CurmElo: The theory and practice of a forced-choice approach to producing preference rankings

Soham Sankaran* & Jacob Derechin*†, Nicholas A. Christakis.

* co-first authors

† corresponding author

October 26, 2019

Abstract

We introduce CurmElo, a forced-choice approach to producing a preference ranking of an arbitrary set of objects that combines the Elo algorithm with novel techniques for detecting and correcting for (1) preference heterogeneity induced polarization in preferences among raters, and (2) intransitivity in preference rankings. We detail the application of CurmElo to the problem of generating approximately preference-neutral identifiers, in this case four-letter and five-letter nonsense words patterned on the phonological conventions of the English language, using a population of Amazon Mechanical Turk workers. We find evidence that human raters have significant non-uniform preferences over these nonsense words, and we detail the consequences of this finding for social science work that utilizes identifiers without accounting for this effect. In addition, we describe how CurmElo can be used to produce rankings of arbitrary features or dimensions of preference of a set of objects relative to a population of raters.

1 Introduction

In this paper, we detail the theory and practice of CurmElo, a forced-choice based approach to producing a preference ranking of an arbitrary set of objects. CurmElo was originally designed for the purpose of producing sets of approximately preference-indifferent identifiers, which we define as identifiers that are relatively equally preferred across a population of subjects. In our original use case, those identifiers were sets of nonsense words of four and five letters.

This work has three motivations. The first motivation is that when eliciting preference, forced-choice based questions are preferable to Likert-style scales in a number of circumstances. The second motivation is that confounding preference for identifiers of various kinds rears its head in numerous unexpected places in social science research, and that it is essential to use some explicit form of preference elicitation, ideally using the population targeted by the research as raters, to control for these effects. The third is that preference heterogeneity induced polarization in preferences among raters and also intransitivity in preference rankings can render naive attempts to control for identifier preference inadequate, and that some method for dealing with these issues is necessary before the preference rankings can be used.

In the section below, we outline the three topics and detail our initial motivating use case for CurmElo, the production of approximately preference-indifferent four-letter and five-letter non-sense identifiers. In the rest of the paper, we use this motivating use case to demonstrate how CurmElo incorporates these insights into a comprehensive method for preference elicitation.

2 Motivations

2.1 Why forced choice?

In many applications, preferences are elicited from raters using Likert-type questions or scales (Likert 1932). Whether numerical (e.g., “rate your preference from 1-5”) or descriptive (e.g., ‘strongly disagree’-‘strongly agree’), the variance inherent in individual perception of points in the scale and ordering effects with regard to samples presented to each subject make it challenging to extract an accurate group-level preference ranking from the data in aggregate. Normalization procedures used to compensate for these issues run the risk of imposing arbitrary ordering based on the specifics of the algorithm used – for example, it has been found that the effect of question wording (positive vs. negative wording) does not generalize across different scales. (Kam 2018) As a result, there is some controversy around the use of such scales, especially single Likert questions as opposed to comparisons across multiple questions, to measure preference and sentiment (Jamieson et al. 2004; Gliem and Gliem 2003; Carifio and Perla 2008). Additionally, within the Likert scale literature, there

are significant inconsistencies about what the optimal size of the scale is. Some empirical results suggest that, consistent with the predictions of information theory, scales with greater numbers of points (1-7 vs 1-11 for example) are better (Alwin 1997), while other empirical results suggest precisely the opposite, that scales with more points tended to be less reliable. (Revilla, Saris, and Krosnick 2014; Alwin and Krosnick 1991) Moreover, the optimal parity – even or odd – of the scale used is also contentious: while a sizeable number of deployments of Likert-type surveys appear to use odd-parity scales, research into these instruments suggests that survey participants will often times use a middle option, only available on an odd-parity scale, to express that they don’t know or don’t have an opinion about the question instead of an actual opinion corresponding with the middle value, even when an “I don’t know” option is available, in many cases materially changing the final results (Sturgis, Roberts, and Smith 2014).

A diverse body of research (Ray 1980; Jackson, Wroblewski, and Ashton 2000; Bartlett, Quay, and Wrightsman Jr 1960; Bartram 2007) including publicly described but unpublished work within the technology industry (Roettgers 2017) suggests that a series of forced-choice binary questions – yes or no, this one or that one – extracts more accurate information about preferences than Likert scales when samples of questions and raters are suitably large.

In our particular use case, a Likert-based preference elicitation method would likely be even more unreliable due to the unfamiliarity of the raters with the identifiers they are being asked to compare – unlike familiar objects like actual English words or, say, human faces, they may have no solid internal baseline for preference for these nonsense words, whereas comparing two identifiers requires no such preexisting knowledge or baseline. CurmElo presents forced-choice questions to raters to avoid these issues.

2.2 Why preference-rank identifiers?

CurmElo was originally designed for the purpose of producing sets of approximately preference-indifferent identifiers, which we define as identifiers that are relatively equally preferred across a population of subjects. In our original use case, those identifiers were sets of nonsense words of four and five letters. While at first glance it may seem reasonable to expect that preference across a set of nonsense words generated randomly will not differ significantly, it is well established that people have innate preferences for particular numbers, letters, and strings of numbers and letters – examples of this include the name-letter effect, where people prefer letters in their own name over others (Nuttin Jr 1985), and the people’s documented preference for the number seven over other single-digit numbers (Heywood 1972; Kubovy and Psotka 1976). Research from cognitive science suggests that the map between the form of a word and its meaning is not entirely arbitrary (Dingemanse, Blasi, Lupyan, Christiansen, and Monaghan 2015), and that human raters impute

category information to nonsense words in systematically different ways (Lupyan and Casasanto 2015). The existence of these preferences is also illustrated by work on the passwords people choose for online services (Riddle, Miron, and Semo 1989; Bonneau 2012).

It seems likely that this sort of identifier preference extends not just to nonsense words, but potentially to any class of object that might be used as an identifier: images, sounds, physical objects, colours, etc. There is work in psychology that suggests that novel and nonsense stimuli of many kinds can prime people just as much as sensical and familiar stimuli (Duckworth, Bargh, Garcia, and Chaiken 2002). This has serious implications for the use of identifiers in experimental social science.

Here is non-exhaustive set of examples of experimental social science work where we believe that identifier preference may be a confounder: work employing the Minimal Group Paradigm (Tajfel 1970; Diehl 1990; Billig and Tajfel 1973), and more generally any work where groups need identifiers; work involving inter-subject interaction where subjects have identifiers; work involving goals or target that need identifiers (our motivating use case fits in this and the previous category); work involving participants reading or listening to narratives where identifiers are used for specific characters. In work of these kinds, we believe that identifier preference, if left unaccounted for, might significantly skew results by heterogeneously changing effect size on a per-identifier basis, as well as make replication difficult due to cross-population preference heterogeneity. We suspect that identifier preference may be unacknowledged confounder for a large number of experiments in these areas.

As such, we believe that preference needs to be explicitly dealt with in some fashion in any social science work where preference for identifiers can be a confounder. This may take several forms. In certain experiments, such as in our motivating use case, one might control for identifier preference by using approximately preference-indifferent identifiers. In other settings, it might be useful to produce identifiers that are quantifiably different, up to a specified tolerance, on some dimension of preference for the raters, for example to measure interaction effects between identifier preference and some other variable.

Preference-conscious identifier generation may also be of value in other empirical or applied circumstances where the objective is to name people, objects or places in such way as to accord them neutral or specific preference of some kind, such as in game design, fiction writing, and bias training.

In this paper, we detail the motivation for the development of CurmElo for our specific use case, that is, issues with identifier preference we observed in our network science experiments, as well demonstrate that randomness in identifier generation and selection do not sufficiently mitigate these effects. We then propose a workable solution.

2.3 Why consider preference heterogeneity induced polarization and intransitivity in preference rankings?

CurmElo uses a version of the Elo algorithm to convert a set of forced-choice binary comparisons within a set of objects into ratings for each of those objects to form a totally-ordered ranking of the set. Consider the case where we want to find preference-indifferent objects of some kind. If we were to interpret these rankings naively, we would extract a subset of objects from the middle of the ranking distribution that are sufficiently similar in rating and call those objects preference-indifferent. It may be the case, however, that some of these objects are not so much preference-indifferent as ‘polarizing’, that is, strongly preferred by one subset of the population and strongly dispreferred by another. This sort of heterogeneity in preference may be the result of some hard-to-detect form of population heterogeneity, and could be a significant confounder if the objects are being used as identifiers in experiments, for example.

CurmElo uses a novel technique based on counting breaks in win-percentage monotonicity in Elo rankings to detect latent heterogeneity and identify polarizing objects. Crucially, this method is distinct from other formulations of the latent population heterogeneity problem since we need to measure no identifying characteristics of the populations other than their choices (Pearl 2017), and as such this could be a valuable method of measuring population-level heterogeneity via preference.

Transitivity is the property that given, some objects a , b , and c , where a is preferred to and ranked above b , and b is ranked above and preferred to c , then a will be preferred to c . To see why breaks in transitivity matter, consider a case where we want to run an experiment to investigate the interaction between identifier preference and some other variable or test condition. Now imagine that our set of identifiers is objects a , b , and c , except that now there is a transitivity break manifesting as a preference cycle such that c is preferred to a . This would completely disrupt any attempt to use preference as an independent variable in the experiment since the ranking is no longer coherent – one cannot say, for example, that a is always most preferred since in this case this is dependent on what it is being compared to – and thus analysis of data collected using these identifiers can produce problematic results. CurmElo uses a technique based on counting breaks in preference transitivity in Elo rankings to identify sets of objects that break transitivity.

2.4 Our motivating use case: four and five letter nonsense identifiers

As part of our ongoing work on human consensus, we needed to find arbitrary identifiers of four-letter and five-letter length to be used as subject and goal identifiers respectively in an experiment where subjects were communicating with each other about a set of goals. We initially decided to use a truncated version of each subject’s Amazon Mechanical Turk HIT ID, a somewhat arbitrary

number-letter string, as subject identifiers, and to use integers from 1-12 to identify goals.

In our initial pilot experiments, we found strong preference effects for specific identifiers, in particular the numbers 9, 1, and 5, with multiple participants subjects clustering asymmetrically around specific identifiers. It has been found that when people are asked to pick a "random" number from an interval, there is clumping around specific parts of interval, so this result is consistent with what is already known (Heywood 1972; Kubovy and Psotka 1976). This made it very difficult isolate the mechanisms of the process we were investigating independently of the identifiers being used. Second, the random subject identifiers, despite their arbitrariness, seemed subject to heterogeneous preference effects that may have strongly influenced the likelihood of subjects to form consensus with other subjects based on their identifiers. These effects were both difficult to predict on a per-identifier basis and difficult to control for in our analysis.

Compounding these issues was the intelligibility of the subject identifiers – random number-letter strings such as ‘341AXM’ are not particularly easy to parse, pronounce, or remember. This effect made our experiment needlessly harder for participants, and may also have explained some of the heterogeneous preference effects given that subject identifiers that seemed closer to familiar words or names, such as ‘30EJON’ (not a real example), appeared to be more preferred.

Given this state of affairs, we set out to produce two distinct sets of identifiers that were both intelligible to our subjects and approximately preference-neutral across the population of subjects we were using for our experiment.

To produce our preference-indifferent identifiers, we first generated two very large sets of four and five-letter identifiers using formats based on the general rules of English phonology to ensure that they were reasonably pronounceable and memorable. For each set, we removed any identifier that had (as of January 2018) been used previously across the large and representative English-language Google Ngrams corpus (Michel, Shen, Aiden, Veres, Gray, Pickett, Hoiberg, Clancy, Norvig, Orwant, et al. 2011). Then, we used a version of the Elo rating system, initially formulated by Arpad Elo to rank chess players (Elo 1978), to derive ratings for each identifier from individual pairwise comparisons to form a population-level ranking. After this, we applied a novel technique based on monotonicity breaks to remove identifiers that might be polarizing but achieve middling values in Elo ratings. Finally, we extracted a set of similarly-rated identifiers from the middle of the ranking distribution. These are our preference-indifferent identifiers.

Our approach allows us both to make claims about which identifiers are equally preferred by raters and also to make claims about which identifiers, overall, are more or less preferred by the raters. For instance, the following pairs of identifiers are equally preferred: (1) camaz and bumak; (2) lujaf and piqez; and (3) cixuq and quhuq. But the first pair is much liked by the raters, the third pair much disliked, and the middle pair has middling ratings (the subjects are neutral about

neutral choices, as it were). Finally, we also provide (in the appendix) a list of 1,000 4-letter and 5-letter identifiers and their ratings, which might be useful to others, from social scientists to fiction writers, facing similar objectives.

