# The Evolution of Egalitarian Socio-Linguistic Conventions

*By* Suresh Naidu, Sung-Ha Hwang, and Samuel Bowles*

Half a century ago two works laid the foundation for modern sociolinguistics. In his *Conventions*, David Lewis (1960) studied a speech community using a system of symbols and grammatical rules and coordinating on the mapping from symbols to states of the world in order to communicate effectively. Economists and others built on Lewis' work to erect the modern theory of conventions, their persistence and occasional transformations Young (1998). Subsequent research has used evolutionary dynamics to show that successful languages are informationally efficient under the constraints imposed by human cognitive and sensory systems.

At about the same time Brown and Gilman published their "Pronouns of Power and Solidarity" exploring the fact that "a man's consistent pronoun style gives away his class position" and that this "non-reciprocal power semantic" had been the norm in many Indo European languages for at least half a millennium despite recurrent contestation by egalitarian language innovators. (Brown and Gilman (1960))

Subsequent research has documented a wide variety of pronominal markers of status, in particular what Brown and Gilman called the T-V distinctions (e.g. "Tu" vs "Vous" in French), the semantics of which typically involve an ambiguity in that the V pronoun may denote superordinate status as well as plurality (Tabellini (2008)). The T-V status markers are far from unique as linguistic features of group interests and identity. Labov (2011) and other sociolinguists have established, for example, that class-based accents are pervasive.

Here we use recently developed models of conventions to model the evolution of the non reciprocal power semantic studied by Brown and Gilman. Our objective is to provide a framework that is consistent both with the long term persistence of the T-V distinction (despite the ambiguity intrinsic to the dual use of the V pronoun for status and plurality) and with its recent displacement in many languages by an egalitarian pronoun convention. To do this we consider a battle of the sexes coordination game in a population composed of two classes in which communication of relative status imposes subjective costs on the subordinate class and also may serve as a socially valuable coordination device.

We specify conditions under which evolutionarily successful languages are unambiguous and egalitarian in the sense of not communicating inferior status by the choice of pronouns, for example. But if, as is often the case, the subordinate population is large relative to the elite and linguistic innovations are intentional (rather than mutation-like accidents) then ambiguous and unequal conventions are likely to emerge and persist over a long period, consistent with the history of the T-V conventions.

## I. Bottom-up Intentional Transitions

### A. *Race and Gender Language Conventions*

The terms for referring to race and gender identities have been transformed since the 1960s, largely due to attempts to shape the language made by activists and intellectuals. The transition from "colored" to "negro" to capitalized "Negro/Afro-American" to "Black/African-American" was not the result of spontaneous linguistic innovation, but instead deliberate campaigning. Booker T. Washington and the NAACP advocated for linguistic change in the early 20th century, with a victory scored when the New York Times imposed capitalized Negro in its style guides on March 7, 1930. Black power activists in the 1960s debated the merits of various terms and *Ebony* magazine joined them in promoting "Black". This change was soon reflected in the overall culture, as can be seen from Figure 1.

Gendered language has also been transformed, where the universal "he" has been cur-

* Naidu, Columbia University; Hwang, Korea Advanced Institute of Science and Technology; Bowles, Santa Fe Institute.

tailed in recent years. As with the transformation of racial language, the change was deliberately promoted by many independent deviations from the status quo (**?**). The magazine $Ms$ titled to promote a new salutation of women was founded by two feminist activists Gloria Steinem and Dorothy Pitman Hughes. The year it was first published, 1972, the Modern English Handbook confirmed the traditional pronoun use: "He, alone, is usually preferred", while modern guides often suggest "he or she" and suggest that gender not be presumed. A result of this norm is that gender-inclusive subjects have increased (as in Figure 1). (Curzan, 2014)

### B. *Honorifics and Modes of Address During Political Transitions*

To examine more long-run changes, we look at the evolution of honorifics, and particularly the T-V distinctions of status when used as a singular pronoun. There has been a decline of the honorific (non-reciprocal) use of "V" among strangers, where superiors are addressed by "V" and subordinates by "T". Kachru and Smith (2008) "there is an increasing tendency to address all intimates, regardless of status, with the T-pronoun, and all strangers with the V-pronoun," a trend identified Brown and Gilman in their initial paper.

