What's in a Name?

Saku Aura
University of Missouri and CESifo

and

Gregory D. Hess
Claremont McKenna College and CESifo [*]

February 2004

Abstract

Plenty. This paper analyzes two broad questions: Does your first name matter? And how did you get your first name anyway? Using data from the National Opinion Research Centers (NORC's) General Social Survey, including access to respondents first names from the 1994 and 2002 surveys, we extract the important "first name features" (FNF), e.g. popularity, number of syllables, phonetic features, Scrabble score, 'blackness (i.e. the fraction of people with that name who are black), etc ... We then explore whether these first name features are useful explanatory factors of an respondent's exogenous background factors (sex, race, parent's education, etc...) and lifetime outcomes (e.g. financial status, occupational prestige, perceived social class, education, happiness, and whether they became a parent before 25). We find that first name features on their own do have significant predictive power for a number of these lifetime outcomes, even after controlling for a myriad of exogenous background factors. We find evidence that first name features are independent predictors of lifetime outcomes that are likely related to labor productivity such as education, happiness and early fertility. Importantly, however, we also find evidence based on the differential impacts of gender and race on the blackness of a name and its popularity that suggest that discrimination may also be a factor.

JEL Codes: D1, J1, J7.

Keywords: Names, Identity, Discrimination.

What's in a name?
That which we call a rose by any other name would smell as sweet.
William Shakespeare, <u>Romeo and Juliet</u>.

So I give ya that name and I said goodbye
I knew you'd have to get tough or die
And it's that name that helped to make you strong
Johnny Cash (1969), <u>A Boy Named Sue</u>.

1 Introduction

Expectant parents lie awake at night, consult books, and some even hire a consultant to choose their new child's name. Is it similar to the process that manufacturers undertake when branding a new product? Viagara pretty much speaks for itself, but to what extent does Gregory, Saku or even Jamaal convey information and/or meaning. Our study explores two main questions. First, does a person's name convey information about their background? Second, does a person's name have an impact on the person's long run economic outcomes, such as income, education, fertility, social standing, happiness or prestige?

Recently economists have turned to the analysis and impact of an individuals first name.[1] The emphasis of this analysis has been on the impact of having an African American Name. More formally, these studies contrast the outcomes of an individual having a name that signal high probability of being black with outcomes individual that has a name that signals a high probability of being white. In our dataset, this would mean comparing the outcomes of otherwise similar individuals with a name like Mark (exclusively a white name in our data) with Marcus (exclusively a black name in our data) or comparing Alice (white name) with Tanisha (black name).

---

[1]Names have been for a longer period a topic of research for other social scientist and linguistics. Sociologist Stanley Lieberson, in work with co-authors, has considered many of the question we list in our opening paragraph, e.g. Lieberson & Bell (1992), Lieberson & Kelly (1995) and Lieberson (2000). They provide an extended descriptive discussion on what influences parent's choice of first names for their children and extensively document static and dynamic features of first-name distributions in the US across time. While it does not do full justice to their rich analysis, it could be said that their framework is mostly concentrated around the question on what influences pure parental preference for first names as opposed to the more structural economic mechanisms considered here.

In a compelling study, Bertrand and Mullainathan (2003) undertake a carefully controlled experimental study on the extent to which an African American first name affects ones job prospects. Constructing a set of synthetic resumes that differ based on whether the applicant has an African American name,[2] the researchers send these resumes to actual job openings listed in the newspapers of a number of major cities. Importantly, they find that in order to receive one job interview, a resume with a black name needs to be sent to 15 openings. In contrast, to receive one job opening, a resume with a white name needs to be sent to only 10 openings. Furthermore, by using auxiliary information from birth certificates from Massachusetts that also list mother's education Bertrand and Mullainathan construct measures of expected maternal education level for each name in their sample. Their results imply that it is the racial information conveyed in the name and not the parental background factor signaled by the name (parental education) that potential employers are using as the basis to select between resumes.[3] They conclude that this shows that racial discrimination is pervasive in the job market.

A similar set of issues is addressed in another recent study by Fryer and Levitt (2003). They make two main contributions to our understanding of the economic role of first names. First, they outline four approaches to the economic analysis of first names: namely, ignorance, price theory model, a signalling model, and an identity model. The first model speaks for itself: it assumes that parents ignore or are unaware of any consequences of their child's first name. The price model, assumes that parents choose their child's name to maximize their expected utility. If indeed, an individual of social group A is more likely to be in society dominated by group A (B) individuals, then their parents will choose names that are more in keeping with names from group A (B), as such a choice is likely to increase their child's economic opportunities. The signaling model, suggests that parents signal their type (i.e. are they a member of group A or B) by choosing names that are consistent with the features of those in type A or B. In doing so, individuals can learn about the parent's type through their choice of child's name. Finally, the identity model attempt to address the presumed shortcoming of the signaling model: namely, that sending signals is generally assumed to be costly. However, if ethnic names are not costly, perhaps the economic model that

---

[2]The authors do vary the applicants level of expertise.

[3]Bertrand and Mullainathan also found little evidence that names as a social background marker have an effect on the probability that a resume leads to a callback from a potential employer.

2

needs to be adopted should allow for a positive benefit to ethnic names. Such an approach is used by Akerloff and Kranton (2000). They provide a model and set of examples whereby an individual takes on actions that may even appear to be detrimental, in order to improve their sense of self as a member of an identified group. Applied to name choice situation, their model is closer to the framework of Lieberson and coauthors, since the choice of the name is driven by a direct argument of utility function (identity) and not by a indirect effect through the outcomes of social or market interactions like in the models of Fryer and Levitt.

In addition to their contribution to understanding the economics of naming one's children they also provide an intriguing analysis of whether the blackness of a respondent's name (that is, the conditional probability that a respondent's name is held by someone that is black) affects a respondent's fertility outcomes. Using birth records from the State of California, Fryer and Levitt (2003) primarily provide an analysis of the effect of the blackness of a respondent's name on unwelcome birth outcomes – low birth weight babies, unmarried parents, lack of private health insurance, and total children born. In general, they find that the blackness of the mother's name is related to unwelcome birth outcomes.[4] Using the broad time series of their data, however, Fryer and Levitt can match a mother's information to her own birth, and in doing so they demonstrate that the blackness of a mother's names is associated with her own parent's outcomes (i.e. was she born out of wedlock, was her own mother a teenager, ...). In particular they show that the blackness of a woman's name is related to the poor birth outcomes of her parents. They conclude from these findings that since there is evidence that the blackness of a name reflects in part the individual's background that they grew up in, it is likely to be related to factors that could realistically affect their labor productivity. As such, they reconcile their work with Bertrand and Mullainathan by noting that while a black name may get fewer interviews, interviewers are likely to be inferring (and rightfully so) something about the individuals background and perhaps labor productivity.

Our study differs from these studies in the following three important ways. First, rather than focus on the issue of African American names, we explore broader linguistic features of names that

---

[4]Fryer and Levitt (2003) do provide a results where the birth mother's and father's educational attainments at the time of birth, and where zip-code specific income data are the dependent variables. However, the vast majority of their findings are focused on birth outcomes.

may convey meaning and/or may bring out discrimination. For instance, women with diminutive sounding names that end in an 'ee' sound, may also face glass walls and ceilings in their job market prospects.[5] Second, unlike the aforementioned studies that are limited in the literal outcomes they can consider (i.e. interviews and early births), our use of the General Social Survey (GSS) allows us to consider a broader set of economic outcomes. More specifically, we investigate the extent to which a respondent's first name features affect his or her years of formal education, self reported financial relative position as well as social class, to have a child before 25, and occupational prestige. Third, because we have data for both men and women, we can examine the gender differences between lifetime outcomes and first name features. Together, our use of broader linguistic features of names coupled with a more comprehensive set of lifetime outcomes for both male and female respondents make our study an important contribution to the understanding of the role that first names play in our economic and social lives.

Our analysis produces three important empirical findings. First, there is a strong empirical relationship between an individual's first name and their background. Second, there is a weaker empirical relationship between an individual's lifetime outcomes and their first names. Taken together, these first two findings are consistent with those of Fryer and Levitt (2003) – names may convey information about an individual's labor market productivity.

However, our third empirical finding is that both non-black non-whites with 'blacker' names as well as blacks with more popular (i.e. predominantly 'whiter') names have significantly worse financial outcomes. This last piece of evidence can be interpreted in light of a subtle form of discrimination: namely, while black names come with discrimination and identity costs and benefits for black individuals, non-black non-whites with 'blacker' names face the costs of such names though not the benefits. A similar identity/discrimination channel would also hold for blacks with more popular (i.e. whiter) names. This final point is consistent with Bertrand and Mullainathan's (2003) finding of a link between first names and discrimination, though it does not provide conclusive proof of discrimination.

---

[5]Such a study may not in general be possible using other data sets. For example, the data in Lieberson and Mikelson (1995) and Fryer and Levitt (2001) only have a fixed number of the first letters of the name, so that many endings are not observable.

2 Empirical Analysis

Our empirical analysis section is broken into three main parts. Sub-section 2.1 provides a description of our data as well as the features we extract from an individual's reported first name. In sub-sections 2.2 we provide the simple empirical regularities about first name features. Finally, in sub-section 2.3 we provide an empirical analysis of how first name features affect a respondent's lifetime outcomes.