3 Preference Data Collection

We first formulated two sets of candidate identifiers using phonological formats that mimic name identifiers the English language (Chomsky and Halle 1968).

For one set of identifiers, we generated all 4-character identifiers of the type:
vowel-consonant-vowel-consonant (VCVC).

For example: ayiz, erik.

There were 11025 of these four-letter identifiers in total.

For the second set of identifiers, we generated all 5-character identifiers of the type:
consonant-vowel-consonant-vowel-consonant (CVCVC).

For example: yezak, roman.

There were 231525 of these five-letter identifiers in total.

We then implemented the following procedure: (1) we removed any identifier that occurred once or more (as of January 2018) in the Google Ngram corpus of published work in English (Michel, Shen, Aiden, Veres, Gray, Pickett, Hoiberg, Clancy, Norvig, Orwant, et al. 2011) (this left 118061 of the five-letter and 5969 of the four-letter identifiers), then (2) we randomly selected 1000 of the remaining identifiers from each of the sets.

These sets of 1000 identifiers were then randomly matched up against each other (within and not across each set), in pairs on our custom-built CurmElo software platform. For each of the two sets, we used 400 unique US-based raters on Amazon Mechanical Turk to perform head-to-head preference comparisons of pairs of identifiers within the set. Note that these AMT raters were recruited from the same population from which we recruit participants in the experiments in which we would subsequently use the identifiers produced using this process, so we have high internal validity for the preference rankings. Each rater was shown 50 random pairs of non-identical identifiers, one pair at a time, and asked the following each time: “Which of the two names below do you prefer? Please do not answer randomly.”

There were 400 workers used for each set, and each worker was shown 50 pairs of identifiers. Given that there are 1000 identifiers in total, each identifier ended up with an average of 40 comparison data points. There is some variation in this number, but no identifier ended up with significantly fewer than 30 comparisons.

Figure 1: A screenshot of the CurmElo interface with two candidate CVCVC identifiers

3.1 Querying other features and dimensions of preference using CurmElo

While in this use case, we asked raters which name identifier they preferred in general, one could use CurmElo to query any other specific feature or dimension of preference. For example, one might ask “Which name sounds better?” or “Which name makes you happier?” or even “Which name seems reddest?”. If your objects are pictures of faces, one might ask “Which face appears angriest?” or “Which face is sharpest?”. The rankings created from the data thus collected would then correspond to the ranking of the objects relative to that specific feature (redness, anger, sharpness) or dimension of preference (sounds, happiness-of-feeling).

3.2 Using non-textual objects

CurmElo can be deployed for any set of objects that can practicably be exposed to raters. In an online-only setting such as Amazon Mechanical Turk, anything displayable on a webpage, including but not limited to audio, images, video, and interactive animations may be used. In a lab setting, physical objects may be used given that they can be uniquely identified and randomized systematically.

4 Theory

4.1 The Elo Algorithm

The Elo algorithm produces a relative rating across a set of objects. The algorithm is initialized by setting all objects to a some common initial rating, R_0 . Then, objects are matched against each other, with some external input determining an outcome where one objects ‘wins’ and the other ‘loses’. In CurmElo, a match is simply a comparison of two objects by a human participant (the external input) being asked to choose a winner and loser among them. Different applications may use different matching systems; for example, if Elo ratings are used for some sort of competitive activity, it may make sense to match objects – in this case players – with similar ratings. In our setting, we use random matching, as it allows the Elo ratings to quickly converge to their stationary

distribution(Jabin and Junca 2015). Consider objects a and b along with their corresponding Elo ratings R_a and R_b . If a and b are matched, and object a wins, the ratings are updated as follows(Elo 1978; Aldous 2017):

$$R'_a = R_a + k \frac{1}{1 + e^{\frac{R_a - R_b}{R_D}}} \quad (1)$$

$$R'_b = R_b - k \frac{1}{1 + e^{\frac{R_a - R_b}{R_D}}} \quad (2)$$

If a and b are matched, and object b wins, the ratings are updated as follows(Elo 1978; Aldous 2017):

$$R'_a = R_a - k \frac{1}{1 + e^{\frac{R_b - R_a}{R_D}}} \quad (3)$$

$$R'_b = R_b + k \frac{1}{1 + e^{\frac{R_b - R_a}{R_D}}} \quad (4)$$

In this setting, k and R_D are free parameters, used to tune how sensitive that rating is to the results of new matches. It is possible to use a broader class of update functions other than $k \frac{1}{1 + e^{\frac{R_a - R_b}{R_D}}}$ as long as it satisfies the conditions for a strong utility distribution, which will be discussed in the next section(Block, Marschak, et al. 1959; Aldous 2017). We use the logistic update function because it is commonly used for Elo applications.

This process continues until all matches – in our case, comparisons – are complete, and we refer to the Elo ratings after all matches have occurred to be the “final Elo rating.” In contrast to applications in sports or gaming, where the number of matches is exogenously built in to the structure of a tournament, in social scientific applications the number of matches can be chosen by the researcher depending on how big a sample of comparisons is needed. Jabin and Junca show that in settings with a large number of objects and intrinsic win probabilities that are not time dependent (such as our motivating example), the distribution of Elo ratings converges to a stationary distribution that represents the underlying preference(Jabin and Junca 2015).

4.2 Stochastic Preferences

Preference is a primitive that underlies many important social phenomena. In this sections, we discuss the basic formalism of deterministic and stochastic preferences.

A preference $\succeq$ must be complete and transitive in order to admit a utility representation . Let A be the finite set of objects. Completeness requires that $\forall a, b \in A$ either $a \succeq b$ or $b \succeq a$. Transitivity requires that $\forall a, b, c \in A$ if $a \succeq b$ and $b \succeq c$ then $a \succeq c$. (Mas-Colell, Whinston, Green, et al. 1995)

In many real systems, choices are stochastic and not deterministic, so the definitions of preferences and transitivity must be extended to accommodate the fact that, in a choice between a

and b , where $a \succeq b$, b will still sometimes be chosen. Block and Marschak extend the notion of preferences by stipulating that when choosing between a and b , $a \succeq b$ if and only if a is chosen with probability greater than or equal to 50% (Block, Marschak, et al. 1959).

Cattelan shows three different ways to apply the definition of transitivity to stochastic choice: Weak Stochastic Transitivity; Moderate Stochastic Transitivity; Strong Stochastic Transitivity. (Cattelan 2012) Let π_{ab} be the probability that a is chosen when the agent is presented with a choice between a and b . Consider $\forall a, b, c \in A$ when $\pi_{ab} \geq .5$ and $\pi_{bc} \geq .5$ if $\pi_{ac} \geq .5$; then $\succeq$ satisfies Weak Stochastic Transitivity; if $\pi_{ac} \geq \min(\pi_{ab}, \pi_{bc})$, then $\succeq$ satisfies Moderate Stochastic Transitivity; or if $\pi_{ac} \geq \max(\pi_{ab}, \pi_{bc})$, then $\succeq$ satisfies Strong Stochastic Transitivity. (Cattelan 2012) Let u_a and u_b be the utility representations for objects a and b respectively. The stochastic definition of preferences also imposes requirements on the probabilities a given object is chosen. Let $\pi_{ab} = W(u_a - u_b)$, where W is the win probability function. (Aldous 2017) W corresponds to the Block and Marschak strong utility distribution and has the following properties: $W : \mathbb{R} \rightarrow (0, 1)$, W is continuous, W is strictly increasing, $\lim_{u \rightarrow \infty} W(u) = 1$, and $W(-u) + W(u) = 1 \forall u \in \mathbb{R}$. (Block, Marschak, et al. 1959; Aldous 2017)

4.3 Heterogeneous Preference, Polarization, and Transitivity Breaks

As discussed in the motivation section, the rankings produced using the Elo algorithm may be subject to the problem of ‘polarizing’ objects resulting from heterogeneous preference. This is the situation where, for some given object, one subset of the population has a strong preference for it and another subset has a strong dispreference for it, and this is not accounted for in the Elo rating. This would manifest in the object being chosen more or less often than its rating would suggest against certain objects, and signals some unobserved heterogeneity within the population. We call this the “polarization in ratings problem” and provide a method to detect when an object is polarized, as well as latent heterogeneity in preference more generally. This method is distinct from other formulations of the latent population heterogeneity problem since we measure no identifying characteristics of the populations other than their choices. (Pearl 2017)

We also provide a method to detect whether an object induces intransitivity in a preference ranking via calculating a normalized ‘transitivity breaks score’ of the number of transitivity breaks in the ranking the object is involved in.

Our methods work on the basis that while, in theory, the Win Probability function must be monotonically increasing and the ratings must satisfy stochastic transitivity for stochastic preferences to be well defined, in practice this is not always the case. Heterogeneity in preference can induce breaks in the monotonicity in the win rate among objects and, intuitively speaking, we ‘count’ the number of these breaks to estimate a normalized ‘polarization score’ (min of 0.0, max

of 1.0) for a given object. In addition, real preferences rankings of various kinds may well be truly intransitive to some degree, and we similarly ‘count’ the number of transitivity breaks an object is involved in to estimate a normalized ‘transitivity breaks score’ (min of 0.0, max of 1.0) for it.

For applications where a well-behaved preference ranking is essential, in particular in order to rely on the predictions of much of the work referenced in the theory section, it is necessary to remove polarizing and transitivity breaking objects.

We first present a Pairwise Polarization Estimator based on monotonicity breaks.

4.3.1 Pairwise Polarization Estimator

The monotonicity assumption of the Elo algorithm is that, for a given object, it should have a higher win rate when compared against lower Elo objects than higher Elo objects. Thus, for a given object, we assess its win rate when compared against all other objects in the set. Next, we look at all pairs of these win rates to see if they match up to the expectations of higher Elo win rates being smaller than lower Elo win rates. We count all violations of this assumption normalized by the number of possible ways this rule could be broken. The process is formalized below.

Assume that there are N total objects for agents to choose from and they are presented in menus of size two. Thus, for each menu, the agent has a choice between two objects, i and j . Let W_{ij} represent the rate at which object i is chosen compared against object j . W_{ijk} refers to the k th sample of a Bernoulli random variable which is 1 if object i is chosen and 0 otherwise, when compared against object j . $\bar{W}_{ij}$ represents the sample estimator of W_{ij} . Let R_i represent the final Elo rating of object i . N is the total number of objects. $\mathcal{N}$ represents the normalization factor and represents the total number of possible breaks in monotonicity implied by the Elo rating. P is the estimator for pairwise polarization.

$$\bar{W}_{ij} = \frac{1}{n} \sum_{k=0}^n W_{ijk} \quad (5)$$

$$\mathcal{N} = \sum_{l=0}^{N-1} (l-1) \quad (6)$$

$$P_i = \frac{1}{\mathcal{N}} \sum_{j \neq i} \sum_{\substack{k \\ R_k < R_j}} \mathbb{1}_{(\bar{W}_{ij} - \bar{W}_{ik} > 0)} \quad (7)$$

4.3.2 Use of Quantiles

In our use case, and in many practical applications where there are a reasonably large number of objects, it would be prohibitively expensive to get enough data points comparing any specific pair to use pairwise estimators with any degree of reliability. Instead, we rely on dividing the objects in the ranking distribution into quantiles, and perform comparisons between a single object and

quantiles to estimate the polarization score of the object.

4.3.3 Quantile Polarization Estimator

Let $\{Q_1, \dots, Q_q\}$ be the q -quantiles of the final Elo distribution R . By convention, quantiles with higher integer values contain lower rated objects. So in a setting with 5 Quantiles, Q_5 refers to the bottom 5th of the Elo distribution and Q_1 refers to the top 5th of the Elo distribution. Let W_{iQ_j} represent the rate at which object i is chosen compared against objects in Q_j . W_{iQ_jk} refers to the k th sample of a Bernoulli random variable which is 1 if object i is chosen and 0 otherwise, when compared against an object in Q_j . By convention, we assume that there were a total of n comparisons of object i against objects in Q_j .

$$\bar{W}_{iQ_j} = \frac{1}{n} \sum_{k=0}^n W_{iQ_jk} \quad (8)$$

$$\mathcal{N} = \sum_{l=0}^q (l-1) \quad (9)$$

$$P_i = \frac{1}{\mathcal{N}} \sum_{j \leq q} \sum_{\substack{k \\ j < k < q}} \mathbb{1}_{(\bar{W}_{iQ_j} - \bar{W}_{iQ_k} > 0)} \quad (10)$$

4.3.4 Quantile Transitivity Breaks Estimator

To count transitivity breaks, for all pairs of quantiles we count the number of times the object is stochastically preferred to a given quantile, while simultaneously not stochastically preferred to a lower quantile than the given quantile. We normalize this count by the number of ways this is possible to produce a ‘transitivity breaks score’ (min of 0.0, max of 1.0). We formalize this process below.