A good example is French, as shown in Figure 1, which plots the relative frequency of "Vous" to "Tu" (case-insensitive) in the Google N-grams database. While this is naturally noisy and imperfect data[1], it shows a relative increase in the "Tu" form around the French Revolution. While we cannot distinguish between formal and informal symmetric (vs asymmetric use), this is consistent with revolutionary norms of egalitarianism that prevailed during the Revolution.

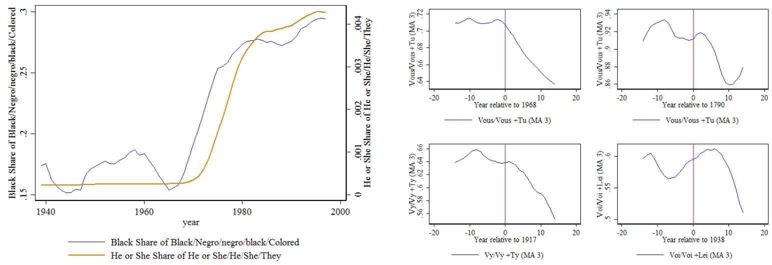
The Committee for the Public Safety denounced "vous" as a feudal anachronism; Robespierre *tu*'d the Assembly's President. Anderson (2007) writes "the idea of using *tu* in all circumstances was first proposed in an article in the $MercureNational$ on December 14, 1790....No laws were passed registering the mandatory use of *tu* but...[it]..began to spread. Now the baker's apprentice could address his master and clients in a familiar form, a practice that had been strictly forbidden." The data also a sharp increase following the student movements of 1968, with activists again deliberately using "Tu" to address superiors.

The Russian revolution provides another example as can be seen in the Russian corpus in Figure 1. The Russian military abolished the use of the honorific within the army in 1917. This diffused into a larger linguistic change (Corbett, 1976). A demand of revolutionary workers, for example during the Lena strike of 1912, was to be addressed in the polite mode Stites (1988). During and after the revolution Russian intellectuals and activists intentionally began using the informal mode of address universally.

Deliberate attempts to change conventions back to inegalitarian ones sometimes fail. Mussolini's fascist movement in Italy attempted to move Italians away from the pronoun *Lei*, which was honorific and to some effeminate. The Partito Nazionale Fascista restricted the use of "Lei" among members pushing "voi" instead and then mandated the same practice for public employees in 1939. As can be seen from the figure, this reform was unsuccessful, with the secular relative decline in the use of "voi" only temporarily reversed during the two decades of Mussolini's regime.

Like Mussolini's intervention, deliberate bottom up challenges to status quo conventions also typically fail. The Society of Friends (Quakers) founded in the mid 17th century by George Fox raised the banner of Plain Speech, according to which the informal "Thou" or "Thee" was prescribed for all social interactions rather than the asymmetric formal "you" or reverential "Ye". He wrote: "... when the the Lord sent me forth into the world, He forbade me to put off my hat to any, high or low: and I was required to Thee and Thou all men and women without any respect to rich or poor, great or small." But little came of it, and the informal pronouns eventually were all but abandoned, while "you" lost its status connotations.

[1]The frequencies are calculated from the universe of books digitized by Google, see Michel et al. (2011) and discussion in Appendix.




*Note:* Left: Transitions in group titles: Relative increase in fraction Black and 3-word phrases that are "he or she" following 1960s. Right: Clockwise from top left 15 year windows around political events: Vous/(Tu+Vous) relative frequency in the 2012 French Google N-grams corpus around 1968 and 1790. Russian Vy/(Vy + Ty) around 1917. Italian Voi/(Voi+Lei) around 1938.

FIGURE 1.

## II. An Evolutionary Linguistic Model

Our model is a contribution to an evolutionary socio-linguistics drawing on our previous work on intentional evolutionary equilibrium selection and on a rich literature in evolutionary linguistics which we cite more fully in an on line appendix (in which we also present a more complete technical representation of our model). We represent a language as a convention, that is a mutual best response of speakers who may adopt differing languages. By the evolutionary success of a language convention we mean roughly the likelihood over a very long period of time that a population will coordinate on that particular convention when speakers typically best respond by conforming to the status quo convention but occasionally (with probability $\epsilon$) innovate. [2]

We model language evolution as decentralized process in which the common-language coordination that may occur is an emergent property of uncoordinated interactions, not a mandated result. But when individuals deviate from the prevailing convention, they do so not in error but instead intentionally, adopting an alternative convention in which the would be better of, were the rest of the population to do the same.