2.1 The Data

In this section, we begin by describing the data employed in the paper. The GSS is an annual survey which asks respondents specific questions about their financial status, standard demographic variables (e.g., age, education, etc...), family information and background, as well as many more areas. The respondents are not re-interviewed across years so that that data are repeated cross sections rather than a panel. In two years, 1994 and 2002, the respondents were asked their first names which was coded by the interviewers.[6] There are a total number of just under six thousand respondents who reported names in the 1994 and 2002 surveys. Summary statistics for the key variables are reported in the Table 1, and a description of the variables we examine is presented in the Data Appendix.

The key variable of interest for this study is the extraction of important first name features from a respondent's name. The names data are recorded by interviewers and only clear keyboard errors were fixed, such as if the number '1' was struck instead of an 'l' or if the word 'respondent' was recorded as the first name. The first name features we create are as follows. Following Fryer and Levitt (2003), the Blackness index, BIND, is the probability of being black conditional on having the name in our sample. This was calculated only for names that have at least two occurrences in the sample.[7] Again, as shown in their study, this variable is associated with poor outcomes for the

---

[6]This data was made available to us based on a joint understanding between the authors and the National Opinion Research Center that the data would remain strictly confidential.

[7]So as not to introduce correlation between our measured BIND index and the error term, the BIND is calculated for each individual's name omitting their own name observation in the calculation. To note, the results presented

respondent's birth outcomes as well as the one into which she was born.

In addition, we also allow for a rich set of more linguistically based features of first names. First, the popularity, POP, of a name was calculated as one minus the respondent's name rank in gender specific Social Security Administrations' 'Top 1,000 popularity list' for the birth decade of the individual divided by 1000.[8] The series runs in ascending from 0 to 1, and has the value 0 if the name does not appear on the list, and 1 if it has the top name, namely $1 - ((rank - 1)/1000)$. To account for unusual spellings of the name, UNUSUAL-SP, is a dummy variable equal to one if the name is spelled unusually, and zero otherwise. A name was determined to be unusually spelled if multiple spellings were pronounced the same way (e.g. Jeffrey and Geoffrey), the more popular name appeared at least four times, and the less popular spelling occurred half or fewer times as the most popular spelling. $SYLLABLES$ is the number of syllables in a name, coded by following the rules of syllabication.[9] Finally, NICKNAME is a dummy variable if the name was listed as a nickname by Wallace (1992).

We also include a number of additional potentially important linguistic features of names that have currently not been explored and yet may convey information to employers and/or reflect aspects of an individual's upbringing. The following dummy variables are defined to be equal to one if their criteria is met and zero otherwise: AH-END if the name ends with an 'ah' sound (e.g. Laura), 0H-END if the name ends with an 'oh' sound (e.g. Mario), EE-END if the name ends with an 'ee' sound (e.g. Brittany), VOW-BEG if the name begins with a vowel (e.g. Abigail), and CON-END if the name ends with a consonant (e.g. Meredith). Finally, we adopt SCRABBLE as a measure of name's linguistic complexity. It is measured by the simple word score the name would receive from the Scrabble, the popular board game. Taken together, these linguistic feature

---

below are very similar if this adjustment is not made. For unique names, the BIND is set equal to the probability that the respondent is black, conditional on having a unique name. This treatment differs from that in Fryer and Levitt (2003), though the findings below do not depend on this data construction decision. Also note that given the few number of Asians in the sample and that Hispanics are coded as white in the GSS, the black names index, BIND, is very strongly negatively correlated with a similarly constructed white names index, and is also negatively correlated with the Popularity of a name discussed below. See Fryer and Levitt (2003) for a discussion of other names indices.

[8]The Social Security Administration lists the top 1,000 names by decade and by gender on their website, http://www.ssa.gov/OACT/babynames/.

[9]For an example of the rules of syllabication see http://english.glendale.cc.ca.us/syllables.html. We also coded the number of letters in a name but do not include it in our analysis as it is very highly correlated (approximately 0.9) with the number of syllables.

provide a broader scope for discerning potential information that may be projected from and/or onto a name.

Next, we now turn to a description of the six lifetime outcomes that we examine in this study. Our measures of lifetime outcomes are a wide-ranging scope of economic and social standing outcomes that have clear economic content and implications and which may be affected by a respondent's first name. First, FINRELA is the respondent's household financial income relative to others. It is measured on a self reported increasing scale of 1 to 5.[10] Second, OCCPREST is the respondent's occupational prestige score based on 1980 occupational classifications and income, and it is on an increasing scale from 17 to 86. Third, we define CLASS to be the respondent's social class relative to others. It is measured on a self reported scale of 1 to 4, with 1 being lower class, 2 being working class, 3 being middle class and 4 being upper class. Fourth, since Fryer and Levitt point to the importance of birth outcomes as they related to their black names index, we define CHB25 to be a dummy variable equal to 1 if an over-25 year old respondent reports having a child before 25, and zero otherwise. This variable is defined for both male and female respondents. Our final two measures of lifetime outcomes are happiness, measured on a an increasing scale of 1 ('not too happy'), 2 ('pretty happy') or 3 ('very happy'), and EDUC which is the number of years of education that the respondent has obtained.[11,12]

Additional explanatory variables which we use to control for observable factors that can systematically influence a respondent's lifetime outcomes are the respondent's age, AGE, labor market experience, $LMEX$ which is equal to $AGE - EDUC - 5$ and sex, MALE.[13] We also have additional information of a respondent's background. For instance, we code the dummy variables BLACK if the respondent lists this as their race and OTHER is for those who report a race other

---

[10]FINRELA is measured on a self reported scale of 1 to 5, with 1 being that your financial status is well below the mean, 2 being that it is below the mean, 3 being that it is about the mean, 4 being that it is above the mean, and 5 being that it is far above the mean. While the GSS does report family income for some years, it does not do so for all years.

[11]The actual question is: "Taken all together, how would you say things are these days would you say that you are:" very happy (3), pretty happy(2), or not too happy (1)? Note that the GSS codes happiness so that an increase in the score makes one unhappy. To ease the presentation, we changed the order of the scale.

[12]We also experimented with measuring education by the highest degree obtained. They are not included as the results were broadly similar.

[13]We also experimented with placing an upper bound on $LMEX$ to account for retirement. Such a change does not affect the results presented below.

than white or black.[14] In addition, MAEDUC and PAEDUC are the number of years of the respondent's parent's formal education. Also, if either of a respondent's parents were born outside of the U.S., then ABROAD is set equal to one, and is zero otherwise.[15] Again, the details of the data are provided in the Data Appendix as well as in Table 1.

2.2 Empirical Regularities

Table 1 provides a number of basic statistics for the main data of interest. In particular, we report the mean, standard deviation, median, minimum, maximum and number of observations for each of the 11 first name features, $FNF = \{$POPULATION, SYLLABLES, SCRABBLE, UNUSUAL-SP, NICKNAME, AH-END, 0H-END, EE-END, VOW-BEG, CON-END, BIND$\}$, our six measures of lifetime outcomes, $LO = \{$FINRELA, OCCPREST, CLASS, HAPPY, CHBF25, and EDUC $\}$, and our nine exogenous background factors, $EBF = \{$BLACK, OTHERRACE, MALE, YRBORN, PAEDUC, MAEDUC, SIBS, ABROAD$\}$.[16] Both BLACK, OTHERRACE and MALE are of direct interest as both gender and race are often important decipherable information from a name. Approximately 14 percent of the sample are black, 44 percent are male and the average age is in their mid 40's.

---

[14]Unfortunately, in the GSS not all demographic questions are asked for both the respondent and his/her spouse.

[15]The results below are robust to separately including dummy variables for whether the respondent was born abroad or just their parents were born abroad.

[16]We also included a respondent's religion, region, and city size at age 16 as additional variables in the final column of Table 3 and all of Table 5. The inclusion or exclusion of these variables does not affect our reported results.

Table 1: Summary Statistics

FIRST NAME FEATURES (FNF)

VARIABLE	$MEAN$	STD	$MEDIAN$	MIN	MAX	$NOBS$
POPULARITY	0.61	0.37	0.76	0	1	5587
SYLLABLES	2.01	0.74	2.00	0	7	5587
SCRABBLE	9.95	3.78	10.00	1	29	5587
UNUSUAL-SP	0.06	0.24	0.00	0	1	5587
NICKNAME	0.15	0.36	0.00	0	1	5587
AH-END	0.18	0.39	0.00	0	1	5587
OH-END	0.01	0.11	0.00	0	1	5587
EE-END	0.20	0.40	0.00	0	1	5587
VOW-BEG	0.11	0.32	0.00	0	1	5587
CON-END	0.48	0.50	0.00	0	1	5587
BIND	0.14	0.16	0.09	0	1	5587

EXOGENOUS BACKGROUND FACTORS (EBF)

VARIABLE	$MEAN$	STD	$MEDIAN$	MIN	MAX	$NOBS$
BLACK	0.14	0.34	0.00	0	1	5587
OTHERRACE	0.05	0.22	0.00	0	1	5587
MALE	0.44	0.50	0.00	0	1	5587
YRBORN	51.60	17.62	54.00	5	84	5571
PAEDUC	11.18	4.12	12.00	0	20	4065
MAEDUC	11.25	3.48	12.00	0	20	4806
SIBS	3.67	3.06	3.00	0	35	5568
ABROAD	0.17	0.38	0.00	0	1	5587
AGE	46.13	17.21	43.00	18	89	5570

LIFETIME OUTCOMES (LO)

VARIABLE	$MEAN$	STD	$MEDIAN$	MIN	MAX	$NOBS$
FINRELA	2.89	0.86	3.00	1	5	4227
OCCPREST	43.50	13.66	42.00	17	86	5334
CLASS	2.47	0.65	2.00	1	4	5552
HAPPY	2.17	0.62	2.00	1	3	4261
CHB25	0.35	0.48	0.00	0	1	1704
EDUC	13.25	2.97	13.00	0	20	5570

Note: The sample statistics average(MEAN), standard deviation (STD), median (MEDIAN), minimum (MIN) and maximum (MAX). NOBS is the number of observations. The data were taken from the 1994 and 2002 samples from the General Social Survey (GSS).