Let $W_{Q_iQ_j}$ represent the rate at which objects in Q_i is chosen compared against objects in Q_j . $W_{Q_iQ_jk}$ refers to the k th sample of a Bernoulli random variable which is 1 if the object in Q_i is chosen and 0 otherwise when compared against an object in Q_j . We call $W_{Q_iQ_j}$ the Inter Quantile Win Rate. If the Elo rating is behaving as expected, one would expect that $W_{Q_iQ_j} > .5$ if $i < j$. This would imply that quantiles with higher rated words tend to be preferred to quantiles with lower rated words. We assume that $W_{Q_iQ_j} > .5$ if $i < j$ as otherwise that implies that the ratings do not represent the preference. We use the definition of weak stochastic transitivity for this estimator.

$$\bar{W}_{Q_iQ_j} = \frac{1}{n} \sum_{k=0}^n W_{Q_iQ_jk} \quad (11)$$

$$\mathcal{N} = \sum_{l=0}^q (l - 1) \quad (12)$$

$$T_i = \frac{1}{\mathcal{N}} \sum_{j \leq q} \sum_{\substack{k \\ j < k < q}} \mathbb{1}_{[(\bar{W}_{iQ_j} < .5) \wedge (\bar{W}_{iQ_k} \geq .5)]} \quad (13)$$

5 Data

We analyzed the data using the following parameters: $k = 20$, $R_0 = 1000$, $R_D = 400$. Table 1 shows the summary statistics for the Elo ratings. The mean Elo rating for both identifiers are both close to R_0 . Additionally, this table shows that there is significant variation in the final Elo ratings of both identifiers. Thus, the fact that the identifiers are arbitrary and nonsensical by construction, then subsequently randomly sampled to produce sets of 1000, does not imply that the identifiers are equally preferred. Table 2 shows summary statistics for the Polarization of each identifier. The summary statistics for polarization are quite similar for both the 4-Letter and 5-Letter identifiers. Table 3 shows the summary statistics for Transitivity breaks for the identifiers. It appears that there are about twice as many breaks in win rate monotonicity as there are in transitivity.

Figures 2 and 3 show histograms of the number of identifiers for each Polarization Score for the 4 and 5 character identifiers respectively. Figures 4 and 5 show histograms of the number of identifiers for each Transitivity Breaks Score for the 4 and 5 character identifiers respectively.

For our final sets of approximately preference-indifferent identifiers of 4 and 5 characters, we looked for identifiers with Elo values between the range of about 990-1010 and filtered out all identifiers with polarization values greater than 0.2. We chose this band because the Elo algorithm was initialized at a value of 1000, so these identifiers are very close to the center of the Elo distribution.

Table 1: Summary Statistics for Elo Ratings

	Mean	Standard Deviation	Min	Max
4-Letter Identifiers	1010.183524	89.31063206	777.725104	1396.567372
5-Letter Identifiers	1022.774282	118.8151442	767.036162	1614.585634

Table 2: Summary Statistics for Polarization

	Mean	Standard Deviation	Min	Max
4-Letter Identifiers	0.2756	0.164047406	0	0.8
5-Letter Identifiers	0.2453	0.155634611	0	1

Table 3: Summary Statistics for Transitivity Breaks

	Mean	Standard Deviation	Min	Max
4-Letter Identifiers	0.1016	0.127254209	0	0.6
5-Letter Identifiers	0.0658	0.104696329	0	0.5

Figure 2: Histogram of 4-character Polarization Scores

Figure 3: Histogram of 5-character Polarization Scores

Figure 4: Histogram of 4-character Transitivity Breaks Scores

6 Analysis

6.1 Inter-Quantile Win Rates

One of the assumptions of the Quantile Polarization Estimator is that the Average Win rates for the quantiles against each other satisfies monotonicity. If there are monotonicity breaks at the quantile level, this indicates departure from the stationary distribution. That could indicate either that the one is using too many quantiles, so there are an insufficient number of samples per quantile, or that the overall number of samples is too low. The Inter Quantile Win Rate Matrix is calculating the average win rate of objects in the quantile represented by the rows against objects in the quantile represented by the columns. For our own application, we used quintiles, so the Inter Quantile Win Rate Matrix is a 5×5 . Table 4 shows the Inter Quantile Win Matrix for the Target 5-character identifiers and Table 5 shows the Inter Quantile Win Matrix for the 4-character identifiers. These have the properties as expected: values close to .5 along the diagonal and monotonicity in win rates.

6.2 Analysis of Polarization and Transitivity Breaks

For the 5-character identifiers we found that 92.1% of them had a nonzero polarization score and for the 4-character identifiers 93.7% had a nonzero polarization score. This suggests that some level of polarization is not uncommon in this kind of preference data. This serves to underscore the importance of testing for polarization in preference data. Breaks in Transitivity were less common with 36% of the 5-character identifiers having a nonzero number of Transitivity breaks and 49.5% of the 4-character identifiers having a nonzero number of Transitivity breaks. This suggests that even in preferences over nonsense words, intransitivity in preference must be accounted for.

We tested the distribution of polarization against the distribution of ratings. If there are objects that are very likely to win against highly rated objects and lose against low rated objects, we would expect their final rating to be in the middle of the distribution. If this is the case, we would expect to find a statistically significant and negative coefficient in a regression where centered Elo ratings are the explanatory variable. Alternatively, if polarization is higher at the tails of the Elo distribution, we would expect the coefficient in the quadratic model to be positive. If Elo rating is not predictive of polarization, we would expect either non-statistically-significant or precisely identified zeros in both linear and quadratic models.

Table 8 shows the results of the linear regression model for the 5-Letter Identifiers. The coefficient on centered Elo ratings is small and not statistically significant. Table 9 shows the results of the quadratic model for the 5-letter Target Identifiers. The coefficient on centered Elo ratings squared is small, positive and not statistically significant. Based on these results, there is no clear

Figure 5: Histogram of 4-character Transitivity Breaks Scores

Table 4: Inter Quantile Win Rates for 5-Letter Identifiers

	5	4	3	2	1
5	0.495723	0.368918	0.300589	0.231730	0.190450
4	0.623822	0.499651	0.418548	0.350965	0.266467
3	0.697948	0.583093	0.503599	0.428252	0.309600
2	0.762648	0.654072	0.551350	0.494170	0.386135
1	0.814080	0.744856	0.679105	0.619651	0.498692

Table 5: Inter Quantile Win Rates for 4-Letter Identifiers

	5	4	3	2	1
5	0.499399	0.401221	0.323861	0.282018	0.227306
4	0.599208	0.503693	0.435614	0.377869	0.292094
3	0.674439	0.549963	0.506442	0.428634	0.328867
2	0.720698	0.627391	0.555608	0.509095	0.388713
1	0.768008	0.706379	0.662835	0.610826	0.501591

relationship between the Elo ratings and the polarization scores. Table 10 shows the results of the linear model for the 4-letter Subject Identifiers. The coefficient on the centered Elo ratings is small and not statistically significant. Table 11 shows the results of the quadratic model for the Subject Identifiers. The coefficient on the centered Elo rating squared is negative, small and not statistically significant. These results are also consistent with the hypothesis that whatever is causing the polarization is uniformly distributed across rating. Based on our setting, it is likely that this is due to unobserved heterogeneity in the population of raters used here. This finding may be relevant to other work using populations of US-based Amazon Mechanical Turk workers, especially work involving preference.

7 Phonological Preference and Polarization: A Further Illustrative Application

Preferences over the identifiers in our corpus could be due to phonological aspects of the identifiers. For example, raters may prefer identifiers that more like a well formed English word than not. Given these types preferences would operate at the linguistic level, one would not expect them to contribute to polarization given that all raters are expected to agree on the phonological conventions of English. Within phonology, other experiments have been conducted using human raters evaluating nonsense words, and have found that how the word is constructed influences how acceptable raters find the word (Ohala and Ohala 1986; Coleman and Pierrehumbert 1997; Frisch, Large, and Pisoni 2000; Bailey and Hahn 2001; Hammond 2004; Albright 2008). Importantly these studies were assessing how much like a real word the raters thought the nonsense words were and presented the words aurally (Bailey and Hahn also ran an experiment with only visual stimulus). (Ohala and Ohala 1986; Coleman and Pierrehumbert 1997; Frisch, Large, and Pisoni 2000; Bailey and Hahn 2001; Albright 2008). These results may not necessarily map onto preferences in affinity over nonsense words. For example, one may recognize that "moist" is a proper English word but that does not necessarily imply that they like it.

Given our particular application, and just for completeness, we tested the impact of the five following phonological constructions on both Elo Rating and Polarization: The first consonant in the word is a nasal (Initial Nasal); the last consonant the word is a voiced obstruent (Terminal Voiced Obstruent); the last consonant the word is voiceless (Terminal Voiceless); the last consonant the word is a fricative (Terminal Fricative); and the last consonant the word is a stop (Terminal Stop). We only use single letter vowels and consonants in our data set, so, for our purposes, the nasals are: ('m', 'n'), the fricatives are: ('f', 's', 'v', 'z'), the stops are: ('p', 't', 'k', 'b', 'd', 'g'), the voiced obstru-

ents are: ('b','d','g','v','z'), and the voiceless consonants are: ('p','t','k','f','s','h','c','x') (Hammond 1999).

Phonological Cue Theory predicts that word terminal fricatives should be preferred, word terminal stops should be dispreferred, nasals early in the word should be preferred, and voiced obstruent in the word terminal position should dispreferred (Wright 2004). Table 6 summarizes the results for our regressions of the phonological constructions on Elo, and the individual models are detailed in Tables 12 - 29. Table 7 summarizes the results for our regressions of the phonological constructions on Polarization, and the individual models are detailed in Tables 30 - 35. The construction Initial Nasal has the most robust effect on Elo, with a statistically significant and positive coefficient in all models. This is consistent with what Phonological Cue Theory predicts. For the rest of the constructions, the results were mixed and *not* entirely consistent with Phonological Cue Theory. With respect to polarization, we find that only the constructions Initial Nasal and Terminal Fricative have a statistically significant relationship. The construction Initial Nasal was found to reduce polarization, which is in line with our predictions, but Terminal Fricative was found to increase polarization, which was surprising. The coefficients for Terminal Fricative in the Elo regression were consistent with the predictions of Phonological Cue Theory, so we expected the presence of this construction to reduce polarization. This suggests that the polarization process is more complex than we expected, and that work in phonology may be unknowingly affected by polarization problems.

It is important not to over-interpret our results given that this was not initially designed as a phonology experiment. For example, we planned to test whether sibilant fricatives in the word initial position impacted the elo ratings, but it turns out none of the potential words in both the 4-letter and 5-letter survived our Google Ngram filter. Since the Ngram filter involves a comparison to real English words, it is possible that the corpora suffer from significant selection bias. In addition, phonological experiments are typically conducted with aural stimuli, and here we have raters visually reading the words. Nonetheless, we still see that our design and the CurmElo system can be of use to experimental phonologists. Of the experiments we surveyed, only Ohala and Ohala use forced choice paired comparison for ratings (Ohala and Ohala 1986), additionally Frisch, Large and Pisoni had a trial that used a binary rating for words (Frisch, Large, and Pisoni 2000); and the rest of the studies use Likert Scales (Coleman and Pierrehumbert 1997; Bailey and Hahn 2001; Albright 2008). We believe that, in this setting, forced choice will perform better than Likert scales for rating applications. It is also worth noting that our number of raters is much larger than those of the experiments we surveyed: Ohala and Ohala had 16 raters in one experiment and 21 raters in a second experiments(Ohala and Ohala 1986); Coleman and Pierrehumbert had 6 raters

Table 6: Elo Linguistics Results Summary

		4-Letter	5-Letter
Initial Nasal	Statistically Significant Sign Consistent	All Models + Yes	All Models + Yes
Terminal Voiced Obstruent	Statistically Significant Sign Consistent	All Models + No	No Models Mixed No
Terminal Voiceless	Statistically Significant Sign Consistent	Some Models + Yes	No Models + Yes
Terminal Fricative	Statistically Significant Sign Consistent	All Models + Yes	No Models + Yes
Terminal Stop	Statistically Significant Sign Consistent	All Models + No	No Models + No

Table 7: Polarization Linguistics Results Summary

		4-Letter	5-Letter
Initial Nasal	Statistically Significant Sign	Yes -	No -
Terminal Voiced Obstruent	Statistically Significant Sign	No +	No -
Terminal Voiceless	Statistically Significant Sign	No +	No -
Terminal Fricative	Statistically Significant Sign	No +	Yes +
Terminal Stop	Statistically Significant Sign	No +	No -

(Coleman and Pierrehumbert 1997); Frisch, Large and Pisoni had two experiments with 24 raters in each arm; and Bailey and Hahn had one experiment with 24 raters and a second experiment with 12 raters (Bailey and Hahn 2001). While some of these results have been shown to replicate, (Hammond 2004; Albright 2008) the number of raters per experiment is still quite low and there may still be reproducibility and generalizability issues that have not been uncovered. The CurmElo system can straightforwardly be adapted to accommodate audio stimuli, so we believe it would be possible to design phonology experiments using CurmElo with a large number of raters relatively easily.