We build on the model of language of Nowak, Plotkin and Krakauer (1999), but extend it in a number of ways. We incorporate two populations, $A$ and $B$, where both population sizes $N_A$ and $N_B$ are large and $N_A = \eta N_B$, where $\eta$ is the relative group size of $A$. We assume that the payoffs are asymmetric so as to capture a Battle of the Sexes logic: a common language convention is preferred by both populations, but they differ on which convention they prefer .

We divide the space of objects to be communicated into "Regular" (R) and "Status-relevant" (S). Members of the A population are randomly paired with B's and may with equal probability be a sender or a receiver. A language strategy is a probability matrix mapping objects to symbols (the sending matrix), and the transpose of that matrix is the "receiver" matrix that decodes symbols back into objects. For example, a sender who utters "letter" could with some probability intend "one of the items making up the alphabet" and with the complementary probability intend to convey "a written message."

Communication is successful if the object signalled by the sender is decoded correctly by the receiver. Since the receiver matrix is the transpose of the sending matrix, communication occurs with highest probability when both agents are coordinating on the same language. Unsuccessful communication gets a payoff of 0.

Agents get a payoff of 1 from successfully communicating the $R$ objects. When communicating about in the status relevant objects there is some total payoff (to be distributed between the A and B member of the pair) to successfully communicating status differences. For example passage through a doorway may be coordinated by the norm that the higher status person goes first, and the observance of this norm and its as-

[2] More technically an evolutionarily successful language is stochastically stable. A Nash equilibrium (for example a language convention) is *stochastically stable* if the resulting ergodic distribution of strategies in the population puts positive mass on that equilibrium as $\epsilon \to 0$.

sociated benefits (avoiding collisions or endless deferring to the other) may be communicated by some aspect of the language on which the two coordinate, such as the T-V distinction.

We denote this total benefit by $\rho$, and we imagine systems of economic and social interaction in which this might be a considerable magnitude. For example, if costly conflicts in dyads are sometimes avoided by the mutual recognition of status differences (e.g. the rule that subordinates must cede to the dominant member (as in many primates) then $\rho$ would be substantial. $\rho$ may also reflect payoffs to communicating status that are unrelated to status itself (e.g. using T-V may result in higher outside market rewards (Lazear (1999); Clingingsmith (2015)).

The member of group $A$ gets payoff $\theta\rho < 1$, while the $B$ member of the interaction gets payoffs $(1-\theta)\rho$ , where $\theta < \frac{1}{2}$. Thus group $A$ is the relatively "low status" group. In the example they derive some benefit by avoiding the doorway collision, but at the cost of publicly acknowledging their social inferiority.

We consider just two languages in which we let $P^e$ be an egalitarian language that does not let an agent communicate status and $P^u$ be an inegalitarian language that does let an agent communicate status with with probability $x$. So if the sender using $P^u$ utters "I will see you (using a V pronoun) later," it could mean the plural "you and your family" with probability $1-x$ or the singular "you, my recognized superior" with probability $x$.

Because in $P^u$ V may therefore designate either plurality or status, there is a probability of miscommunication even when the A to B match are individuals both adopting $P^u$. In this case both agents using $P^u$ will understand each other with probability $(1-x)^2$ when communicating the Regular object and $x^2$ when communicating the Status object. Our assumption of 2 strict Nash equilibria imply payoff restrictions that make $x$ to be a monotone measure of clarity (lack of ambiguity) in communicating status (i.e. $\frac{1}{2} < x < 1$).

By contrast, both agents using $P^e$ will never mis-communicate plurality, but they will also never communicate status. When an agent using $P^e$ encounters an agent playing $P^u$ they receive only the payoff from communicating plurality, and then only when the Agent playing $P^u$ communicates plurality. In the Appendix, we derive the payoffs from language coordination and mis-coordination more formally, and show that we can represent the payoffs as a simple coordination game in Table 1.

We are interested in the case in which there are two strict Nash equilibria, corresponding to the two conventions $P^e$ and $P^u$ which from the payoffs in the table is assured if $x > \frac{1}{1+\rho\theta}$, which we assume. We now turn to determining which equilibrium is selected under an explicit evolutionary dynamic.