The name variables are somewhat unusual and so deserve our initial attention. Table 1 reveals a number of interesting features about the data. First, the average popularity of names in our sample is .61 with a median of .75, suggesting a lot of name clustering at the top of the popularity index. Second, the linguistic features of names suggest that, on average, a name has 2 syllables, a scrabble score of 10, with about 6 percent having a non-standard or unusual spelling. Moreover, about half of all names end in a constant and almost 80 percent start with a constant (1 minus the fraction that begin in a vowel). And of the names that end in a vowel sound, however, they appear to be evenly split between 'ee' endings and 'ah' endings (each about 20 percent of the sample), with 'oh' sounding name endings being relatively rare (about 1 percent of the time). The average probability of having black name is estimated around 14 percent, suggesting that our method of estimating that reproduces the sample probability of being black very accurately. Finally, the median value for BIND is much lower, however, at about .09, consistent with skewed distribution of BIND, where some names have very high values of BIND.

The remaining demographic data are somewhat standard. The respondent's of parents have an average level of 11 years of education, while respondents have approximately two more years of formal education. Moreover, 17 percent of the sample was either born outside of the U.S. or their parents were, and 35 percent of those over 25 had a child before they were 25. Also, on average respondent's report that they are between working class and middle class, and that they are 'somewhat happy'. Finally, the average respondent's marker of their relative income status when growing up is very similar to their average perceived financial relative position, 2.8 out of a scale of 1 to 5, with a standard deviation of .6. Finally, the average prestige occupation score is 43, with a range between 17 and 86.

Table 2 provides the raw correlation between these key first name features variables of interest. Rather than remark on all the possible sets of correlations, we provide information of the data relationships as they pertain to the relationships between First Name Features in addition Exogenous Background Factors and Lifetime Outcomes. First, there is a significant amount of correlation between the First Name Features. For instance, names that have a higher popularity score have typically higher scrabble scores, are less likely to end in with an 'ah' or 'oh' ending

sound, and are less likely to be nicknames or start with a vowel. Additionally, first names with a higher score for the BIND is also significantly correlated with ending with an 'ah' sound, not ending with a consonant and having more syllables.[17]

Second, First Name Features and Exogenous Background Factors are frequently significantly correlated with one another. Indeed, most of the correlations reported in the rows labeled BLACK through ABROAD are statistically significant. Interestingly, POP is positively correlated with the mother and father's education levels. It is also negatively related to having more siblings and not having both parents born in the U.S. The blackness index is also lower among men and it is negatively related to parent's education.[18] Also, as has been noted by others, First Name Features contain a great deal of gender and race information. Indeed, the correlation of First Name Features with BLACK, MALE and OTHER are typically statistically significant.

Finally, First Name Features are not as significantly correlated with Lifetime Outcomes as they were with Exogenous Background Factors– that is, the correlations in the rows labelled FINRELA through EDUC are less frequently significant as those above. A notable exceptions is that BIND is negatively correlated with FINRELA, OCCPREST, CLASS, HAPPY and EDUC, and positively correlated with having a child before 25. Also, POPULARITY is positively correlated with attaining more education and having a reduced likelihood of having a child before 25.

---

[17]First name features are also correlated across time within a family. Though not shown, the GSS data contained also first name information on the other household members. We used this information to form parent child pairs to study intergenerational transmission of the first name features considered here using intergenerational correlations or contingency tables as indicators of intergenerational persistence. The analysis was performed for the whole sample, for African American sub sample and for matched gender sub sample (where parent and child were of same gender). The overall result of this analysis is that most first name features are statistically significantly correlated across generations, but that this correlation is a moderate one. The only exception to this moderate correlation is the correlation of the Blackness index across generations in the entire sample (and matched gender sub sample), where the intergenerational correlation is approximately .20. However, the intergenerational correlation in the African American sub sample is only .08, implying that the high correlation in the overall sample is driven by the presence of the African American minority in that sample.

[18]Fryer and Levitt (2003) also find a negative association between BIND and MALE.

Table 2: Correlations of First Name Features with the Data Used in the Analysis

	POPULAR-ITY	SYLL-ABLES	SCRABBLE	UNUSUAL-SP	NICK-NAMES	AH-END	OH-END	EE-END	VOW-BEG	CON-END	BIND
SYLLABLES	.03										
SCRABBLE	.16c	.29c									
UNUSUAL-SP	−.17c	−.02	.00								
NICKNAMES	−.16c	−.33c	−.04c	−.01							
AH-END	−.05c	.37c	−.16c	−.05c	−.18c						
0H-END	−.07c	.12c	−.01	−.02	−.04c	−.05c					
EE-END	.02	.06c	.09c	.08c	.24c	−.23c					
VOW-BEG	−.07c	.16c	−.17c	.01	−.07c	.01	−.06c				
CON-END	.13c	−.27c	.04c	−.07c	−.02	−.42c	−.11c	−.05c			
BIND	−.18c	.17c	.00	−.08c	.02	.14c	.05c	−.48c	.00		
BLACK	−.10c	.11c	.03	−.01	−.05c	.08c	.00	.03	.07c	−.18c	
OTHERRACE	−.09c	.08c	−.02	−.03	−.03	.06c	.15c	−.03	.04c	−.09c	.17c
MALE	.21c	−.32c	.03	−.06c	.16c	−.38c	.10c	−.17c	−.05c	.47c	.05c
YRBORN	.14c	−.01	.06c	.04c	.07c	.07c	.01	.02	−.09c	−.02	−.09c
PAEDUC	.10c	−.07c	.02	.03	.05c	−.03	−.08c	.00	−.07c	.08c	−.09c
MAEDUC	.13c	−.08c	.04	.05c	.05c	−.04c	−.10c	.01	−.10c	.08c	−.08c
SIBS	−.07c	.05c	−.02	−.04c	−.02	.02	.04c	.02	.05c	−.04c	.08c
ABROAD	−.08c	.11c	−.03	.01	−.06c	.04c	.15c	−.04c	.10c	−.03	.04c
AGE	−.14c	.01	−.06c	−.04c	−.07c	−.07c	.00	−.02	.09c	.02	.03
FINRELA	.04	−.05c	−.01	−.02	.04	−.07c	.00	−.01	−.03	.08c	−.06c
OCCPREST	.03	−.04c	.01	.01	−.01	−.05c	−.03	−.03	−.01	.06c	−.05c
CLASS	.01	−.03	.00	.00	.01	−.06c	−.01	.01	−.01	.03	−.07c
HAPPY	.02	−.03	.03	−.01	.02	−.05c	.02	.01	−.03	.02	−.06c
CHBF25	−.10c	.08c	−.05	.02	−.06	.07c	.01	.06	.01	−.11c	.11c
EDUC	.06c	−.05c	.01	.01	.01	−.03	−.03	−.03	−.05c	.06c	−.09c

Note: See Table 1. ^{c} indicates that the relationship is different from zero at or below the .01 level of statistical significance, respectively.

2.3 Estimation Results

While Tables 1 and 2 in sub-section 2.2 describe the empirical regularities of the data, in this sub-section we adopt a more standard regression format for disentangling the potential role that names signal about background and which can affect an individual's future. We begin by analyzing the extent to which a respondent's Exogenous Background Factors (EBF) can be predicted from an individual's first name features. To do so, we estimate the following regression specification:

$$EBF_i = \beta_0 + \beta_1 \cdot FNF_i + \epsilon_i \tag{1}$$

where EBF_i is respondent $i's$ 9 exogenous background factors, and FNF_i is respondent $i's$ 11 first name features described above.[19] These regression results for expression (1) are presented in Table 3, and indicate the extent to which an outsider would be able to infer an individual's **exogenous** background characteristics just from the characteristics of an individual's name.[20,21]

---

[19]We also ran the regressions in this analysis by including quadratic terms for the relevant first name features as well as allowing for asymmetric effects by measuring these same variables as the absolute deviation from the median. In general, these effects were not significant and so were excluded from the results reported below.

[20]Of course, the dependent variables in many of these regressions are discrete. However, rather than estimate probit or ordered probit specifications, we provide estimates using the simple OLS estimator. To note, when one estimates these empirical models using probit or ordered probit estimators, the results of the linear effects are similar to those presented in the following tables. Moreover, the pattern of statistical significance is unaffected by using this alternative estimation strategy. Hence, to keep matters simple we present the OLS estimates below though the alternative estimates are available upon request.

[21]The standard errors were estimated allowing for clustering based on the individuals names that occur more than twice and by assigning all unique names to the same cluster. This is easily justified as follows. The regression of Y_i on dummy variables for actual first names (920 of them that occur twice in the full sample) , AFN_i, allows for an F-test on their joint significance. However this type of test from the regression $Y_i = AFN_i + u_i$ does not reveal much interesting information about the qualitative and common features of names other than provide a somewhat interesting rank of names. However if AFN_i is related to first name features, FNF_i with estimated error, e_i, then the regression of $Y_i = FNF_i + e_i + u_i$ would involve an error, e_i, that was distinct for each name. Accordingly, the variance of this error may differ across names, which suggests that we should allow for clustering along this dimension, and we do so throughout the paper.