8 Sociocultural Preference

It is well known that preferences over identifiers can be socially mediated. For example heterogeneous response has been documented in audit studies attempting to evaluate ethnoracial bias based on randomly assigning names to putative applicants for jobs (Bertrand and Mullainathan 2004; Booth, Leigh, and Varganova 2012). Audit studies have also shown that the name of the applicant can affect responses to rental applications (Carpusor and Loges 2006; Edelman, Luca, and Svirsky 2017). These naming preferences go both ways as there is significant evidence that patterns of naming children vary based on education and race (Lieberson and Bell 1992; Lieberson and Mikelson 1995; Fryer Jr and Levitt 2004). Thus, one might expect there to be heterogeneity in the preferences in the identifiers in corpora of words such as ours based on these sociocultural factors. It is not entirely straightforward to test this, but we believe a potentially informative approach would be to use CurmElo to produce a ranking of words relative to the features of ‘blackness’ or ‘whiteness’ (in the racial sense) or other axes. Such efforts might be useful in future audit studies.

9 Conclusion

In this paper, we introduced CurmElo, a forced-choice approach to producing a preference ranking of an arbitrary set of object that combines the Elo algorithm with a novel technique for detecting and correcting for heterogeneity and polarization in preferences among raters.

We detailed the application of CurmElo to the problem of generating approximately preference-neutral identifiers, in this case four and five letter nonsense words that are patterned on the phonological conventions of the English language. We provided evidence that human raters have significant preferences over even a randomly selected set of identifiers that were arbitrary and nonsensical by construction, indicating that some method of preference-ranking is necessary to control for preference. We also demonstrate the existence of significant polarization in identifier

preference in our population of US-based Amazon Mechanical Turk raters, indicating both that this heterogeneous preference could have been a significant and tricky confounder if left unaddressed.

We further demonstrated that the preference ranking produced is only somewhat consistent with the predictions of existing work in phonological preference, in particular that polarization appears to affect phonological features of words that are predicted to increase preference by Phonological Cue Theory, suggesting that experiments in phonology based on preference would benefit from using CurmElo to detect and control for such polarization. While our CurmElo phonology experiments have much larger subject populations and numbers of data points than the phonology work we reference, our experiments were not originally designed for phonological analysis and as such suffer from selection (real words removed) and presentation (visual versus aural) issues, so they are limited.

We believe that the polarization-corrected Elo framework we detail is a theoretically strong method for generating preference rankings. In particular, we see it as superior to Likert scales for the purposes of extracting a population’s preference ranking of a large number of objects. We believe that CurmElo could be deployed confidently across a wide range of settings where there may be unobserved heterogeneity in the target population, and that it is a robust method for preference elicitation generally, and identifier generation specifically, across a variety of domains.

We also believe that approximately preference-indifferent identifiers should be used in any social science work where preference for identifiers can be a confounder, for example for subject and group identifiers in work employing the Minimal Group Paradigm or Vignette Studies involving arbitrary names. We believe that identifier preference is an unacknowledged confounder for many experiments of this nature, in particular in experiments in using Amazon Mechanical Turk populations, for which we have already demonstrated significantly non-uniform identifier preference and preference polarization. CurmElo can be used to produce rankings of arbitrary features or dimensions of preference of a set of objects relative to a population of raters.

References

- Albright, Adam. 2008. "From clusters to words: grammatical models of nonce word acceptability." *Handout of talk presented at 82nd LSA, Chicago* .
- Aldous, David. 2017. "Elo ratings and the sports model: A neglected topic in applied probability?" *Statistical Science* 32:616–629.
- Alwin, Duane F. 1997. "Feeling thermometers versus 7-point scales: Which are better?" *Sociological Methods & Research* 25:318–340.
- Alwin, Duane F and Jon A Krosnick. 1991. "The reliability of survey attitude measurement: The influence of question and respondent attributes." *Sociological Methods & Research* 20:139–181.
- Bailey, Todd M and Ulrike Hahn. 2001. "Determinants of wordlikeness: Phonotactics or lexical neighborhoods?" *Journal of Memory and Language* 44:568–591.
- Bartlett, Claude J, Lorene Childs Quay, and Lawrence S Wrightsman Jr. 1960. "A comparison of two methods of attitude measurement: Likert-type and forced choice." *Educational and Psychological Measurement* 20:699–704.
- Bartram, Dave. 2007. "Increasing validity with forced-choice criterion measurement formats." *International Journal of Selection and Assessment* 15:263–272.
- Bertrand, Marianne and Sendhil Mullainathan. 2004. "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination." *American economic review* 94:991–1013.
- Billig, Michael and Henri Tajfel. 1973. "Social categorization and similarity in intergroup behaviour." *European Journal of Social Psychology* 3:27–52.
- Block, Henry David, Jacob Marschak, et al. 1959. "Random orderings and stochastic theories of response." Technical report, Cowles Foundation for Research in Economics, Yale University.
- Bonneau, Joseph. 2012. "The science of guessing: analyzing an anonymized corpus of 70 million passwords." In *Security and Privacy (SP), 2012 IEEE Symposium on*, pp. 538–552. IEEE.
- Booth, Alison L, Andrew Leigh, and Elena Varganova. 2012. "Does ethnic discrimination vary across minority groups? Evidence from a field experiment." *Oxford Bulletin of Economics and Statistics* 74:547–573.
- Carifio, James and Rocco Perla. 2008. "Resolving the 50-year debate around using and misusing Likert scales." *Medical education* 42:1150–1152.
- Carpusor, Adrian G and William E Loges. 2006. "Rental Discrimination and Ethnicity in Names." *Journal of Applied Social Psychology* 36:934–952.

- Cattelan, Manuela. 2012. "Models for paired comparison data: A review with emphasis on dependent data." *Statistical Science* pp. 412–433.
- Chomsky, Noam and Morris Halle. 1968. "The sound pattern of English." .
- Coleman, John and Janet Pierrehumbert. 1997. "Stochastic phonological grammars and acceptability." *arXiv preprint cmp-lg/9707017* .
- Diehl, Michael. 1990. "The minimal group paradigm: Theoretical explanations and empirical findings." *European review of social psychology* 1:263–292.
- Dingemanse, Mark, Damián E Blasi, Gary Lupyan, Morten H Christiansen, and Padraic Monaghan. 2015. "Arbitrariness, iconicity, and systematicity in language." *Trends in cognitive sciences* 19:603–615.
- Duckworth, Kimberly L, John A Bargh, Magda Garcia, and Shelly Chaiken. 2002. "The automatic evaluation of novel stimuli." *Psychological science* 13:513–519.
- Edelman, Benjamin, Michael Luca, and Dan Svirsky. 2017. "Racial discrimination in the sharing economy: Evidence from a field experiment." *American Economic Journal: Applied Economics* 9:1–22.
- Elo, Arpad E. 1978. *The rating of chessplayers, past and present*. Arco Pub.
- Frisch, Stefan A, Nathan R Large, and David B Pisoni. 2000. "Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords." *Journal of memory and language* 42:481.
- Fryer Jr, Roland G and Steven D Levitt. 2004. "The causes and consequences of distinctively black names." *The Quarterly Journal of Economics* 119:767–805.
- Gliem, Joseph A and Rosemary R Gliem. 2003. "Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales." Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education.
- Hammond, Michael. 1999. *The Phonology of English: A Prosodic Optimality-Theoretic Approach: A Prosodic Optimality-Theoretic Approach*. Oxford University Press, UK.
- Hammond, Michael. 2004. "Gradience, phonotactics, and the lexicon in English phonology." *International Journal of English Studies* 4:1–24.
- Heywood, Simon. 1972. "The popular number seven or number preference." *Perceptual and Motor Skills* 34:357–358.
- Jabin, Pierre-Emmanuel and Stéphane Junca. 2015. "A continuous model for ratings." *SIAM Journal on Applied Mathematics* 75:420–442.

- Jackson, Douglas N, Victor R Wroblewski, and Michael C Ashton. 2000. "The impact of faking on employment tests: Does forced choice offer a solution?" *Human Performance* 13:371–388.
- Jamieson, Susan et al. 2004. "Likert scales: how to (ab) use them." *Medical education* 38:1217–1218.
- Kam, Chester Chun Seng. 2018. "Why Do We Still Have an Impoverished Understanding of the Item Wording Effect? An Empirical Examination." *Sociological Methods & Research* 47:574–597.
- Kubovy, Michael and Joseph Psotka. 1976. "The predominance of seven and the apparent spontaneity of numerical choices." *Journal of Experimental Psychology: Human Perception and Performance* 2:291.
- Lieberson, Stanley and Eleanor O Bell. 1992. "Children's first names: An empirical study of social taste." *American Journal of Sociology* 98:511–554.
- Lieberson, Stanley and Kelly S Mikelson. 1995. "Distinctive African American names: An experimental, historical, and linguistic analysis of innovation." *American Sociological Review* pp. 928–946.
- Likert, Rensis. 1932. "A technique for the measurement of attitudes." *Archives of psychology* .
- Lupyan, Gary and Daniel Casasanto. 2015. "Meaningless words promote meaningful categorization." *Language and Cognition* 7:167–193.
- Mas-Colell, Andreu, Michael Dennis Whinston, Jerry R Green, et al. 1995. *Microeconomic theory*, volume 1. Oxford university press New York.
- 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* 331:176–182.
- Nuttin Jr, Jozef M. 1985. "Narcissism beyond Gestalt and awareness: The name letter effect." *European Journal of Social Psychology* 15:353–361.
- Ohala, John J and Manjari Ohala. 1986. "Testing hypotheses regarding the psychological manifestation of morpheme structure constraints." *Experimental phonology* pp. 239–252.
- Pearl, Judea. 2017. "Detecting Latent Heterogeneity." *Sociological Methods & Research* 46:370–389.
- Ray, John J. 1980. "The comparative validity of Likert, projective, and forced-choice indices of achievement motivation." *The Journal of Social Psychology* 111:63–72.
- Revilla, Melanie A., Willem E. Saris, and Jon A. Krosnick. 2014. "Choosing the Number of Categories in Agree–Disagree Scales." *Sociological Methods & Research* 43:73–97.

- Riddle, Bruce L, Murray S Miron, and Judith A Semo. 1989. "Passwords in use in a university timesharing environment." *Computers & Security* 8:569–579.
- Roettgers, Janko. 2017. "Netflix Replacing Star Ratings With Thumbs Ups and Thumbs Downs." *Variety* <http://variety.com/2017/digital/news/netflix-thumbs-vs-stars-1202010492/>.
- Sturgis, Patrick, Caroline Roberts, and Patten Smith. 2014. "Middle alternatives revisited: how the neither/nor response acts as a way of saying "i don't know"?" *Sociological Methods & Research* 43:15–38.
- Tajfel, Henri. 1970. "Experiments in intergroup discrimination." *Scientific American* 223:96–103.
- Wright, Richard. 2004. "A review of perceptual cues and cue robustness." *Phonetically based phonology* pp. 34–57.

10 Tables and Figures

10.1 Main Regression Tables

Dep. Variable:	rank_breaks	R-squared:	0.000
Model:	OLS	Adj. R-squared:	-0.001
Method:	Least Squares	F-statistic:	0.002043
Date:	Fri, 11 May 2018	Prob (F-statistic):	0.964
Time:	16:51:59	Log-Likelihood:	-1860.8
No. Observations:	1000	AIC:	3726.
Df Residuals:	998	BIC:	3735.
Df Model:	1		

	coef	std err	z	P> z	[0.025	0.975]
Intercept	2.4526	0.051	48.497	0.000	2.353	2.552
centered	1.826e-05	0.000	0.045	0.964	-0.001	0.001

Omnibus:	84.487	Durbin-Watson:	2.036
Prob(Omnibus):	0.000	Jarque-Bera (JB):	108.155
Skew:	0.714	Prob(JB):	3.27e-24
Kurtosis:	3.745	Cond. No.	123.