We impose the dynamics in Hwang, Naidu and Bowles (2016)[3], where agents myopically play best responses to the distribution of strategies in the population, and have the opportunity to play idiosyncratic strategies with probability $\epsilon$. The occasion for these "innovations" is random but their direction is not. To capture the purposeful nature of language change we say that agents are less likely to play idiosyncratically when the status quo convention is one they prefer. The parameter $\iota$ measures the degree to which the errors are "intentional": supported only on the strategies that if widely adopted would result in an equilibrium that is better than the current one.

Besides varying the degree of intentionality with $\iota$, we allow for asymmetric population sizes: the number of Group A is $\eta$ times the number of Group B. As in Hwang et al. (2016), these modifications to the stochastic evolutionary equilibrium selection context allow the theory to be used to model equilibrium selection in environments where there are conflicts of interests between groups that differ in size, payoffs, and level of internal organization and mobilization.

The underlying dynamical process is ergodic so that the population never gets 'locked into' one language or the other, but will persist for a long time in one before making a transition to the other, only to return after another long period of stasis. Our measure of the "evolutionary success" of a language can be judged by the expected duration of a long period of time that the population will spend under that particular convention, compared to the same statistic for the other population. Evolutionary success in a language is equivalent to persistence in this

[3] As $x \to 1$ this game reduces to the game studied in that paper.

| | | Group B | |
|---|---|---|---|
| | | $P^e$ | $P^u$ |
| Group A | $P^e$ | $1, 1$ | $1-x, 1-x$ |
| | $P^u$ | $1-x, 1-x$ | $(1-x)^2+\rho\theta x^2, (1-x)^2+\rho(1-\theta)x^2$ |

Table 1—**Payoffs in the Language Game.**

*Note:* Both agents using $P^e$ will communicate the Regular object correctly, obtaining payoffs 1. When an agent using $P^e$ encounters an agent playing $P^u$ (or agent using $P^u$ encounters an agent using $P^u$), they only receive payoff 1 from communicating plurality, which occurs with probability $1-x$. Finally, both agents using $P^u$ will only understand each other with probability $(1-x)^2$ when communicating the Regular object and $x^2$ when communicating the Status object. Thus both agents obtain payoff 1 from communicating the Regular object with probability $(1-x)^2$, while group A agents obtain payoff $\rho\theta$ and group B agents obtain payoff $\rho(1-\theta)$ each with probability $x^2$.

dynamic and is inversely related to the probability in any period of making a transition to the other language.

Some of our results are unsurprising. The unequal language will be more persistent, the more valuable is the communication of status differences (the greater is $\rho$)and the lesser the ambiguity in accomplishing this (the greater is $x$), illustrating the basic point that reduced ambiguity increases evolutionary stability of a given convention, holding everything else constant.

Other results are more counter intuitive. The dynamic we have modeled, like that in Young (1998) favors languages that are egalitarian: a unequal language with a higher $\rho$ may be less persistent than a more equal variant of $P^u$ with high $\theta$

The reason that more equal convention is more likely to be evolutionarily successful is the following. First, as $\theta$ increases, the payoff to group $A$ agents at $P^u$ convention increases and $P^u$ is less unfavorable to group $A$ agents. Thus, this raises the threshold number of group $B$ agents inducing group $A$ agents to adopt $P^e$ language as best-responses and makes transition from $P^u$ convention more difficult. Second, an increase in $\theta$ also decreases the payoffs to agents $B$ at $P^u$ convention and the relative payoff advantages of $P^u$ convention (compared to $P^e$ convention) to group $B$ agents are reduced. This decreases the threshold number of group $A$ agents inducing agents $B$ to play $P^e$ as best-responses and expedites transition toward $P^e$ convention. The two effects of a more equal convention work in in opposite direction, the first stabilizing $P^u$ convention, and the second destabilizing it. But for $\theta < \frac{1}{2}$, the first effect is larger, so the more equal convention, the population will spend more time at $P^u$ convention.

But our results show that this egalitarian tendency in the underlying dynamic is a rather model specific property. Directed linguistic innovations ($\iota$ large) and small population sizes ($\eta$ either very small or very large) can stabilize even vague ($x = \frac{1}{2}$) status distinctions in language of little value.