Table 3: Respondent's Background Regressed on First Name Features Only

	BLACK	OTHERRACE	MALE	YRBORN	PAEDUC	MAEDUC	SIBS	ABROAD
POPULARITY	−0.096[c]	−0.052[b]	0.275[c]	7.794[c]	1.025[c]	1.192[c]	−0.514[c]	−0.073[b]
	[0.017]	[0.024]	[0.092]	[2.668]	[0.320]	[0.196]	[0.101]	[0.037]
SYLLABLES	0.029[c]	0.013[b]	−0.109[c]	−0.849	−0.234[a]	−0.234[b]	0.103	0.053[c]
	[0.009]	[0.006]	[0.033]	[0.755]	[0.131]	[0.103]	[0.067]	[0.012]
SCRABBLE	0.003[b]	−0.001	0.003	0.263[a]	0.015	0.021	−0.01	−0.004[a]
	[0.001]	[0.001]	[0.006]	[0.151]	[0.027]	[0.021]	[0.012]	[0.002]
AH-END	0.014	0.022[a]	−0.202[c]	6.566[c]	0.529[a]	0.283	0.019	−0.024
	[0.020]	[0.011]	[0.059]	[1.540]	[0.285]	[0.230]	[0.169]	[0.027]
OH-END	−0.092	0.269[c]	0.688[c]	5.944[b]	−1.617[a]	−2.118[c]	0.817[b]	0.413[c]
	[0.070]	[0.059]	[0.083]	[2.473]	[0.855]	[0.565]	[0.358]	[0.060]
EE-END	−0.011	0.000	−0.085	1.384	0.396	0.231	0.267	−0.042[a]
	[0.020]	[0.009]	[0.064]	[1.662]	[0.292]	[0.235]	[0.184]	[0.022]
VOW-BEG	0.025	0.014	−0.008	−3.021[b]	−0.524[b]	−0.757[c]	0.311[a]	0.069[c]
	[0.016]	[0.011]	[0.060]	[1.385]	[0.230]	[0.194]	[0.159]	[0.021]
CON-END	−0.049[c]	0.003	0.332[c]	1.299	0.677[c]	0.458[c]	0.016	−0.012
	[0.016]	[0.008]	[0.053]	[1.219]	[0.250]	[0.174]	[0.132]	[0.024]
UNUSUAL-SP	−0.045[c]	−0.031	−0.017	5.584[c]	0.856[c]	0.997[c]	−0.639[c]	0.003
	[0.016]	[0.027]	[0.045]	[1.774]	[0.305]	[0.171]	[0.162]	[0.021]
NICKNAME	−0.040[b]	−0.012	0.210[c]	5.291[c]	0.497[b]	0.436[b]	−0.203[a]	−0.026
	[0.017]	[0.011]	[0.062]	[1.409]	[0.234]	[0.178]	[0.115]	[0.021]
BIND		0.021	0.179[c]	1.200	−1.516[c]	−0.732[a]	1.246[c]	−0.016
		[0.022]	[0.067]	[2.296]	[0.513]	[0.379]	[0.343]	[0.035]
NOBS	5587	5587	5587	5571	4065	4806	5568	5587
R^2	0.03	0.04	0.36	0.1	0.03	0.05	0.02	0.05
P: FNF	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
P: FNF ex BIND	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: [c], [b], and [a] indicate that the relationship is different from zero at or below the .01, .05 and .10 level of statistical significance, respectively. All regressions include a constant and a dummy variable for the 1994 survey. P: FNF is the p-value from the F-test that all the coefficients on First Name Features are jointly zero. P: FNF ex BIND is the p-value from the F-test that all the coefficients on First Name Features, excluding BIND, are jointly zero. Standard errors, reported in brackets, were calculated based on clustering by actual first names that appear at least two times or more, with those observed once clustered together.

14

The results in Table 3 provide the very strong impression that one can infer a great deal about a respondent's background from their first name. As such, this confirms the results of the simple correlations presented in Table 2. For instance, respondents who are Black, as reported in the first column of results, have less popular names, more syllables and higher Scrabble scores, fewer 'oh' endings, are more likely to end in a vowel, and have higher blackness index. This same pattern also exists for those that are neither white nor black (i.e. either), except that they have more 'ah' endings and more 'oh' endings, not fewer ones. Also, as indicated in the bottom rows of the table, in both of these equations the R^2 is .04 and the p-value from an F-test that the coefficients on First Name Features are all equal to zero is well below .001. As well, even if BIND is excluded from the F-test, the remaining First Name Features are jointly significant at below the .001 level. The remaining rows indicate a number of key features of the relationship between background factors and first names. First, names indicate a great deal of information on gender and the year when one was born. In both of the regressions, the R^2's are relatively large (over .35 for Male).[22] Also, a respondent's higher parental education background can be partially inferred from higher popularity, fewer syllables, more standard spellings, more nicknames, fewer 'oh' endings, not starting with a vowel, ending with a consonant and having a lower Blackness Index. This latter result, on BIND, is actually quite large: moving from a purely non-black name to a fully black name is associated with a Father having 2 fewer years of formal education and a Mother having 1 year less.

While it has been important to establish that first names confer a great deal of information about a person's background, it is also essential to understand whether any linkages exist between first name features and an individual's lifetime outcomes. As discussed above, the lifetime outcomes we consider from the GSS data available are denoted LO_i. To begin to unravel this connection, we first estimate the following specification[23]

---

[22]The point we are making here, is that purely linguistic features (non-gender specific) of the names do convey information about the gender. Naturally, most names are gender-specific and thus convey gender information.

[23]Again, while these lifetime outcomes are discrete, we estimate the parameters using OLS with clustered errors for ease of presentation. The results are similar if we estimate the parameters using probit or ordered-probit.

Table 4A: Respondent's Lifetime Outcomes Regressed on First Name Features

	FINRELA	OCCPREST	CLASS	HAPPYIND	CHILDBF25	EDUC
POPULARITY	0.062[b]	0.686	−0.006	0.008	−0.107[c]	0.393[c]
	[0.031]	[0.567]	[0.021]	[0.020]	[0.034]	[0.121]
SYLLABLES	−0.001	−0.236	0.016	−0.020	0.033	−0.089
	[0.022]	[0.336]	[0.016]	[0.017]	[0.024]	[0.070]
SCRABBLE	−0.005	0.022	−0.003	0.004[a]	−0.008[a]	−0.001
	[0.003]	[0.050]	[0.003]	[0.002]	[0.004]	[0.013]
AH-END	−0.071	−1.476[b]	−0.115[c]	−0.027	0.010	−0.017
	[0.048]	[0.640]	[0.035]	[0.030]	[0.060]	[0.140]
OH-END	0.101	−3.100[c]	−0.094	0.179[b]	−0.025	−0.423
	[0.111]	[1.143]	[0.088]	[0.081]	[0.113]	[0.423]
EE-END	0.023	−1.177[a]	−0.019	0.027	0.062	−0.119
	[0.042]	[0.627]	[0.033]	[0.031]	[0.050]	[0.141]
VOW-BEG	−0.065[a]	−0.277	−0.020	−0.036	−0.025	−0.328[b]
	[0.038]	[0.751]	[0.030]	[0.026]	[0.036]	[0.146]
CON-END	0.111[c]	0.362	−0.011	0.013	−0.039	0.188
	[0.040]	[0.504]	[0.024]	[0.025]	[0.048]	[0.122]
UNUSUAL-SP	−0.029	0.675	−0.014	−0.018	−0.005	0.301[a]
	[0.065]	[0.748]	[0.044]	[0.030]	[0.039]	[0.167]
NICKNAME	0.072[a]	−0.611	0.004	0.008	−0.084[b]	0.036
	[0.040]	[0.626]	[0.029]	[0.028]	[0.035]	[0.119]
BIND	−0.218[c]	−3.487[c]	−0.258[c]	−0.195[c]	0.251[c]	−1.148[c]
	[0.078]	[1.168]	[0.059]	[0.060]	[0.085]	[0.259]
NOBS	4227	5334	5552	4261	1704	5570
R^2	0.01	0.01	0.01	0.01	0.04	0.02
P: FNF	0.000	0.000	0.000	0.000	0.000	0.000
P: FNF ex BIND	0.000	0.000	0.131	0.024	0.000	0.000

Note: See Table 3.

$$LO_i = \beta_0 + \beta_1 \cdot FNF_i + \epsilon_i \tag{2}$$

where, again, FNF_i represents the respondent's first name features described above.

Our estimation results of expression (2) are reported in Table 4A. There are a number of key findings to the table. First, in all cases, the first name variables are jointly statistically significant at conventional levels. Second, the R^2's are typically lower in these regressions as compared to those in Table 3 where exogenous background characteristics were the dependent variables. Third, more popular names are associated with better lifetime outcomes: that is, more education, occupational prestige and income, and a reduced likelihood of having a child before 25. Moreover, names with higher values for BIND are associated with poorer lifetime outcomes: that is, less education, occupational prestige, happiness, social class and income, and an increased likelihood of having a child before 25. As well, 'ah' and 'oh' ending sounds in a name are also related to poorer lifetime outcomes, though popularity is related to better lifetime outcomes.