Table 8: 5-Letter Identifiers Linear Regression

Dep. Variable:	rank_breaks	R-squared:	0.000
Model:	OLS	Adj. R-squared:	-0.001
Method:	Least Squares	F-statistic:	0.1152
Date:	Fri, 11 May 2018	Prob (F-statistic):	0.734
Time:	16:51:41	Log-Likelihood:	-1860.7
No. Observations:	1000	AIC:	3725.
Df Residuals:	998	BIC:	3735.
Df Model:	1		

	coef	std err	z	P> z	[0.025	0.975]
Intercept	2.4444	0.056	43.648	0.000	2.335	2.554
np.power(centered, 2)	5.908e-07	1.74e-06	0.339	0.734	-2.82e-06	4e-06

Omnibus:	84.774	Durbin-Watson:	2.036
Prob(Omnibus):	0.000	Jarque-Bera (JB):	108.655
Skew:	0.715	Prob(JB):	2.55e-24
Kurtosis:	3.749	Cond. No.	3.59e+04

Table 9: 5-Letter Identifiers Quadratic Regression

Dep. Variable:	rank_breaks	R-squared:	0.000
Model:	OLS	Adj. R-squared:	-0.001
Method:	Least Squares	F-statistic:	0.02374
Date:	Fri, 11 May 2018	Prob (F-statistic):	0.878
Time:	16:52:06	Log-Likelihood:	-1913.4
No. Observations:	1000	AIC:	3831.
Df Residuals:	998	BIC:	3841.
Df Model:	1		

	coef	std err	z	P> z	[0.025	0.975]
Intercept	2.7570	0.052	53.254	0.000	2.656	2.858
centered	-9.626e-05	0.001	-0.154	0.878	-0.001	0.001

Omnibus:	44.566	Durbin-Watson:	1.921
Prob(Omnibus):	0.000	Jarque-Bera (JB):	49.735
Skew:	0.546	Prob(JB):	1.59e-11
Kurtosis:	3.053	Cond. No.	90.4

Table 10: 4-Letter Identifiers Linear Regression

Dep. Variable:	rank_breaks	R-squared:	0.006
Model:	OLS	Adj. R-squared:	0.005
Method:	Least Squares	F-statistic:	4.120
Date:	Fri, 11 May 2018	Prob (F-statistic):	0.0426
Time:	16:51:49	Log-Likelihood:	-1910.5
No. Observations:	1000	AIC:	3825.
Df Residuals:	998	BIC:	3835.
Df Model:	1		

	coef	std err	z	P> z	[0.025	0.975]
Intercept	2.6811	0.062	43.107	0.000	2.559	2.803
np.power(centered, 2)	9.28e-06	4.57e-06	2.030	0.042	3.19e-07	1.82e-05

Omnibus:	42.909	Durbin-Watson:	1.932
Prob(Omnibus):	0.000	Jarque-Bera (JB):	47.692
Skew:	0.535	Prob(JB):	4.40e-11
Kurtosis:	3.040	Cond. No.	1.84e+04

Table 11: 4-Letter Identifiers Quadratic Regression

10.2 Linguistic Regression Tables

Dep. Variable:	elo	R-squared:	0.015			
Model:	OLS	Adj. R-squared:	0.014			
Method:	Least Squares	F-statistic:	15.07			
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.000110			
Time:	19:47:17	Log-Likelihood:	-6188.3			
No. Observations:	1000	AIC:	1.238e+04			
Df Residuals:	998	BIC:	1.239e+04			
Df Model:	1					
	coef	std err	z	P> z	[0.025	0.975]
const	1016.3492	4.059	250.385	0.000	1008.393	1024.305
initial_nasal	40.1567	10.343	3.883	0.000	19.885	60.428
	Omnibus:	127.444	Durbin-Watson:	0.033		
	Prob(Omnibus):	0.000	Jarque-Bera (JB):	200.368		
	Skew:	0.864	Prob(JB):	3.09e-44		
	Kurtosis:	4.351	Cond. No.	2.81		

Table 12: 5-Letter Identifiers Elo vs Initial Nasal

Dep. Variable:	elo	R-squared:	0.001			
Model:	OLS	Adj. R-squared:	0.000			
Method:	Least Squares	F-statistic:	1.116			
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.291			
Time:	19:47:17	Log-Likelihood:	-6195.5			
No. Observations:	1000	AIC:	1.239e+04			
Df Residuals:	998	BIC:	1.240e+04			
Df Model:	1					
	coef	std err	z	P> z	[0.025	0.975]
const	1019.6074	5.003	203.814	0.000	1009.802	1029.412
terminal_voiceless	7.9770	7.552	1.056	0.291	-6.824	22.778
Omnibus:	124.464	Durbin-Watson:	0.003			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	189.968			
Skew:	0.862	Prob(JB):	5.61e-42			
Kurtosis:	4.261	Cond. No.	2.45			

Table 13: 5-Letter Identifiers Elo vs Terminal Voiceless Consonant

Dep. Variable:	elo	R-squared:	0.000
Model:	OLS	Adj. R-squared:	-0.001
Method:	Least Squares	F-statistic:	0.009826
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.921
Time:	19:47:17	Log-Likelihood:	-6196.0
No. Observations:	1000	AIC:	1.240e+04
Df Residuals:	998	BIC:	1.241e+04
Df Model:	1		

	coef	std err	z	P> z	[0.025	0.975]
const	1022.9633	4.342	235.596	0.000	1014.453	1031.474
terminal_obstruents	-0.8554	8.629	-0.099	0.921	-17.768	16.057

Omnibus:	121.775	Durbin-Watson:	0.001
Prob(Omnibus):	0.000	Jarque-Bera (JB):	183.882
Skew:	0.851	Prob(JB):	1.18e-40
Kurtosis:	4.230	Cond. No.	2.55

Table 14: 5-Letter Identifiers Elo vs Terminal Voiced Obstruent

Dep. Variable:	elo	R-squared:	0.016
Model:	OLS	Adj. R-squared:	0.014
Method:	Least Squares	F-statistic:	8.151
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.000308
Time:	19:47:17	Log-Likelihood:	-6187.9
No. Observations:	1000	AIC:	1.238e+04
Df Residuals:	997	BIC:	1.240e+04
Df Model:	2		

	coef	std err	z	P> z	[0.025	0.975]
const	1013.6635	5.129	197.632	0.000	1003.611	1023.716
initial_nasal	39.8241	10.388	3.834	0.000	19.464	60.184
terminal_voiceless	6.8990	7.528	0.916	0.359	-7.856	21.654

Omnibus:	129.467	Durbin-Watson:	0.036
Prob(Omnibus):	0.000	Jarque-Bera (JB):	205.081
Skew:	0.872	Prob(JB):	2.93e-45
Kurtosis:	4.372	Cond. No.	3.07

Table 15: 5-Letter Identifiers Elo vs Initial Nasal and Terminal Voiceless Consonant

Dep. Variable:	elo	R-squared:	0.016
Model:	OLS	Adj. R-squared:	0.013
Method:	Least Squares	F-statistic:	5.454
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.00101
Time:	19:47:17	Log-Likelihood:	-6187.8
No. Observations:	1000	AIC:	1.238e+04
Df Residuals:	996	BIC:	1.240e+04
Df Model:	3		

	coef	std err	z	P> z	[0.025	0.975]
const	1012.2185	6.751	149.930	0.000	998.986	1025.451
initial_nasal	39.8237	10.384	3.835	0.000	19.472	60.175
terminal_voiceless	8.3442	8.710	0.958	0.338	-8.726	25.415
terminal_obstruents	3.9430	9.837	0.401	0.689	-15.337	23.223

Omnibus:	130.985	Durbin-Watson:	0.036
Prob(Omnibus):	0.000	Jarque-Bera (JB):	208.912
Skew:	0.877	Prob(JB):	4.32e-46
Kurtosis:	4.392	Cond. No.	3.67

Table 16: 5-Letter Identifiers Elo vs Initial Nasal, Terminal Voiceless Consonant, and Terminal Voiced Obstruent

Dep. Variable:	elo	R-squared:	0.001			
Model:	OLS	Adj. R-squared:	-0.000			
Method:	Least Squares	F-statistic:	1.033			
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.310			
Time:	19:47:17	Log-Likelihood:	-6195.5			
No. Observations:	1000	AIC:	1.240e+04			
Df Residuals:	998	BIC:	1.240e+04			
Df Model:	1					
	coef	std err	z	P> z	[0.025	0.975]
const	1021.0435	4.198	243.217	0.000	1012.815	1029.272
terminal_fricative	9.5621	9.408	1.016	0.309	-8.878	28.002
Omnibus:	123.381	Durbin-Watson:	0.003			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	187.486			
Skew:	0.858	Prob(JB):	1.94e-41			
Kurtosis:	4.248	Cond. No.	2.70			

Table 17: 5-Letter Identifiers Elo vs Terminal Fricative

Dep. Variable:	elo	R-squared:	0.001			
Model:	OLS	Adj. R-squared:	-0.000			
Method:	Least Squares	F-statistic:	0.5403			
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.462			
Time:	19:47:17	Log-Likelihood:	-6195.8			
No. Observations:	1000	AIC:	1.240e+04			
Df Residuals:	998	BIC:	1.241e+04			
Df Model:	1					
	coef	std err	z	P> z	[0.025	0.975]
const	1021.1161	4.498	227.036	0.000	1012.301	1029.931
terminal_stop	6.0078	8.173	0.735	0.462	-10.011	22.026
Omnibus:	123.301	Durbin-Watson:	0.002			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	187.515			
Skew:	0.857	Prob(JB):	1.91e-41			
Kurtosis:	4.251	Cond. No.	2.45			

Table 18: 5-Letter Identifiers Elo vs Terminal Stop

Dep. Variable:	elo	R-squared:	0.002
Model:	OLS	Adj. R-squared:	0.000
Method:	Least Squares	F-statistic:	1.042
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.353
Time:	19:47:17	Log-Likelihood:	-6195.0
No. Observations:	1000	AIC:	1.240e+04
Df Residuals:	997	BIC:	1.241e+04
Df Model:	2		

	coef	std err	z	P> z	[0.025	0.975]
const	1017.9530	5.301	192.025	0.000	1007.563	1028.343
terminal_fricative	12.6527	9.949	1.272	0.203	-6.848	32.153
terminal_stop	9.1710	8.641	1.061	0.289	-7.766	26.107

Omnibus:	125.877	Durbin-Watson:	0.005
Prob(Omnibus):	0.000	Jarque-Bera (JB):	193.449
Skew:	0.866	Prob(JB):	9.84e-43
Kurtosis:	4.281	Cond. No.	3.24

Table 19: 5-Letter Identifiers Elo vs Terminal Fricative and Terminal Stop

Dep. Variable:	elo	R-squared:	0.017
Model:	OLS	Adj. R-squared:	0.014
Method:	Least Squares	F-statistic:	5.872
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.000565
Time:	19:47:17	Log-Likelihood:	-6187.3
No. Observations:	1000	AIC:	1.238e+04
Df Residuals:	996	BIC:	1.240e+04
Df Model:	3		

	coef	std err	z	P> z	[0.025	0.975]
const	1011.6812	5.443	185.867	0.000	1001.013	1022.349
terminal_fricative	12.7265	9.837	1.294	0.196	-6.554	32.007
terminal_stop	8.6200	8.615	1.001	0.317	-8.265	25.505
initial_nasal	40.0655	10.360	3.867	0.000	19.760	60.371

Omnibus:	131.246	Durbin-Watson:	0.037
Prob(Omnibus):	0.000	Jarque-Bera (JB):	209.601
Skew:	0.878	Prob(JB):	3.06e-46
Kurtosis:	4.396	Cond. No.	3.30

Table 20: 5-Letter Identifiers Elo vs Terminal Fricative, Terminal Stop, and Initial Nasal

Dep. Variable:	elo	R-squared:	0.014			
Model:	OLS	Adj. R-squared:	0.013			
Method:	Least Squares	F-statistic:	13.69			
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.000228			
Time:	19:47:17	Log-Likelihood:	-5903.3			
No. Observations:	1000	AIC:	1.181e+04			
Df Residuals:	998	BIC:	1.182e+04			
Df Model:	1					
	coef	std err	z	P> z	[0.025	0.975]
const	1007.3665	2.898	347.604	0.000	1001.686	1013.047
initial_nasal	43.3391	11.715	3.700	0.000	20.379	66.300
Omnibus:	61.410	Durbin-Watson:	0.030			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	74.083			
Skew:	0.588	Prob(JB):	8.19e-17			
Kurtosis:	3.627	Cond. No.	4.07			