The intuition behind this is that transitions are induced by extreme realizations of the stochastic process generating innovations, in which a large fraction of a given population does not best respond. The extreme realizations are more likely in small populations for the same reason that the variance around a sample mean is greater the smaller is the sample.

This makes it more likely that it is the members of a small population whose innovations will induce a transition. But if their innovations are random rather than intentional, they will as likely induce a transition away from their preferred convention as towards it.

If we add to the advantage of small size the fact that people innovate intentionally, so that they only induce transitions in a direction from which they benefit, then we have our main result. A relatively a small population, when deliberately innovating in opposition to an egalitarian status quo convention can easily destabilize conventions that are not in their favor. Evolutionarily successful linguistic conventions – including the non reciprocal power semantic studied here – need not excel in clarity of communication, as the long persistence of the asymmetrical use of T-V pronouns suggests[4]

[4]In Hwang, Naidu and Bowles (2016) we show how these results extend to games played on arbitrary networks. We provide conditions on the topology of arbitrary networks that can make $P^u$ the more persistent, whereby a small well-connected set of $A$'s can induce a large set of $B$'s to change

Our final finding is that a population that has a higher rate of idiosyncratic play will be favored and will spend more time speaking the convention they prefer when linguistic change is intentional. This unsurprising result appears consistent with the changing gender and racial language that we have reviewed, especially in light of the fact that in these cases media and government played important roles in amplifying the influence of a few innovators.

## REFERENCES


**Anderson, James Maxwell.** 2007. *Daily life during the French Revolution.* Greenwood Publishing Group.

**Brown, Roger, and Albert Gilman.** 1960. "The pronouns of power and solidarity. Style in language, ed. by Thomas A. Sebeok, 253-76."

**Clingingsmith, David.** 2015. "Are The World's Languages Consolidating? The Dynamics and Distribution of Language Populations." *The Economic Journal.*

**Corbett, GC.** 1976. "Address in Russian." *Journal of Russian Studies*, , (31): 3–15.

**Curzan, Anne.** 2014. *Fixing English: Prescriptivism and language history.* Cambridge University Press.

**Hwang, Sung-Ha, Suresh Naidu, and Samuel Bowles.** 2016. "Social conflict and the evolution of unequal conventions."

**Kachru, Yamuna, and Larry E Smith.** 2008. *Cultures, contexts, and world Englishes.* Routledge.

**Labov, William.** 2011. *Principles of linguistic change, cognitive and cultural factors.* Vol. 3, John Wiley & Sons.

**Lazear, Edward P.** 1999. "Culture and Language." *Journal of Political Economy*, 107(S6): S95–S126.

**Lewis, David.** 1960. *Convention: A philosophical study.* John Wiley & Sons.


their language with only a few idiosyncratic innovations.


**Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K Gray, Joseph P Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, et al.** 2011. "Quantitative analysis of culture using millions of digitized books." *science*, 331(6014): 176–182.

**Nowak, Martin A, Joshua B Plotkin, and David C Krakauer.** 1999. "The evolutionary language game." *Journal of Theoretical Biology*, 200(2): 147–162.

**Stites, Richard.** 1988. *Revolutionary dreams: Utopian vision and experimental life in the Russian revolution.* Oxford University Press.

**Tabellini, Guido.** 2008. "Institutions and Culture." *Journal of the European Economic Association*, 6(2-3): 255–294.

**Young, H. Peyton.** 1998. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions.* Princeton University Press.

## Appendix

"When I use a word," Humpty Dumpty said in rather a scornful tone, "it means just what I choose it to mean neither more nor less." "The question is," said Alice, "whether you can make words mean so many different things." "The question is," said Humpty Dumpty, "which is to be master– that's all." -Lewis Carroll. Through the Looking-Glass, and What Alice Found There

### A1. *Literature on which we draw*

Since the initial contribution of Lewis, evolutionary linguists and biologists studying models of language have developed a rich set of evolutionary models describing the distribution of languages, the emergence of words, syntax, and universal grammar, and the evolution of grammatical conventions such as verb regularization (Pagel 2009, 2013, Nowak and Krakauer 1999, Nowak et al. 2001, Michel et al. 2010, Ahern et al. 2016). These models have a population of individuals who play languages, modelled as mappings from objects to signals, and who must successfully coordinate on a language in order to communicate. Nowak and Trapa (2004) show that evolutionarily stable languages are strict Nash equilibria where the number of signals are equal to the number of symbols, and each symbol is emitted with probability 1 conditional on the state. Empirically, studies of language evolution have documented "ultraconserved" words from the last ice age that remain in related forms in the current language distribution (Pagel et al. 2013), and that much linguistic evolution happens in sharp, punctuated bursts (Atkinson et al. 2008).