As demonstrated in Table 4A, first names just by themselves convey information about a respondent's lifetime outcomes. An important question that has come up in the literature, however, is whether the link between first names and lifetime outcomes survives the inclusion of additional information such as labor market experience and/or education (e.g. the analysis of Bertrand & Mullainathan). In other words, does your first name features affect one's lifetime outcomes after learning information that may be readily available in a job resume.[24] We estimate, therefore, the following equation and report the results in Table 4B.

$$LO_i = \beta_0 + \beta_1 \cdot FNF_i + \gamma_1 LMEX_i + \gamma_2 LMEX_i^2 + \gamma_3 EDUC_i + \epsilon_i \tag{3}$$

---

[24]Since employers are not suppose to discriminate based on age, and this information (like race) is unlikely to be directly available on a resume, we use labor market experience instead. While it is also illegal to discriminate based on gender, we include this since it is likely to be easily inferred from a resume. However, the results in Table 4B do not depend on the inclusion of MALE as an explanatory variable.

Table 4B: Respondent's Lifetime Outcomes Regressed on First Name Features and Resume Information

	$FINRELA$	$OCCPREST$	$CLASS$	$HAPPY$	$CHBF25$
POPULARITY	0.026	0.760[a]	0.031	0.002	−0.045[a]
	[0.029]	[0.392]	[0.023]	[0.021]	[0.026]
SYLLABLES	0.016	−0.212	0.014	−0.018	0.012
	[0.020]	[0.270]	[0.014]	[0.017]	[0.019]
SCRABBLE	−0.005[a]	0.044	−0.001	0.005[a]	−0.005
	[0.003]	[0.042]	[0.002]	[0.002]	[0.003]
AH-END	−0.016	−0.892	−0.082[b]	−0.012	−0.011
	[0.044]	[0.562]	[0.036]	[0.030]	[0.052]
OH-END	0.083	−0.745	0.013	0.193[b]	−0.026
	[0.101]	[1.419]	[0.074]	[0.088]	[0.114]
EE-END	0.05	−0.794	−0.003	0.034	0.040
	[0.039]	[0.560]	[0.028]	[0.030]	[0.042]
VOW-BEG	−0.055	0.067	−0.026	−0.034	−0.036
	[0.037]	[0.605]	[0.026]	[0.026]	[0.036]
CON-END	0.047	0.133	−0.006	−0.003	0.040
	[0.046]	[0.424]	[0.024]	[0.026]	[0.039]
UNUSUAL-SP	−0.033	0.652	−0.003	−0.02	0.022
	[0.057]	[0.513]	[0.038]	[0.029]	[0.037]
NICKNAME	0.064[a]	−0.039	0.044[a]	0.008	−0.060[a]
	[0.038]	[0.481]	[0.025]	[0.029]	[0.031]
BIND	−0.110	−0.727	−0.170[c]	−0.163[c]	0.208[b]
	[0.075]	[0.994]	[0.057]	[0.061]	[0.083]
MALE	0.155[c]	−0.412	−0.030	0.037	−0.181[c]
	[0.033]	[0.372]	[0.021]	[0.023]	[0.027]
LMEX	0.007[c]	0.162[c]	0.000	0.003	0.019[c]
	[0.002]	[0.033]	[0.002]	[0.002]	[0.003]
$LMEX^2/100$	−0.002	−0.091[a]	0.011[c]	−0.001	−0.029[c]
	[0.003]	[0.049]	[0.002]	[0.003]	[0.004]
EDUC	0.097[c]	2.558[c]	0.078[c]	0.025[c]	−0.047[c]
	[0.005]	[0.070]	[0.004]	[0.004]	[0.004]
NOBS	4211	5312	5530	4242	1701
R^2	0.12	0.28	0.12	0.02	0.16
P: FNF	0.072	0.027	0.002	0.02	0.018
P: FNF ex BIND	0.204	0.045	0.083	0.199	0.212
R^2†	0.13	0.28	0.13	0.03	0.16
$P:FNF$†	0.503	0.158	0.208	0.147	0.188
$P:BIND$†	0.522	0.817	0.024	0.074	0.034
$P:POP$†	0.912	0.152	0.721	0.298	0.103

Note: See Table 3. The † indicates that the R-squared's and p-values refer to a regression (not shown) where BLACK and OTHERRACE were included as right hand side regressors.

18

There are three key findings in Table 4B. First, POPULARITY remains a significant explanatory variable in the CHB25 regressions, and it is no longer significant in the FINRELA response though it is now positively related to the OCCPREST response. Based on the statistical significance of education and labor market experience in these equations, it can be inferred that popularity is working through education in the financial relative income response. Second, BIND is statistically significant again in the CLASS, HAPPY and EDUC regressions. Higher values for BIND lead to a lower assessment of social class, happiness and an increased chance of having children before 25. However, as with POPULARITY, BIND is now no longer statistically significant in the income responses. Again, the role of education and labor market experience is clearly soaking up the role that POPULARITY and BIND played in the income response regressions in Table 4A. Third, the First Name Features are jointly statistically different from zero at or below the .1 level for all the lifetime outcomes. It would thus seem that first names retain a strong role overall in determining lifetime outcomes even after controlling for a respondent's labor market experience.

In addition, the bottom panel of Table 4B includes additional information about the sensitivity of these findings to whether race variables BLACK and OTHERRACE are includes as additional explanatory variables to equation (3). These variables might be important to an interview, where a candidate's race may be easily communicated by sight.[25] The bottom panel reports the R^2's and p-value tests for the inclusion of the FNF variables as well as POPULARITY and BIND by themselves for these specifications (not shown). Interestingly, including these race variables only marginally affect the R^2's by less than .01. Second, and perhaps more importantly, BIND is statistically significant at below the .1 level for both the CLASS and CHB25 regressions, though collectively the FNF variables are no longer statistically significant. Finally, POPULARITY is not significant when race variables are included, with the exception of the CHB25 regression.

Of course, education and labor market experience (with the exception of the age component) are endogenous and are well known to affect the lifetime outcomes we consider. To explore the extent to which only purely exogenous components of an individuals life, that is their first name features (FNF) and exogenous background factors (EBF), we estimate the following regression:

---

[25]Of course, other characteristics may be communicated by sight that may affect a candidate's chances for receiving an offer: namely, obesity or height.

$$LO_i = \beta_0 + \beta_1 \cdot FNF_i + \beta_2 EBF_i + \epsilon_i \qquad (4)$$

The estimation results of equation (4) are provided in Table 5.[26] The importance of these regressions is to identify the pure effects of name features on an individual's lifetime outcomes, since presumably their names could be indirectly affecting the results in Table 4B as first name features could be affecting education and labor market experience. The table has two main components: the first 6 columns of the table allow for the standard effects of a first name's popularity and the black names index, whereas the last 6 columns allow for interactive effects between these first name features and the respondent's actual race.

There are three key findings contained in the first set of results in columns 1 through 6. First, POPULARITY is still a significant predictor of having a higher social class and not having a child before 25. Second, BIND is still a significant predictor of lower social class, lower happiness, reduced education and an increased likelihood of having a child before 25. In words, moving from a name that no blacks have to one that only blacks have is associated with .6 years fewer of education. Importantly, these findings hold even though RACE and the full gamut of external background factors, particularly parents education, are included as explanatory variables. Third, First Name Features excluding BIND are jointly significant at below the .05 level in the social class and happiness regressions.

To put these findings in a broader context, we can think of our six lifetime outcome variables as reflecting three types of phenomena: first, FINRELA and PRESTIGE are indicators of actual and future financial outcomes; Second, CHB25 and EDUC are lifetime outcomes that are intermediate economic outcomes in the sense that while they would not enter directly into an individual's utility function, they would definitely impact on an individual's utility. Moreover, these two variables are variables that are likely to be correlated with an individual's labor market productivity. Third, social class and happiness are quasi-economic indicators of both socio-economic standing as well

---

[26]We also included a respondent's religion, region, and city size at age 16 as additional variables in all the columns of Table 5. The inclusion or exclusion of these variables does not affect our reported results.

as reflect a level of lifetime enjoyment that is related to economic utility.[27] There is the further possibility that these quasi economic outcomes may also be correlated with labor productivity: simply unhappy workers and those that feel that they are lower class (or even upper class) may have differential labor productivities. Based on the findings in columns 1 through 6 of Table 5, controlling for a myriad of exogenous family background characteristics, a first name's popularity and/or 'blackness' appear to have an impact on intermediate economic outcomes that are likely correlated with labor productivity but not on actual economic outcomes. It would thus appear that our finding support Fryer and Levitt's (2003) reconciliation of their findings with those of Bertrand and Mullainathan's (2003) findings of apparent resume discrimination. Namely, that the 'blackness' of a name is correlated with factors that can affect labor productivity which could in turn be reflected in discrimination at the resume level. As we demonstrate, however, this potential channel of discrimination does not have an impact on pure economic outcomes in our sample.

The remaining part of Table 5, however, further investigate these findings. Namely, we re-estimated specification (4) and allow for interactive effects between BLACK and OTHERRACE with POPULARITY and BIND.[28] In particular, we are interested in whether the impact of a popular or black sounding name on an individual's lifetime outcomes differs depending upon whether one is black or not. In particular, one would be more inclined to conclude that there was evidence in favor of discrimination based on the blackness of a name if the effect of the variable on income were stronger for black respondents. In contrast, one would be more inclined to believe that identity were an important component of an individual's name if, for example, non-black individuals had worse lifetime outcomes if their names registered high on the blackness index. In other words, blacks may face outside costs and inside benefits to having a 'blacker name', though a non-black would only realize the former.

---

[27]For a more comprehensive treatment of the happiness literature, see, among others, Blanchflower and Oswald (2004) and Frey and Stutzer (2002).