Table 21: 4-Letter Identifiers Elo vs Initial Nasal

Dep. Variable:	elo	R-squared:	0.003			
Model:	OLS	Adj. R-squared:	0.002			
Method:	Least Squares	F-statistic:	2.694			
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.101			
Time:	19:47:17	Log-Likelihood:	-5909.2			
No. Observations:	1000	AIC:	1.182e+04			
Df Residuals:	998	BIC:	1.183e+04			
Df Model:	1					
	coef	std err	z	P> z	[0.025	0.975]
const	1006.6304	3.532	285.034	0.000	999.709	1013.552
terminal_voiceless	9.6553	5.882	1.641	0.101	-1.874	21.185
Omnibus:	59.339	Durbin-Watson:	0.006			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	70.363			
Skew:	0.585	Prob(JB):	5.26e-16			
Kurtosis:	3.565	Cond. No.	2.42			

Table 22: 4-Letter Identifiers Elo vs Terminal Voiceless Consonant

Dep. Variable:	elo	R-squared:	0.011
Model:	OLS	Adj. R-squared:	0.010
Method:	Least Squares	F-statistic:	12.26
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.000484
Time:	19:47:17	Log-Likelihood:	-5905.0
No. Observations:	1000	AIC:	1.181e+04
Df Residuals:	998	BIC:	1.182e+04
Df Model:	1		

	coef	std err	z	P> z	[0.025	0.975]
const	1004.3789	3.394	295.911	0.000	997.726	1011.031
terminal__obstruents	21.0314	6.007	3.501	0.000	9.257	32.806

Omnibus:	68.273	Durbin-Watson:	0.023
Prob(Omnibus):	0.000	Jarque-Bera (JB):	84.435
Skew:	0.622	Prob(JB):	4.63e-19
Kurtosis:	3.691	Cond. No.	2.45

Table 23: 4-Letter Identifiers Elo vs Terminal Voiced Obstruent

Dep. Variable:	elo	R-squared:	0.018
Model:	OLS	Adj. R-squared:	0.016
Method:	Least Squares	F-statistic:	8.767
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.000168
Time:	19:47:17	Log-Likelihood:	-5901.5
No. Observations:	1000	AIC:	1.181e+04
Df Residuals:	997	BIC:	1.182e+04
Df Model:	2		

	coef	std err	z	P> z	[0.025	0.975]
const	1003.1573	3.593	279.201	0.000	996.115	1010.199
initial_nasal	44.7962	11.742	3.815	0.000	21.781	67.811
terminal_voiceless	11.1807	5.858	1.909	0.056	-0.301	22.663

Omnibus:	60.388	Durbin-Watson:	0.037
Prob(Omnibus):	0.000	Jarque-Bera (JB):	72.137
Skew:	0.588	Prob(JB):	2.17e-16
Kurtosis:	3.592	Cond. No.	4.42

Table 24: 4-Letter Identifiers Elo vs Initial Nasal and Terminal Voiceless Consonant

Dep. Variable:	elo	R-squared:	0.038
Model:	OLS	Adj. R-squared:	0.035
Method:	Least Squares	F-statistic:	12.44
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	5.44e-08
Time:	19:47:17	Log-Likelihood:	-5891.0
No. Observations:	1000	AIC:	1.179e+04
Df Residuals:	996	BIC:	1.181e+04
Df Model:	3		

	coef	std err	z	P> z	[0.025	0.975]
const	989.1805	4.890	202.280	0.000	979.596	998.765
initial_nasal	42.8709	11.457	3.742	0.000	20.416	65.326
terminal_voiceless	25.2412	6.727	3.752	0.000	12.057	38.425
terminal_obstruents	32.3465	6.839	4.730	0.000	18.943	45.750

Omnibus:	72.454	Durbin-Watson:	0.081
Prob(Omnibus):	0.000	Jarque-Bera (JB):	90.873
Skew:	0.643	Prob(JB):	1.85e-20
Kurtosis:	3.726	Cond. No.	4.55

Table 25: 4-Letter Identifiers Elo vs Initial Nasal, Terminal Voiceless Consonant, and Terminal Voiced Obstruent

Dep. Variable:	elo	R-squared:	0.003			
Model:	OLS	Adj. R-squared:	0.002			
Method:	Least Squares	F-statistic:	2.770			
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.0964			
Time:	19:47:17	Log-Likelihood:	-5909.3			
No. Observations:	1000	AIC:	1.182e+04			
Df Residuals:	998	BIC:	1.183e+04			
Df Model:	1					
	coef	std err	z	P> z	[0.025	0.975]
const	1007.9550	3.195	315.490	0.000	1001.693	1014.217
terminal_fricative	11.3121	6.797	1.664	0.096	-2.010	24.635
Omnibus:	62.066	Durbin-Watson:	0.006			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	74.614			
Skew:	0.596	Prob(JB):	6.28e-17			
Kurtosis:	3.609	Cond. No.	2.63			

Table 26: 4-Letter Identifiers Elo vs Terminal Fricative

Dep. Variable:	elo	R-squared:	0.013			
Model:	OLS	Adj. R-squared:	0.012			
Method:	Least Squares	F-statistic:	13.31			
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.000278			
Time:	19:47:17	Log-Likelihood:	-5904.1			
No. Observations:	1000	AIC:	1.181e+04			
Df Residuals:	998	BIC:	1.182e+04			
Df Model:	1					
	coef	std err	z	P> z	[0.025	0.975]
const	1003.6893	3.363	298.432	0.000	997.097	1010.281
terminal_stop	22.3171	6.117	3.648	0.000	10.328	34.307
Omnibus:	63.532	Durbin-Watson:	0.028			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	76.680			
Skew:	0.604	Prob(JB):	2.23e-17			
Kurtosis:	3.617	Cond. No.	2.43			

Table 27: 4-Letter Identifiers Elo vs Terminal Stop

Dep. Variable:	elo	R-squared:	0.021
Model:	OLS	Adj. R-squared:	0.019
Method:	Least Squares	F-statistic:	10.74
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	2.44e-05
Time:	19:47:17	Log-Likelihood:	-5899.8
No. Observations:	1000	AIC:	1.181e+04
Df Residuals:	997	BIC:	1.182e+04
Df Model:	2		

	coef	std err	z	P> z	[0.025	0.975]
const	997.6954	4.021	248.151	0.000	989.815	1005.575
terminal_fricative	21.5718	7.222	2.987	0.003	7.417	35.727
terminal_stop	28.3109	6.502	4.354	0.000	15.568	41.054

Omnibus:	68.137	Durbin-Watson:	0.046
Prob(Omnibus):	0.000	Jarque-Bera (JB):	83.539
Skew:	0.627	Prob(JB):	7.24e-19
Kurtosis:	3.657	Cond. No.	3.26

Table 28: 4-Letter Identifiers Elo vs Terminal Fricative and Terminal Stop

Dep. Variable:	elo	R-squared:	0.036			
Model:	OLS	Adj. R-squared:	0.033			
Method:	Least Squares	F-statistic:	11.80			
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	1.35e-07			
Time:	19:47:17	Log-Likelihood:	-5892.2			
No. Observations:	1000	AIC:	1.179e+04			
Df Residuals:	996	BIC:	1.181e+04			
Df Model:	3					
	coef	std err	z	P> z	[0.025	0.975]
const	994.6756	4.073	244.231	0.000	986.693	1002.658
terminal_fricative	21.6764	7.167	3.025	0.002	7.630	35.723
terminal_stop	28.7500	6.468	4.445	0.000	16.073	41.427
initial_nasal	44.1755	11.624	3.800	0.000	21.393	66.958
Omnibus:	69.990	Durbin-Watson:	0.077			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	86.930			
Skew:	0.632	Prob(JB):	1.33e-19			
Kurtosis:	3.700	Cond. No.	4.35			

Table 29: 4-Letter Identifiers Elo vs Terminal Fricative, Terminal Stop, and Initial Nasal

Dep. Variable:	rank_breaks	R-squared:	0.002
Model:	OLS	Adj. R-squared:	0.000
Method:	Least Squares	F-statistic:	1.283
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.278
Time:	22:26:53	Log-Likelihood:	-1859.6
No. Observations:	1000	AIC:	3725.
Df Residuals:	997	BIC:	3740.
Df Model:	2		

	coef	std err	z	P> z	[0.025	0.975]
const	2.4947	0.068	36.550	0.000	2.361	2.628
initial_nasal	-0.2034	0.129	-1.578	0.115	-0.456	0.049
terminal_voiceless	-0.0231	0.100	-0.232	0.817	-0.218	0.172

Omnibus:	83.175	Durbin-Watson:	2.034
Prob(Omnibus):	0.000	Jarque-Bera (JB):	105.922
Skew:	0.709	Prob(JB):	9.99e-24
Kurtosis:	3.730	Cond. No.	3.07

Table 30: 5-Letter Identifiers Polarization vs Initial Nasal and Terminal Voiceless Consonant

Dep. Variable:	rank_breaks	R-squared:	0.002			
Model:	OLS	Adj. R-squared:	-0.001			
Method:	Least Squares	F-statistic:	0.8567			
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.463			
Time:	22:26:53	Log-Likelihood:	-1859.6			
No. Observations:	1000	AIC:	3727.			
Df Residuals:	996	BIC:	3747.			
Df Model:	3					
	coef	std err	z	P> z	[0.025	0.975]
const	2.4963	0.084	29.599	0.000	2.331	2.662
initial_nasal	-0.2034	0.129	-1.576	0.115	-0.456	0.049
terminal_voiceless	-0.0247	0.112	-0.221	0.825	-0.244	0.194
terminal_obstruents	-0.0044	0.134	-0.033	0.974	-0.267	0.258
Omnibus:	83.181	Durbin-Watson:	2.034			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	105.940			
Skew:	0.709	Prob(JB):	9.89e-24			
Kurtosis:	3.730	Cond. No.	3.67			

Table 31: 5-Letter Identifiers Polarization vs Initial Nasal, Terminal Voiceless Consonant, and Terminal Voiced Obstruent

Dep. Variable:	rank_breaks	R-squared:	0.007			
Model:	OLS	Adj. R-squared:	0.004			
Method:	Least Squares	F-statistic:	2.034			
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.107			
Time:	22:26:53	Log-Likelihood:	-1857.5			
No. Observations:	1000	AIC:	3723.			
Df Residuals:	996	BIC:	3743.			
Df Model:	3					
	coef	std err	z	P> z	[0.025	0.975]
const	2.4497	0.071	34.569	0.000	2.311	2.589
terminal_fricative	0.2501	0.140	1.784	0.074	-0.025	0.525
terminal_stop	-0.0348	0.111	-0.314	0.753	-0.252	0.183
initial_nasal	-0.2019	0.129	-1.566	0.117	-0.455	0.051
Omnibus:	81.069	Durbin-Watson:	2.040			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	102.251			
Skew:	0.701	Prob(JB):	6.26e-23			
Kurtosis:	3.699	Cond. No.	3.30			

Table 32: 5-Letter Identifiers Polarization vs Terminal Fricative, Terminal Stop, and Initial Nasal

Dep. Variable:	rank_breaks	R-squared:	0.005			
Model:	OLS	Adj. R-squared:	0.003			
Method:	Least Squares	F-statistic:	2.715			
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.0667			
Time:	22:26:53	Log-Likelihood:	-1910.7			
No. Observations:	1000	AIC:	3827.			
Df Residuals:	997	BIC:	3842.			
Df Model:	2					
	coef	std err	z	P> z	[0.025	0.975]
const	2.7255	0.065	42.232	0.000	2.599	2.852
initial_nasal	-0.3777	0.201	-1.881	0.060	-0.771	0.016
terminal_voiceless	0.1496	0.110	1.360	0.174	-0.066	0.365
Omnibus:	42.601	Durbin-Watson:	1.915			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	47.276			
Skew:	0.532	Prob(JB):	5.42e-11			
Kurtosis:	3.053	Cond. No.	4.42			

Table 33: 4-Letter Identifiers Polarization vs Initial Nasal and Terminal Voiceless Consonant

Dep. Variable:	rank_breaks	R-squared:	0.006				
Model:	OLS	Adj. R-squared:	0.003				
Method:	Least Squares	F-statistic:	2.053				
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.105				
Time:	22:26:53	Log-Likelihood:	-1910.4				
No. Observations:	1000	AIC:	3829.				
Df Residuals:	996	BIC:	3848.				
Df Model:	3						
		coef	std err	z	P> z	[0.025	0.975]
const		2.6859	0.086	31.338	0.000	2.518	2.854
initial_nasal		-0.3832	0.201	-1.907	0.056	-0.777	0.011
terminal_voiceless		0.1894	0.124	1.533	0.125	-0.053	0.432
terminal_obstruents		0.0915	0.127	0.723	0.470	-0.157	0.340
Omnibus:		42.704	Durbin-Watson:		1.914		
Prob(Omnibus):		0.000	Jarque-Bera (JB):		47.399		
Skew:		0.533	Prob(JB):		5.10e-11		
Kurtosis:		3.055	Cond. No.		4.55		

