The Logic of Decision, 2nd edn. University of Chicago Press, Chicago (1983)

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

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

Lailvaux, S.P., Reaney, L.T., Backwell, P.R.Y.: Dishonest signalling of fighting ability and multiple performance traits in the fiddler crab uca mjoebergi. Funct. Ecol. **23**(2), 359–366 (2009)

Menner, R.J.: Hypercorrect forms in American English. Am. Speech **12**(3), 167 (1937)

Podesva, R.J.: The California vowel shift and gay identity. Am. Speech **86**(1), 32–51 (2011)

Vander Klok, J.: The Javanese language at risk? Perspectives from an East Java village. Lang. Doc. **13**, 300–345 (2019)

Wolff, J.U., Poedjosoedarmo, S.: Communicative Codes in Central Java. Southeast Asia Program Publications, Ithaca, New York (1982)

Zanuttini, R., Martin, K.: Invariant be. Yale Grammatical Diversity Project: English in North America (2017)