[28]In Table 6 below we also examine the effects of gender.

Table 5: Respondent's Lifetime Outcomes Regressed on First Name Features and All Background Information

	(1) FIN-RELA	(2) OCC-PREST	(3) CLASS	(4) HAPPY	(5) CHBF25	(6) EDUC	(7) FIN-RELA	(8) OCC-PREST	(9) CLASS	(10) HAPPY	(11) CHBF25	(12) EDUC
POPULARITY	−0.039 [0.040]	0.127 [0.569]	0.032 [0.024]	−0.026 [0.032]	−0.045 [0.034]	−0.05 [0.100]	−0.009 [0.045]	0.667 [0.731]	0.023 [0.026]	−0.035 [0.035]	−0.048 [0.039]	0.02 [0.113]
POP × BLA							−0.219ᵃ [0.116]	−3.861ᵇ [1.836]	0.082 [0.076]	0.028 [0.092]	−0.045 [0.103]	−0.116 [0.316]
POP × OTH							−0.037 [0.174]	−1.251 [2.634]	0.010 [0.080]	0.067 [0.112]	0.040 [0.165]	−0.816ᵇ [0.340]
SYLLABLES	0.039 [0.027]	−0.085 [0.417]	−0.001 [0.018]	−0.034 [0.023]	0.024 [0.035]	0.059 [0.084]	0.038 [0.027]	−0.105 [0.417]	−0.001 [0.018]	−0.036 [0.022]	0.024 [0.036]	0.058 [0.085]
SCRABBLE	−0.006 [0.004]	0.106ᵃ [0.057]	0.000 [0.003]	0.007ᵇ [0.003]	−0.009ᵇ [0.004]	−0.003 [0.012]	−0.006 [0.004]	0.107ᵃ [0.057]	0.000 [0.003]	0.008ᵇ [0.003]	−0.008ᵃ [0.004]	−0.003 [0.012]
AH-END	−0.045 [0.050]	−0.542 [0.747]	−0.092ᵇ [0.041]	0.042 [0.041]	−0.031 [0.076]	−0.13 [0.141]	−0.045 [0.050]	−0.517 [0.749]	−0.095ᵇ [0.041]	0.045 [0.041]	−0.024 [0.078]	−0.128 [0.141]
OH-END	0.057 [0.153]	0.436 [1.820]	0.179ᵇ [0.078]	0.260ᶜ [0.092]	−0.056 [0.113]	−0.375 [0.549]	0.084 [0.157]	0.51 [1.812]	0.179ᵇ [0.078]	0.269ᶜ [0.095]	−0.049 [0.113]	−0.389 [0.546]
EE-END	0.013 [0.051]	−0.797 [0.697]	0.002 [0.033]	0.035 [0.038]	0.008 [0.063]	−0.121 [0.136]	0.012 [0.051]	−0.784 [0.695]	0.000 [0.033]	0.036 [0.038]	0.008 [0.064]	−0.122 [0.137]
VOW-BEG	−0.062 [0.050]	0.622 [0.738]	−0.018 [0.035]	−0.017 [0.034]	−0.024 [0.044]	−0.065 [0.137]	−0.064 [0.050]	0.595 [0.741]	−0.017 [0.035]	−0.016 [0.034]	−0.023 [0.045]	−0.068 [0.137]
CON-END	0.002 [0.049]	0.291 [0.555]	−0.023 [0.027]	−0.022 [0.031]	0.024 [0.050]	0.011 [0.120]	0.003 [0.049]	0.3 [0.553]	−0.023 [0.027]	−0.023 [0.031]	0.022 [0.051]	0.008 [0.120]
UNUSUAL-SP	−0.119ᵃ [0.071]	0.07 [0.762]	−0.027 [0.039]	−0.054 [0.041]	0.009 [0.047]	−0.014 [0.144]	−0.119ᵃ [0.071]	0.138 [0.761]	−0.029 [0.039]	−0.058 [0.041]	0.008 [0.046]	0.001 [0.149]
NICKNAME	0.051 [0.050]	−0.664 [0.645]	0.037 [0.026]	−0.01 [0.037]	−0.028 [0.039]	−0.039 [0.121]	0.054 [0.051]	−0.634 [0.645]	0.038 [0.026]	−0.016 [0.037]	−0.032 [0.040]	−0.028 [0.123]
BIND	−0.089 [0.102]	−1.482 [1.424]	−0.159ᵇ [0.068]	−0.192ᵇ [0.085]	0.254ᵇ [0.107]	−0.548ᵇ [0.263]	−0.024 [0.126]	−2.273 [1.725]	−0.098 [0.079]	−0.209ᵇ [0.099]	0.223ᵃ [0.120]	−0.642ᵇ [0.319]
BIND × BLACK							0.003 [0.246]	4.142 [3.539]	−0.238 [0.170]	0.236 [0.222]	0.247 [0.222]	0.245 [0.726]
BIND × OTH							−0.946ᵇ [0.446]	0.238 [4.906]	−0.172 [0.269]	−0.553ᵇ [0.248]	−1.519ᵇ [0.754]	0.585 [0.832]

Continued on next page.

22

Table 5: Continued

	(1) FIN-RELA	(2) OCC-PREST	(3) CLASS	(4) HAPPY	(5) CHBF25	(6) EDUC	(7) FIN-RELA	(8) OCC-PREST	(9) CLASS	(10) HAPPY	(11) CHBF25	(12) EDUC
BLACK	-0.229[c]	-1.315[a]	-0.137[c]	-0.195[c]	0.014	0.068	-0.112	0.056	-0.138[b]	-0.257[c]	-0.006	0.097
	[0.067]	[0.735]	[0.041]	[0.042]	[0.051]	[0.146]	[0.105]	[1.596]	[0.055]	[0.069]	[0.081]	[0.263]
OTHER-RACE	-0.177[b]	1.958[a]	-0.057	-0.175[c]	0.010	0.558[c]	-0.007	2.667	-0.036	-0.123	0.230	0.875[c]
	[0.080]	[1.056]	[0.043]	[0.055]	[0.082]	[0.196]	[0.134]	[1.955]	[0.066]	[0.082]	[0.162]	[0.270]
MALE	0.187[c]	-0.568	-0.031	0.026	-0.156[c]	0.012	0.187[c]	-0.609	-0.031	0.027	-0.152[c]	0.005
	[0.039]	[0.570]	[0.023]	[0.028]	[0.029]	[0.093]	[0.039]	[0.571]	[0.023]	[0.028]	[0.030]	[0.093]
AGE	0.031[c]	0.810[c]	0.005	0.002	0.026[c]	0.172[c]	0.031[c]	0.816[c]	0.005	0.002	0.026[c]	0.172[c]
	[0.005]	[0.068]	[0.003]	[0.005]	[0.006]	[0.020]	[0.005]	[0.068]	[0.003]	[0.005]	[0.006]	[0.020]
$AGE^2/100$	-0.024[c]	-0.667[c]	0.003	0.000	-0.026[c]	-0.161[c]	-0.025[c]	-0.671[c]	0.003	0.000	-0.026[c]	-0.161[c]
	[0.005]	[0.069]	[0.004]	[0.005]	[0.005]	[0.020]	[0.005]	[0.069]	[0.004]	[0.005]	[0.005]	[0.021]
PAEDUC	0.028[c]	0.599[c]	0.028[c]	-0.002	-0.010[b]	0.181[c]	0.028[c]	0.595[c]	0.028[c]	-0.002	-0.010[b]	0.179[c]
	[0.005]	[0.070]	[0.003]	[0.003]	[0.004]	[0.012]	[0.005]	[0.070]	[0.003]	[0.003]	[0.004]	[0.012]
MAEDUC	0.019[c]	0.348[c]	0.011[c]	0.009[b]	-0.008[a]	0.151[c]	0.020[c]	0.354[c]	0.010[c]	0.010[b]	-0.008	0.152[c]
	[0.006]	[0.096]	[0.004]	[0.005]	[0.005]	[0.017]	[0.006]	[0.096]	[0.004]	[0.005]	[0.005]	[0.017]
SIBS	-0.010	-0.333[c]	-0.012[c]	0.000	0.011[b]	-0.128[c]	-0.01	-0.330[c]	-0.012[c]	0.000	0.011[b]	-0.128[c]
	[0.006]	[0.078]	[0.004]	[0.005]	[0.005]	[0.016]	[0.006]	[0.079]	[0.004]	[0.005]	[0.005]	[0.016]
ABROAD	0.076	0.805	0.079[c]	-0.009	-0.079[a]	0.310[b]	0.073	0.827	0.078[c]	-0.012	-0.080[a]	0.319[b]
	[0.050]	[0.897]	[0.030]	[0.037]	[0.043]	[0.139]	[0.050]	[0.892]	[0.030]	[0.038]	[0.043]	[0.140]
NOBS	2866	3661	3773	2880	1157	3778	2866	3661	3773	2880	1157	3778
R^2	0.10	0.12	0.12	0.03	0.13	0.28	0.10	0.12	0.13	0.03	0.14	0.28
P:BIND							0.164	0.513	0.047	0.002	0.007	0.196
P:POP							0.201	0.195	0.378	0.747	0.368	0.111
P:16	0.000	0.000	0.000	0.097	0.000	0.000	0.000	0.000	0.000	0.100	0.000	0.000
P:FNF ex POP and BIND	0.528	0.266	0.004	0.066	0.417	0.952	0.615	0.201	0.004	0.056	0.552	0.965

Note: See Table 3. The regression in column 7-12 also includes background variables from the respondent was 16 years of age. These are whether the individual was raised Jewish, Catholic or Protestant, nine dummy variables for the Census Region they were raised in, and six dummy variables for the city size that they were raised in.