Nowak and Plotkin (2000) use results from coding theory to show that evolutionarily fit languages will efficiently transmit information. Efficiency is measured as the number of bits needed to transmit a message, and so Shannon's coding theorem (Shannon 1948) gives the theoretical upper bound on the efficiency of a channel, given the noisiness of the source encoding and decoding. In natural language, the noise is due to biological and cognitive constraints on signal processing. Christiansen (2015) argues that even complex features of language, such as recursion, can be explained by adaptation to pre-given human constraints, and so the remarkable adaptiveness of humans to language is a result of natural selection of languages according to ease of acquisition and usefulness. In theory, language should attain the informationally efficient bound (Shannon 1948). Indeed, a recent study by Pellegrino, Coupe, and Marsico (2011) finds that the information transmitted per second across spoken languages is quite stable, despite differences in speed of speaking and information per syllable. However, one thing lacking in these models is a justification for the persistence of "vague" or "ambiguous" linguistic conventions with significantly higher entropy than the optimal.

A second strand of literature looks at the economics and political economy of language and language policy. Within economics, the seminal paper is Lazear (1999) who makes the simple point that languages that increase the space of trading opportunities will be adopted, and makes numerous predictions that follow from this. The closest model to ours in Clingingsmith (2014), who models languages as conventions on networks. Clingingsmith shows that language growth follows a GIbrat's law, and that the world's languages are doubly-Pareto distributed.

### A2. *Data on Linguistic Corpora*

We use the Google N-Grams corpus due to ease of access. However

See also Davies (2015) and criticisms by Pechenick et al.(2015).

In the Corpus of Historical American English (COHA)[5], "he or she" has increased in frequency as a share of 3-word phrases from less than 2 per million prior to 1970 to 15 per million in 1990.

[5] Davies, Mark. (2011) N-grams and word frequency data from the Corpus of Historical American English (COHA). Downloaded from http://www.ngrams.info on December 18, 2016.

*A3. The model described informally in the paper*

To describe a language, we let a message sending matrix $P$ be a $(|R|+|S|)\times|W|$ stochastic matrix mapping objects into words and a message receiving matrix $P'$ be a $|W| \times (|R| + |S|)$ stochastic matrix mapping words into object. Here, $P_{ij}$ means that word $j$ is emitted with probability $P_{ij}$ when object $i$ is being communicated and $P'_{ji}$ means that object $i$ is received when word $j$ is heard. Thus $P_{ij}P'_{ji}$ is the probability that two persons successfully communicate object $i$ with word $j$. Then a language $L$ is defined to be $L := (P, P')$. Thus the set of all languages, $\mathcal{L}$, is given by

$$\mathcal{L} := \Delta^{(|R|+|S|)\times|W|} \times \Delta^{|W|\times(|R|+|S|)}$$

To simplify, we consider $L$ such that $L = (P, P^T)$, where $P^T$ is the transpose of $P$. We then can regard a matrix $P$ as a language. Under this assumption, the communication probability of two players using $P$ and $Q$ languages for object $i$ using word $j$ is given by $P_{ij}Q^T_{ji} = P_{ij}Q_{ij}$.

The payoff of an agent playing language $P$ when communicating with an agent playing language $Q$ is:

(A1)

$$U^A(P,Q) = \underbrace{\frac{1}{2}\sum_{i\in R}\sum_{j\in W} P_{ij}Q_{ij}}_{\text{Regular Interaction Sender}} + \underbrace{\frac{1}{2}\sum_{i\in R}\sum_{j\in W} P_{ij}Q_{ij}}_{\text{Regular Interaction Receiver}} + \underbrace{\frac{\theta\rho}{2}\sum_{i\in S}\sum_{j\in W} P_{ij}Q_{ij}}_{\text{Status Interaction Sender}} + \underbrace{\frac{\theta\rho}{2}\sum_{i\in S}\sum_{j\in W} P_{ij}Q_{ij}}_{\text{Status Interaction Receiver}}$$