Table 34: 4-Letter Identifiers Polarization vs Initial Nasal,Terminal Voiceless Consonant, and Terminal Voiced Obstruent

Dep. Variable:	rank_breaks	R-squared:	0.005			
Model:	OLS	Adj. R-squared:	0.002			
Method:	Least Squares	F-statistic:	1.955			
Date:	Wed, 29 Aug 2018	Prob (F-statistic):	0.119			
Time:	22:26:53	Log-Likelihood:	-1910.7			
No. Observations:	1000	AIC:	3829.			
Df Residuals:	996	BIC:	3849.			
Df Model:	3					
	coef	std err	z	P> z	[0.025	0.975]
const	2.7241	0.074	36.706	0.000	2.579	2.870
terminal_fricative	0.0480	0.136	0.354	0.724	-0.218	0.314
terminal_stop	0.1648	0.122	1.351	0.177	-0.074	0.404
initial_nasal	-0.3922	0.202	-1.939	0.053	-0.789	0.004
Omnibus:	43.935	Durbin-Watson:	1.911			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	48.905			
Skew:	0.541	Prob(JB):	2.40e-11			
Kurtosis:	3.067	Cond. No.	4.35			

Table 35: 4-Letter Identifiers Polarization vs Terminal Fricative, Terminal Stop, and Initial Nasal

11 Appendix

4-Letter	5-Letter
ivek	pesam
asiq	jayor
avox	kewos
ejon	mavey
omog	coyam
otif	fexen
ojav	pebel
ebud	jupar
ufix	cosop
anix	lozir
awek	koduk
epek	kovaq
izus	mazob
owek	camaz
ozis	bumak
azop	kujol
ixel	falom
emev	poqit
opaw	gosaw
uzus	hosop
uzok	madik
ovay	jutah
axoy	naroc
ozir	lenim
imeg	lelip
atux	gomos
otiz	nubul
uvap	paqel
awax	nuvis
izop	kozec
okoy	dajav
uboz	hamok

ebak	luwog
icoy	korav
enud	puzat
omaj	coxar
oyek	kehab
ebek	gupac
ibuy	hugum
ezip	nala.j
ovox	jaxam
eviv	nejix
ewop	kayus
equm	helad
ezub	nawam
opaz	neroy
enux	maviz
ikab	nipiq
ewoq	cebis
axah	gixan
emuv	mibob
osaz	banom
okap	lilap
ubob	nadog
evuz	pizaz
uvaq	pukap
oxef	deneh
eluz	melox
ufam	nowan
uweg	pefay
ovaq	hazud
enub	finoy
oliq	qavan
oyix	cazup
uzab	mezed
emoc	carir
ewiz	moruz

ecox	kisop
oxoh	mideh
umeg	bavez
etug	peber
oxim	jomic
abux	kojeb
ageg	dapup
ojay	midux
ayom	koyev
ewut	laral
ipox	luxos
ehoc	cupun
ugot	lobud
ezej	neliz
umaj	luzaf
agud	hawoh
udur	kogas
azuc	jusol
omeb	mefer
ecup	mavaz
eqiv	maqet
uwan	livud
izip	mumek
ezot	pixed
azod	pagol
usiv	jatet
awup	dacoy
uzim	jalav
ovax	fozer
opeg	leyah
upad	casuk
oxaz	hidar
obeq	mikaw
ezac	demox
ukiz	hinaj

oniq	ceyos
esoy	mofex
ebaj	gozik
ayug	maqis
ihub	bajoc
umaw	kimoz
uqiz	jezed
afav	jufah
owat	jolot
odiz	mifun
udog	humeg
uxal	miroj
uraj	panuf
olab	cunim
uwos	naroy
odaj	fazos
edav	kivam
oqal	lezul
emix	meqac
atez	pedax
adug	gupil
aqan	lohaw
ogix	jareh
ataj	povip
afab	gudem
adez	deyal
exaz	mosiv
urib	micax
esug	luqem
ojep	culah
owok	panoc
ufom	jawok
inaz	jaluk
obuk	gidem
adoy	pimac

umix	gemey
ovix	fazov
onoq	kuzac
eyab	koriv
uvir	defos
icaw	bokac
iwid	cufay
evev	luciy
odip	bageg
oqim	fazip
ozec	dafir
iwal	levid
ezif	nizaz
ucaw	kehid
ehum	diriv
ejav	namuc
esok	mogip
axog	mekeb
omeq	mepoh
imeq	nayim
olaq	kusit
oteq	jecid
eraw	bomav
umez	nirab
upez	hurod
oqos	mutiz
oket	hunaq
aqic	meyir
osuk	cudip
aqom	qador
ozit	korit
azij	cinok
ohep	kuyem
aqed	maheg
azub	pogok

avaw	hodev
ojeb	koyem
izax	kenux
odab	muvoc
idoz	famoy
ofaj	pemeg
aqac	duzin
izig	mivon
ojaw	muvit
aduw	gulac
ucog	hikop
esuz	miloc
urix	legom
axeb	davuk
atix	dedow
ezeb	pujom
uzid	pocor
orob	fabox
opaq	jigas
uzif	lulaq
ojad	pizah
avux	gawus
ubaz	ganur
uqil	gumad
atev	pufun
owuk	foraj
owol	kadul
ozaw	muluq
emoz	lifix
uzev	cofic
ajun	hukar
ocuz	jahaw
apaj	mojaw
usay	bejon
odid	hisaf

ojex	necuz
azuy	payom
esiz	gumon
ameq	decek
ejox	kebad
ebop	kulem
okif	nuxos
ijun	geroz
oyok	jixil
uleq	gemud
uxir	diroc
ewun	mojic
icaj	kuwis
axez	neyus
ocaz	jurib
atug	jusev
ezog	mafum
ubix	jajoy
oyak	bulem
arux	bapas
oqey	jexic
areq	mafur
opup	laxip
esog	lamo
oxux	natap
onip	kefoz
aqaq	gunek
ozat	kugos
oqon	qetas
udiz	pudew
ukul	kegit
agug	namot
exoc	juzil
axuz	bemaf
aqij	finij

ovus	negaz
uxer	niwal
egev	kedew
enex	canof
ihol	fanap
ayaf	hopaf
uxix	lugof
ohiv	lejag
osot	cadej
acax	melaq
ozej	golug
aleq	bacoy
izik	hixit
ehuz	bodof
uruq	nidac
igak	jukug
utiq	kopip
eqob	gaqon
azup	cuzoy
ujok	hadof
ataw	cecig
ijex	maded
ezoh	hikas
uxim	laqiz
ozov	mumaf
iqex	jufat
ejul	jahev
ezup	foxuy
ebaf	butaz
eqon	locun
ajag	muhaz
ujen	nofen
efob	pamaq
eqot	lipob
uqad	mopot

acag	baziq
osih	lecu
ozaj	jedap
usex	nisev
uqal	doqor
ejom	fawun
ulot	busix
abej	bizej
unuv	bamuc
okof	mufur
ulob	nofiz
izot	bozit
ogaf	bobeh
ogaw	cusud
obuf	buseq
apec	boliy
ohuz	hagac
esif	dihet
osoy	lajak
isuc	cepoz
opef	miseh
ateg	luyak
aqad	janoz
ibud	fabix
afep	jelag
apew	pakih
oxap	pevid
emiz	nanic
idug	babap
uboy	hocun
omob	forek
ubop	nugeb
ogem	bonij
obeg	fehaz
esul	husig

ujaj	mexoh
izec	galuv
imaw	fajul
anuz	payex
oxat	fanuv
evaw	jupel
imom	camux
izep	mabub
enib	dodeg
odoh	qehad
oxex	firup
ipeq	cusem
ejim	japaf
ojug	bewej
ekit	lokip
azeh	bizuq
ilox	nejux
oxeb	gufas
oyey	luveh
efux	fahah
ajav	bafen
oyoj	butaf
ezex	lidag
ozob	dimal
egif	qovel
ekux	domav
iduh	bajuw
ejiv	bexod
eqoc	gilox
opuf	copox
epof	lisiy
ucaz	dutos
usuk	loqen
emuji	jewem
egur	pomew

isuv	hiwer
iwek	lapaq
imif	kisid
ewof	dawos
akiy	dowir
owef	harok
ejem	gesex
eneg	ciriv
ajof	kenuy
obiz	jiwin
avew	pibab
atey	lofow
ikox	lufim
osob	faroq
oqaz	dawom
ivip	hikip
eruz	jekoq
exeb	qapor
uvub	nobuq
akeb	hayup
efay	pucax
iwiz	lijat
icev	bamam
uqox	gigig
izuj	bavim
ubot	godoh
ureh	lofic
ekaf	neqet
oqet	kujih
owik	fekey
edup	dizus
ojil	dimuz
epum	cetef
upuq	juwac
aseg	cimoz

uqaw	jahuh
uzub	mibom
ixoy	kagat
izoc	bijek
upaj	gakaz
efaq	ducoz
iwof	kuroc
avuk	bocut
umuy	maxih
utiv	fapod
ecuf	manup
ivuz	kuza j
igab	focom
efib	hehum
ozuz	jiroy
edih	gupeq
idix	neyat
obef	jucin
iwoj	husaw
ipob	nalew
asiv	qinik
ekat	qenon
ihak	buzaz
axak	bibil
owet	gemeh
ohax	jizof
izex	jeqon
amup	fuxum
exuh	qukum
ehaf	kajip
uhip	qimix
ekik	biwuk
uzuf	hokuj
ubox	muhaf
usuq	fijot

awur	dayip
owug	hahek
ojon	jirem
ofok	mituh
ivaw	hucaw
ukoz	gitab
ubup	nigis
ibek	hizid
uziz	moruv
ezuv	boveq
opiw	qayuk
ajif	kafux
ogab	qotey
udub	micok
ixeb	lepey
exoz	nibat
ucah	bilaq
umav	dijeh
uwab	bisup
ekew	furub
uyox	japoz
ufub	deqeb
ehax	dukup
ixak	heyoz
oyis	gipum
uvim	gavuc
odod	bucom
onuj	capeb
owud	kazuj
ozix	nupub
otaj	jekox
ayuf	kawir
ujur	noyat
ajeb	niqin
aniw	cuxur

asuh	dijun
ibup	bumuy
ucib	bopad
ezuw	nuyak
uteb	kofup
evaj	mifep
uwec	ponum
ehix	gewom
iboq	lovuj
ojix	qaric
eff	kakox
ocav	cotoj
utaj	jowog
okut	dobot
ihoz	kipoh
axaq	bakux
ocox	bumec
omoq	nelit
eqin	kuqoc
otof	mibad
uxas	febaq
oxop	qacam
orud	fojub
uhim	gapib
aqux	giqar
uyac	pelub
uwap	jinop
asiw	lafec
ogoc	laxub
oqeb	cicum
efok	dosoz
ojel	bihir
ekok	mugeh
unih	fizeb
exuk	qamuh

uloj	bohud
okuc	kexuw
ufuy	lulew
awaf	foyar
ikak	bipam
ehof	hihug
oqan	maxaq
ipuz	nifum
iwak	fugum
icud	notak
ebup	mehor
avep	gujad
uboc	loxab
ifeg	biyom
ufug	maxoj
eteb	fojix
eguk	qupep
ofov	fihar
obeb	bofef
owev	dodas
uvut	padeb
avuv	gigox
izaaj	bumaf
ohiz	fipev
ejij	jehas
eqis	gokuv
eyix	jujuc
iwuq	mudut
uluv	deguv
upuv	hidep
ubag	loxoq
ecow	cudol
odij	qozus
osib	dibed
eqec	domif

etew	ceriy
oqur	muxil
awut	kosob
osah	pelif
odiq	fariw
ohet	qaxol
ituk	lujaf
uhoc	piqez
upip	diwab
uxaq	koxus
ewit	lanoq
uwur	dabaf
ocig	buduf
ocix	jutur
okog	bexoh
uvaj	gafux
ejap	jiyag
epab	fazuf
emob	keleq
ufez	qameh
uvow	japum
ahiq	hexaq
ujes	cibam
ewav	kesuv
ukuy	cokud
omeh	dekip
udej	perap
ekak	hitej
onaf	pisof
otih	hedul
ogep	nuvok
afoq	biwar
ewad	hivoy
osuz	liper
ihib	nunif