Columns 7 through 12 of Table 5 report a number of findings that indicate that the relationship between the race content of a name may not be straightforward. There are three main findings in these columns of the table. First, in column 7, the estimate coefficient on $BLACK \times POP$ and $BIND \times OTHERRACE$ are both negative and statistically significant. That is, black respondents with more popular names, not more black ones, tend to report lower financial status, as do non-black and non-whites with 'blacker' names. For example, a non-black non-white with a name that is deemed fully black reports almost a .8 lower financial status as compared to a non-black non-white with a name devoid of blackness. Such findings are consistent with the identity view that blacker names provide net benefits to blacks, though not to others. Further evidence in support of this is in column 8 where occupational prestige is the left hand side variable. Again, the coefficient on the term $BLACK \times POP$ is statistically significant and negative. Moreover, the coefficient on $BLACK \times BIND$ is positive and statistically significant. Together, these two sets of findings indicate that blacker names by black are associated with more prestigious occupations, but not for blacks with more conformist first names or non-white non-blacks with 'blacker' names. Second, the results in columns 8, 9, 11 and 12 indicate that the basic results for CLASS, CHB25 and EDUC are unaffected by the interaction terms – see F-statistics and/or coefficients on BIND or POP. Third, the negative and statistically significant coefficient on $BIND \times OTHERRACE$ indicates that the results on BIND and HAPPY in column 4 are actually concentrated among non-black non-whites.

Just as the impact of first names features on an individual's lifetime outcomes can be differentiated based on race, the results in Table 6 investigate the extent to which gender also has a role to play. In particular, we separately re-estimated for males and females the specification of (4) that also includes interactive effects between race and POPULARITY and BIND. Columns 1 through 6 report the results for MALES, while the results for FEMALES are reported in columns 7 through 12. These results can be contrasted with the full sample results presented in Table 5 to help understand the separate and interactive roles that race and gender play in the transmission of first name features to lifetime outcomes.

Table 6: Respondent's Lifetime Outcomes Regressed on First Name Features and All Background Information By Gender

	MALES						FEMALES					
	(1) FIN-RELA	(2) OCC-PREST	(3) CLASS	(4) HAPPY	(5) CHBF25	(6) EDUC	(7) FIN-RELA	(8) OCC-PREST	(9) CLASS	(10) HAPPY	(11) CHBF25	(12) EDUC
POPULARITY	-0.052	0.788	0.014	-0.054	-0.039	0.020	0.030	-0.133	0.030	-0.023	-0.038	-0.049
	[0.073]	[1.121]	[0.049]	[0.056]	[0.061]	[0.180]	[0.057]	[0.877]	[0.037]	[0.044]	[0.057]	[0.155]
POP × BLA	-0.180	-1.478	-0.095	-0.017	0.069	0.163	-0.337[b]	-3.234	0.208[a]	0.095	-0.189	-0.113
	[0.176]	[1.976]	[0.157]	[0.156]	[0.131]	[0.474]	[0.155]	[2.787]	[0.116]	[0.119]	[0.167]	[0.390]
POP × OTH	-0.085	0.598	0.037	-0.122	0.160	0.105	-0.037	-2.138	-0.030	0.153	0.024	-1.405[c]
	[0.250]	[4.645]	[0.145]	[0.154]	[0.230]	[0.532]	[0.278]	[2.610]	[0.153]	[0.165]	[0.283]	[0.459]
SYLLABLES	0.016	0.198	-0.018	0.014	0.052	-0.114	0.043	0.231	0.011	-0.068[b]	0.011	0.223[b]
	[0.044]	[0.638]	[0.028]	[0.026]	[0.049]	[0.098]	[0.033]	[0.515]	[0.025]	[0.033]	[0.047]	[0.109]
SCRABBLE	-0.004	0.164[a]	-0.003	0.005	-0.010[a]	0.008	-0.006	0.033	0.001	0.008[a]	-0.006	-0.02
	[0.007]	[0.091]	[0.005]	[0.004]	[0.006]	[0.017]	[0.005]	[0.073]	[0.004]	[0.005]	[0.007]	[0.017]
AH-END	-0.270[b]	0.621	-0.008	-0.020	0.037	-0.145	0.020	-0.560	-0.089[a]	0.083	-0.029	-0.225
	[0.127]	[2.976]	[0.122]	[0.095]	[0.148]	[0.419]	[0.062]	[0.960]	[0.048]	[0.051]	[0.084]	[0.171]
OH-END	0.026	-2.558	0.139	0.075	-0.028	-0.609	0.125	8.451[c]	0.382[c]	0.583[c]	-0.104	1.019[b]
	[0.186]	[3.026]	[0.100]	[0.126]	[0.153]	[0.699]	[0.083]	[2.948]	[0.044]	[0.082]	[0.080]	[0.408]
EE-END	-0.124	-1.993	-0.056	-0.004	0.017	-0.13	0.056	0.172	0.038	0.028	-0.005	-0.109
	[0.089]	[1.305]	[0.056]	[0.060]	[0.101]	[0.249]	[0.057]	[1.071]	[0.042]	[0.049]	[0.061]	[0.165]
VOW-BEG	-0.090	-0.228	-0.082	-0.041	-0.018	-0.146	-0.067	1.214	0.021	0.000	-0.032	0.016
	[0.093]	[0.908]	[0.050]	[0.058]	[0.069]	[0.284]	[0.059]	[1.246]	[0.045]	[0.043]	[0.049]	[0.138]
CON-END	-0.118[b]	-0.147	-0.055	-0.042	0.087	0.057	0.064	0.613	-0.011	-0.020	-0.011	-0.031
	[0.059]	[1.309]	[0.043]	[0.052]	[0.086]	[0.219]	[0.072]	[0.780]	[0.036]	[0.044]	[0.057]	[0.148]
UNUSUAL-SP	-0.004	0.045	-0.02	-0.01	-0.035	0.174	-0.149[b]	0.235	-0.018	-0.072	0.017	-0.078
	[0.153]	[1.193]	[0.047]	[0.071]	[0.070]	[0.245]	[0.065]	[0.986]	[0.057]	[0.049]	[0.054]	[0.173]
NICKNAME	0.004	0.388	0.041	-0.047	-0.004	-0.003	0.108	-1.935[a]	0.009	0.059	-0.054	-0.121
	[0.059]	[0.955]	[0.038]	[0.044]	[0.053]	[0.163]	[0.090]	[1.036]	[0.036]	[0.064]	[0.088]	[0.177]
BIND	0.270	-0.618	-0.110	-0.156	0.070	0.216	-0.240	-3.701	-0.093	-0.279[b]	0.316[b]	-1.138[c]
	[0.192]	[2.747]	[0.119]	[0.162]	[0.241]	[0.450]	[0.166]	[2.257]	[0.107]	[0.133]	[0.157]	[0.437]
BIND × BLACK	-0.604[a]	1.918	-0.064	-0.163	0.398	0.406	0.491	6.626	-0.294	0.545[a]	0.102	0.110
	[0.323]	[6.178]	[0.244]	[0.341]	[0.452]	[1.029]	[0.342]	[4.120]	[0.223]	[0.314]	[0.261]	[0.971]
BIND × OTH	-0.942[a]	2.765	0.210	-0.467	-0.920	0.553	-1.124[a]	-2.639	-0.618[a]	-0.568	-1.572[a]	0.418
	[0.514]	[5.835]	[0.283]	[0.362]	[1.560]	[1.123]	[0.647]	[7.007]	[0.372]	[0.360]	[0.926]	[1.179]

Continued on next page.