(A2)

$$U^B(P,Q) = \underbrace{\frac{1}{2}\sum_{i\in R}\sum_{j\in W} P_{ij}Q_{ij}}_{\text{Regular Interaction Sender}} + \underbrace{\frac{1}{2}\sum_{i\in R}\sum_{j\in W} P_{ij}Q_{ij}}_{\text{Regular Interaction Receiver}} + \underbrace{\frac{(1-\theta)\rho}{2}\sum_{i\in S}\sum_{j\in W} P_{ij}Q_{ij}}_{\text{Status Interaction Sender}} + \underbrace{\frac{(1-\theta)\rho}{2}\sum_{i\in S}\sum_{j\in W} P_{ij}Q_{ij}}_{\text{Status Interaction Receiver}}$$

In the 2-state 1 symbol example in the paper, the resulting payoffs for Group A are:

$$U^A(P^u, P^u) = (1-x)^2 + \rho\theta x^2 \tag{A3}$$
$$U^A(P^e, P^u) = (1-x) \tag{A4}$$
$$U^A(P^u, P^e) = (1-x) \tag{A5}$$
$$U^A(P^e, P^e) = 1 \tag{A6}$$

and the payoffs for Group B are

$$U^B(P^u, P^u) = (1-x)^2 + (1-\theta)\rho x^2 \tag{A7}$$
$$U^B(P^e, P^u) = (1-x) \tag{A8}$$
$$U^B(P^u, P^e) = (1-x) \tag{A9}$$
$$U^B(P^e, P^e) = 1 \tag{A10}$$

We are interested in characterizing the stochastic stability of a language convention. Using methods we discuss in Hwang, Naidu and Bowles (2016) and in the Appendix, we find the resistances for each convention, which measure the relative difficulty of idiosyncratic behavior tipping the system from one equilibrium language to another. The resistance from $P^u$ to $P^e$, $c(P^u, P^e)$, and the resistance from $P^e$ to $P^u$, $c(P^e, P^u)$ are given by

$$c(P^u, P^e) = \left\lceil \eta N^R \frac{(1+\rho(1-\theta))x - 1}{(1+\rho(1-\theta))x} \right\rceil \wedge \iota \left\lceil N^R \frac{(1+\rho\theta)x - 1}{(1+\rho\theta)x} \right\rceil$$
$$c(P^e, P^u) = \left\lceil N^R \frac{1}{(1+\rho\theta)x} \right\rceil \wedge \iota \left\lceil \eta N^R \frac{1}{(1+\rho(1-\theta))x} \right\rceil.$$

From this we obtain the following theorem.

THEOREM 1: *Suppose that $x > \frac{1}{1+\rho\theta}$. We have the following characterizations:*
*(i)* ***(Unintentional** $\iota = 1$ **and equal population size** $\eta = 1$**)** *$P^u$ is stochastically stable if and only if $\left\lceil N^R \frac{(1+\rho\theta)x-1}{(1+\rho\theta)x} \right\rceil > \left\lceil N^R \frac{1}{(1+\rho(1-\theta))x} \right\rceil$*
*(ii) (**Intentional** $\iota = \infty$) $P^u$ is stochastically stable if and only if $\left\lceil \eta N^R \frac{(1+\rho(1-\theta))x-1}{(1+\rho(1-\theta))x} \right\rceil > \left\lceil N^R \frac{1}{(1+\rho\theta)x} \right\rceil$*

To show the first prediction in the text, we use Theorem 1 (i). From this, we obtain that convention $P^u$ is stochastically stable if and only if

$$x > \frac{1}{1+\rho\theta} + \frac{1}{1+\rho(1-\theta)} \tag{A11}$$

To study the effect of $\theta, \rho$ on the stochastic stability of convention $P^u$, we first observe that the right hand side of the inequality in (A11) is decreasing in $\rho$ and $\theta$. Also, obviously a higher value of $x$ is more likely to satisfies the inequality in (A11). Thus, the higher $\theta, \rho$, and $x$, the more stochastically stable $P^u$ convention. The second prediction in the text follows from (ii) of Theorem 1. That is, the larger $\eta$, the more likely $P^u$ (the convention favored by the group $B$) is stochastically stable.