iqem	qirab
uhod	mabod
ofeb	govez
utav	kekap
omih	giniq
utif	haceh
eqol	jafaf
esud	nolur
ayuq	buqov
oyur	burug
uwit	fahij
owep	gonuc
oyic	qomap
onuv	nagiq
ekuh	fapif
aziy	koguf
ahch	kixan
ujip	qobeh
iqom	cukep
uqat	gafig
uluh	qujet
ahuc	hidiz
utib	kamip
aliw	pewom
acuh	dafum
uveq	cuzul
ezaw	dahow
upuj	pexih
ipej	joyuk
ogaj	diqom
ofib	niniq
ewik	meyiy
enag	jaroj
idux	pecoq
inuj	cucuk

iqew	gerev
ezuc	becef
utof	boruy
owav	nefag
apuv	nicoq
unub	facuz
ewuf	jidit
ayuy	jekiy
isiz	fasij
ofuk	fukoh
iyit	kunag
igoc	kacex
ajuy	qemor
ifaf	pivun
ivoh	nomep
okiv	dilux
ecej	cecay
ihaq	biseq
iyeh	bowal
uvud	dedik
ukec	luhid
atuy	qodeb
utog	kimof
akof	nogex
imuh	jataj
ovug	fisok
ehoq	qawep
ucux	badod
ucax	kujod
ikum	fahuk
ekoz	jojuy
akeg	gudik
ovib	foxih
uzuq	gewus
uhux	hetaq

apuv	nakix
ihig	leqop
uyov	juyox
eviz	lihih
abuq	lojut
azuw	mohiw
ojoz	cebuq
uvek	dejiw
ifiz	qoroc
egeb	nipoz
ojuh	baxud
ixid	cukip
ucuf	lagoc
iqed	kigog
ekid	joxiy
equx	jucev
ugos	qemak
ujoz	lezax
aduv	besow
uxuy	guyep
ikuz	hozej
ecur	qecim
uyip	decix
ucod	gosac
eyep	juwem
ewuh	gegiv
owid	bufav
upuz	dihep
ajep	ciqog
iloq	qiruc
ajiv	nuzot
inup	didiv
owig	dofuc
ufit	lowop
atuj	hifan

ocof	fiteb
uboa	doli
ucuw	bedif
edeg	ciziy
ogeb	baqey
ibef	naqup
iwum	linuw
uyut	dokek
oqek	fujed
enaw	cukol
ogog	fizim
edaq	fopiv
eziy	qehab
uzeb	jecir
icag	bijag
edaj	lojed
ahoz	dogoc
ogoq	febum
igah	fafuf
afuz	pikaq
uweq	nojiy
eqez	bodid
iwe	fizuz
iqef	midut
egux	bequv
ovof	nibod
axuj	cunoj
owom	guluw
ulim	dadef
ucay	mumop
ujir	pehem
uwal	kecic
uzuw	qucir
aqov	coyul
ugic	lohuh

icuv	heqij
izux	gapif
uvoq	kiyir
atuh	leqem
oxud	jukig
ixow	fovif
ozuy	qegoy
ekoc	deruq
ava:j	huhuk
acew	pahif
eduw	nuxok
umew	bomip
ivez	fuyoh
ecuh	nonoj
osij	jaqew
ixaf	qegem
ufeb	ho:il
asoz	mugiy
ohuj	fijiv
ikof	gisif
efiq	nuhuz
uvej	bihib
ugeg	jagij
eboj	mufuf
ixal	poj:ik
oqib	lacuc
onoj	hodud
obup	bobod
ewip	humiy
ujus	hodog
ihel	nanuw
ukuk	papiq
usux	mipef
uyux	pezuh
ij:q	qayiv

eqes	dudoq
ufoh	mukiy
ayow	kogiy
ulej	qepar
ecaf	jajix
uvab	jojaq
equh	cagiq
ihuw	kugux
ayiw	hefom
iwac	qamom
ejeb	qonip
iduy	jemej
epuk	giviq
ujim	jociy
uhiz	qibog
ekoq	navep
ehog	pojuv
ohub	gupok
ukaq	joviw
abup	geluq
ofec	fevut
iwip	mijiz
iwiq	mapej
uhuv	pifil
ehuq	lawuv
eyir	luwub
opih	gedij
ugor	duyit
ofuq	husit
idej	newew
uxop	juwos
ugiz	nijef
iwuc	heham
ipib	gimaw
ixac	hufuh

itaw	cikoz
ehoh	dewil
uxam	neriy
uvav	qeguz
uyeh	kugog
oguw	qugac
iqoz	goqaj
uzof	kukij
uxac	moqac
evuq	qexix
axuc	geciv
uqef	qapem
oqut	juzud
uyif	qimec
omuy	jofoc
unaj	jebut
uniz	pituf
aguf	pobuz
ikuk	ququl
oqez	pomiy
oqul	pewep
uyem	nuguk
uwil	fisof
aqiy	jufad
iwuy	dufic
ibip	dajuf
uwaw	neqeq
ahek	qucic
uyis	jufej
utej	muxuk
uqiq	bigaj
uzug	mikox
opoq	kazij
icak	baziy
ipul	femof

uqiw	keqif
ejev	jekeh
ujux	qebux
epuy	gibep
umub	fapih
ufob	peqom
ebip	bewef
ituw	qecut
uhih	kixex
ugiv	fuqeh
ojip	gasuq
otoj	gosod
uneb	qewaj
uyey	gicuk
iduw	jaqej
ewiv	lejjj
ekiy	codej
uvux	kocuj
efoh	cunuh
udud	baqoc
egay	cigup
udoj	kuhix
unuy	huqiv
icuh	pewuk
uroj	gahoz
upuf	duzug
ejep	fimup
ageq	qeyaw
ixod	qofis
iyut	jaxug
oxak	gitif
ayeq	bapef
icih	futuj
ofih	juyed
ezeg	gileq

ubij	miguw
igex	boqon
uyaw	capeh
ewiq	gixub
obuj	meyiq
uneh	geqor
ovuq	pucoc
ozoq	pahih
oqak	qiruq
iviq	cecec
uriy	hepoj
ojux	qoneb
imuw	lunij
igaq	jepiv
ofoc	qawaj
oyiv	qepov
iyej	qimuz
iwom	kigay
upiw	biqix
esuw	qupov
ijol	huvib
ujud	cemif
awih	nagip
ugeh	dedav
uyeb	quhec
ugaq	nuhux
aqey	bapup
uzaj	fuqug
ojig	nacih
ulij	qidew
afeg	jucuj
ukib	difiv
uqex	boguw
uxut	goqeg
uziy	hociy

oqiw	ciyip
uxaw	bowiv
oxuj	fufob
ipuj	gevip
iwet	hajoj
otuj	hofoc
ihif	nekuc
equq	qicot
awus	qexay
upej	furep
iqen	ceqem
adih	fihoq
aqaj	qihux
owiv	muyoy
oqok	foviz
ugej	qaqac
iruj	jujif
eyoj	qekah
ixud	qamoh
egew	jixek
aqeg	noduq
ahef	cicaj
obik	kuxaj
uvep	qoyin
ucuv	qoqaw
iyag	mihid
afoj	guyaj
uraw	ligeg
ufiv	norur
uqeg	lereg
ucoh	qadug
ohix	nosif
iyev	fugiq
ujiv	giheb
uyul	gowoc

oqic	guqur
uwuz	qediý
ujif	jucew
ixub	gugej
oxoq	cewoh
ugez	guxun
ugih	qevoj
ugeq	desaq
iquw	qafun
eqoj	jyuy
uqoj	cefuc
ocij	guheh
iwev	gexud
ubuq	jijob
ugop	ligib
ugaj	gigir
ihog	nujuk
awoj	fukik
ifeh	gosux
iyiq	fideh
iyif	hiziq
igub	hehil
uyez	gusuz
uxev	fodeh
ihiy	nuzoq
iqug	kogiw
ukiw	qejev
uyeq	jilih
ukij	hihuh
ujaq	qexag
ejiw	nutuj
uguf	fakag
uwiq	qanej
ipuw	qezuv
igew	peqef

ijaj	pigef
ubih	noxoj
oqof	qagiw
idiy	qoqiq
ikux	liliw
ifof	nafug
iqet	feget
uvih	guhiz
uwux	cusij
oquv	peqeg
ucuj	qavew
uhew	hejic
ihiq	qafij
afuj	fixej
oxok	cacij
egej	jufeh
iqaj	fowiv
ukuh	joqiw
uduj	qosid
uyuw	qupof
okiw	gijek
uluy	gabiy
egub	biquc
omuq	qijaq
oqig	qoyih
ixij	ditij
iceh	jigej
ojiy	guqip
ujoq	gufek
oqug	jikeq
uriw	qeyaf
uyiw	guhuy
udik	fucuj
iquq	gewih
ufof	qupef

iriw	qaviw
iwih	fepoj
uxej	geqog
uruw	qefib
uzup	gogib
iyaj	leheq
iyuf	cahiy
uxuj	cijej
aqej	hisoq
esih	gqiq
uxuk	cuwuq
ujih	cixuq
oqiq	quhuq
eyik	qixew
iruw	gihin
uyoq	qaqep
igay	gayij
ogiw	gikif
owoq	bibiw
exuw	hegug
usiy	fihuw
uxij	qiyag
ihih	nufiw

12 Note on Replication

12.1 Data Collection with CurmElo

Data Collection can be accomplished in two ways: using the custom CurmElo software and using the CurmElo for Qualtrics script. The software for both of these methods is in the `curmelo-software` subdirectory of the main replication package directory.

12.1.1 Data Collection using the custom CurmElo software

Set up a Firebase Realtime Database instance using the instructions here: <https://firebase.google.com/docs/database/>. Then set `fbase_URL` to the URL of your firebase instance, re-

placing `INSERT_FIREBASE_URL_HERE`. Set a unique `run_name` each time you are ranking a new corpus. Set `filename` to be the filename of your corpus, which should be a text file containing a separate object from the set you want to rank on each line. Currently, it is a text file (`example_objects_to_rank.txt`) containing some three-letter test identifiers. Modify the `html` in the `h2h` function to customize the welcome message and question asked at each comparison. With a bit of work, you can make it display images, play sounds or videos, or show interactive animations – the basic structure of the system shouldn't change. Modify `num_q_per_worker` in `h2h` to change the number of comparisons each rater is asked to make. You can deploy this app using python's built-in test HTTP server or using gunicorn using `run_curmelo.sh` and `run_curmelo_gunicorn.sh`. You can also use other web servers using WSGI or similar. Once you've set up CurmElo, send your raters to `YOUR_DEPLOYMENT_URL/headtohead`.

12.1.2 Data Collection with CurmElo for Qualtrics

We created a python script that can generate a list of pairwise comparison questions that can be imported into Qualtrics and used with the Qualtrics platform. The file `curmelo_for_qualtrics.py` takes a CSV file with all alternatives in one column with no headers as input and outputs a file of questions that can be imported into qualtrics. Inside the script the user can set the text of the comparison question. It is not recommended to use CurmElo for Qualtrics when the set of alternatives is large as there are some randomization settings within Qualtrics that seem to need to be done for each individual question by hand.

12.2 Identifier Generation and Pruning

Scripts to generate and prune the four and five letter identifiers are in the `curmelo-identifier-generation-and-pruning` subdirectory of the main replication package directory. These `generate_and_prune_fiveletter_idens.py` and `generate_and_prune_fourletter_idens.py` scripts do what you'd expect them to.

All software and data files referenced in the sections below are in the `curmelo-data-and-analysis` subdirectory of the main replication package directory.

12.3 Creating an Elo ranking

Use the `elo_calc_fbase.py` script to produce a ranking. The `sanity_check_results_amt.py` script can be used to ensure that there are no inconsistencies between the worker log and the CurmElo data if you use Amazon Mechanical Turk to recruit your raters.

JSON files of our own collected data are also available in the `curmelo-data-and-analysis` subdirectory under the names `4_names_41_400_v2.json` and `5_names_1000_random_400_workers.json`.

The rankings produced by these data files are stored in

`4_names_41_400_v2_mean_1000_kfac_20_rpa_400_elo.json` and

`5_names_1000_random_400_workers_mean_1000_kfac_20_rpa_400_elo.json` respectively.

12.4 Monotonicity breaks (polarization) and transitivity breaks

The file `polarization_final.py` is a direct replication file for recreating the analysis related to Polarization, Transitivity Breaks and the Intra Tranche Win Rate. The file `polarization_functions.py` is designed to be used as an import file and contains documentation for all the functions used in `polarization_final.py`. To reproduce the bar plots used in the paper use the file `bar_plot.py`. We also include the file `elo_analysis.py`, which runs some normality tests on the Elo distributions as well as generates some counterfactual distributions. It turned out that this work was not needed but for completeness sake we included it here.

12.5 Phonological analysis

The file `ling_analysis.py` is a direct replication file for recreating the phonological analysis used in this paper. This script generates dummy variables for the various phonological patterns we test using pandas string manipulation. These are then used in the regressions described in the paper.