25

Table 6: Continued

	MALES						FEMALES					
	(1) $FIN-RELA$	(2) $OCC-PREST$	(3) $CLASS$	(4) $HAPPY$	(5) $CHBF25$	(6) $EDUC$	(7) $FIN-RELA$	(8) $OCC-PREST$	(9) $CLASS$	(10) $HAPPY$	(11) $CHBF25$	(12) $EDUC$
BLACK	−0.019	−3.492	−0.075	−0.131	−0.065	−0.523	−0.155	0.979	−0.169[b]	−0.343[c]	0.041	0.396
	[0.169]	[2.261]	[0.124]	[0.128]	[0.128]	[0.412]	[0.121]	[2.100]	[0.075]	[0.092]	[0.115]	[0.295]
OTHER-RACE	−0.048	0.161	−0.139	−0.113	0.071	0.029	0.056	4.238[b]	0.074	−0.077	0.229	1.404[c]
	[0.206]	[3.699]	[0.117]	[0.124]	[0.331]	[0.477]	[0.215]	[1.760]	[0.122]	[0.118]	[0.208]	[0.374]
AGE	0.033[c]	0.881[c]	0.003	−0.001	0.022[b]	0.188[c]	0.032[c]	0.778[c]	0.008[a]	0.005	0.028[c]	0.159[c]
	[0.007]	[0.113]	[0.005]	[0.006]	[0.009]	[0.023]	[0.007]	[0.096]	[0.004]	[0.006]	[0.007]	[0.024]
$AGE^2/100$	−0.028[c]	−0.703[c]	0.004	0.002	−0.024[c]	−0.176[c]	−0.023[c]	−0.668[c]	0.001	−0.003	−0.027[c]	−0.151[c]
	[0.008]	[0.116]	[0.005]	[0.006]	[0.009]	[0.024]	[0.007]	[0.095]	[0.004]	[0.006]	[0.006]	[0.025]
PAEDUC	0.027[c]	0.607[c]	0.019[c]	−0.003	0.001	0.184[c]	0.027[c]	0.554[c]	0.034[c]	−0.001	−0.017[b]	0.170[c]
	[0.008]	[0.120]	[0.005]	[0.005]	[0.007]	[0.024]	[0.007]	[0.090]	[0.004]	[0.004]	[0.007]	[0.017]
MAEDUC	0.024[b]	0.246	0.019[c]	0.006	−0.016[b]	0.123[c]	0.017[b]	0.425[c]	0.005	0.011[a]	−0.002	0.174[c]
	[0.010]	[0.184]	[0.007]	[0.009]	[0.006]	[0.029]	[0.007]	[0.095]	[0.005]	[0.006]	[0.007]	[0.020]
SIBS	−0.010	−0.174	−0.014[b]	0.008	0.000	−0.103[c]	−0.01	−0.484[c]	−0.010[a]	−0.007	0.022[c]	−0.157[c]
	[0.008]	[0.123]	[0.006]	[0.008]	[0.009]	[0.024]	[0.009]	[0.103]	[0.005]	[0.006]	[0.007]	[0.022]
ABROAD	0.031	1.153	0.054	0.058	−0.064	0.402[a]	0.110	0.537	0.107[b]	−0.068	−0.095	0.201
	[0.078]	[1.151]	[0.042]	[0.052]	[0.056]	[0.207]	[0.068]	[1.163]	[0.042]	[0.053]	[0.064]	[0.187]
NOBS	1307	1665	1691	1315	514	1696	1559	1996	2082	1565	643	2082
R^2	0.11	0.16	0.13	0.04	0.12	0.28	0.11	0.12	0.14	0.06	0.15	0.30
P:BIND	0.138	0.963	0.546	0.121	0.658	0.750	0.066	0.219	0.040	0.011	0.024	0.026
P:POP	0.396	0.764	0.825	0.389	0.847	0.975	0.190	0.442	0.120	0.733	0.260	0.006
P:16	0.001	0.000	0.000	0.002	0.001	0.000	0.000	0.056	0.000	0.137	0.002	0.000
P:FNF ex POP and BIND	0.449	0.235	0.084	0.799	0.440	0.788	0.015	0.003	0.000	0.000	0.811	0.001

Note: See Table 5. The results in columns (1)-(6) are for males, while the results for females are reported in columns (7)-(12).

26

There are two key findings presented in Table 6. First, for males only, the effect of a higher BIND index for BLACK's on their financial relative position is negative and statistically significant – a movement from a non-black name to a fully black name reduces a males social standing by .6, which is slightly less than one standard deviation as reported in Table 1. In contrast, for females only, the effect of a higher POPULARITY index for BLACK's on their financial relative position is negative and statistically significant. However, for both males and females, the impact of a higher BIND index for OTHERRACE (i.e non-blacks non-whites) on their financial relative position is negative and statistically significant. In other words, black males with blacker names have lower financial relative status, as do non-black non-whites regardless of gender. In contrast, black women have lower financial status if they have more popular names. Given the amount of background factors controlled for, it would seem unlikely that these results can be explained away by unobserved labor productivity.

Second, the impact of names appears to have a much stronger effect on women as compared to men.[29] For example, for all the lifetime outcomes listed in columns (7) through (12), the p-values at the bottom of Table 6 indicate that either the BIND variables (P:BIND) is below .1 or the first names features exclusive of POPULARITY and BIND (P:FNF ex POP and BIND) is below .1, or both are. In stark contrast, the p-values for MALES are typically a great deal higher than this cut-off. This fact that first names appear to have a greater impact on female's lifetime outcomes, even after controlling for a myriad of background characteristics, can be further seen in the rows for the impact of BIND and its interaction terms on lifetime outcomes. For females only, BIND is negatively and significantly related to EDUC, HAPPY and positively associated with CHBF25. Thus, there seems to be broader evidence of a link between first names and financial outcomes for women than for men.[30]

Taken together, two clear findings emerge from Tables 5 and 6. First, in many instances, there is consistent evidence that the blackness of a name is related to factors that are related

[29]Interestingly, Bertrand and Mullainathan (2003) report a similar ratio male-to-female callback rates for respondents with African American names and those with white names.

[30]This difference between statistical significance of name features by gender does not seem to be driven by differences in the overall fit of the model or by the differences in sample sizes across genders, since these two variables remain roughly comparable across genders in different specifications.

to labor productivity – education, happiness and early fertility. This finding is consistent with Fryer and Levitt (2003). Second, there are also two pieces of partial evidence that discrimination also plays a role in the impact of first names a view consistent with Bertrand and Mullainathan (2003). For example, black males with blacker names report lower financial status even though they do not, *ceteris paribus*, have significantly lower education, for example. Second, the fact that non-white non-blacks with blacker names have lower financial standing (both males and females) suggests an implicit discrimination cost from having blacker names that for these individuals in not accompanied by whatever benefits result from having a black identity. Further support for this identity based role for discrimination is found in that Black Females with more popular (i.e. whiter) names report higher class but also lower financial status.

3 Interpretations and Conclusion

Our research on the economic impact of first names demonstrates three findings. First, the features of an individual's name reveal a great deal of information their background, even the part of their background that they inherited at birth. Second, the features of an individual's name have a great deal of predictive power about their lifetime outcomes such as income and social status, educational attainment, whether they have a child before 25, etc ... Third, in a great many instances these first names also continue to have predictive power for an individual's lifetime outcomes even when one controls for their exogenous background factors. Indeed, two dimensions that emerge as important determinants of the linkage between first names and lifetime outcomes are race and gender.

Ultimately, there are two potential explanations for the transmission mechanism which links names and lifetime outcomes: namely, either individuals face discrimination or preference based on their first names, or first names are correlated with factors that affect an individual's labor productivity. The former explanation has been favored by Bertrand and Mullainathan (2003) while the latter has been favored by Fryer and Levitt (2003). Certainly our data provides evidence of the latter. We demonstrate that gender and race based components of names are independent determinants of less happy individuals, who are less educated, and who have children before 25.

These outcomes are likely related to workers that are less productive, *ceteris paribus*. However, these factors do not generally affect income, though they affect the respondent's perceived social standing.

Nevertheless, while we do find evidence of the link between the race and gender qualities of a first name and factors that are related to labor productivity, we believe that our study does reveal some indirect evidence of discrimination. In particular, our findings that the blackness of a name and the popularity of a name have differential impacts based on the respondent's race and gender indicates that names can have an impact on lifetime financial outcomes that is disproportionate to their relationships to fundamentals. Specifically, the finding that black males with blacker names have lower financial relative position reflects potential straightforward discrimination that reduces these individuals opportunities, subject to the caveat that we have controlled for a sufficient amount of background factors. In addition, the finding that non-black non-whites demonstrate lower financial relative status potentially demonstrates a more subtle form of discrimination based on identity. Namely, while blacks may face costs and benefits from a blacker name that correspond to the groups they do and don't identify with, non-black non-whites with blacker names would only realize the former and not the latter.

References

Akerloff, George, and Kranton, Rachel, 2000, "Economics and Identity," *Quarterly Journal of Economics,* 115(3), 715-753.

Bertrand, Marianne, and Mullainathan, Sendhil, 2003, "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *NBER Working Paper # 9873.*

Blanchflower, Daniel and Oswald, Andrew, 2004, "Well-Being Over Time in Britain and the USA," *Journal of Public Economics,* forthcoming.

Frey, Bruno and Stutzer, Alois, 2002, Happiness and Economics: How the Economy and Institutions Affect Well-Being, Princeton University Press, Princeton, NJ.

Fryer, Roland , Jr. and Levitt, Steven, 2003, "The Causes and Consequences of Distinctively Black Names," *NBER Working Paper # 9938.*

Levy, David, 1997, "Adam Smith's Rational Choice Linguistics," *Economic Inquiry,* 35(3), 672-678.

Lieberson, Stanley, 2000. *A Matter of Taste: How Names, Fashions and Culture Change,* Yale University Press, New Haven, CT.

Lierberson, Stanley and Bell, Eleanor O, 1992 "Children's First Names: An Empirical Study of Social Taste, " *American Journal of Sociology,* 98(3), 511-554.

Lieberson, Stanley and Mikelson, Kelly, 1995, "Distinctive African American Names: An Experimental, Historical and Linguistic Analysis of Innovation," *American Sociological Review,* 60(6), 928-946.

Wallace, Carol, 1992, *20,001 Names for Baby*, Avon Books, New York, NY.

Data Appendix: Variables Used in the Analysis

Key Variables of Interest

FINRELA The respondent's household financial income relative to others. It is measured on a self reported scale of 1 to 5, with 1 being that your financial status is well below the mean, and 5 being that it is far above the mean.

EDUC Number of years of education.

CLASS The respondent's social class relative to others. It is measured on a self reported scale of 1 to 4, with 1 being lower class, 2 being working class, 3 being middle class and 4 being upper class. being upper class.

POPULARITY Variable running from 0.000 to 1. Has value of one minus the respondent's name rank in gender specific Social Security Administrations' 'Top 1,000 popularity list for the birth decade of the individual divided by 1000. Has value 0 if the name does not appear on the list. Positive regression coefficient then means that Y decreases with popularity.

BIND Blackness index. Probability of being black conditional on having the name in our sample, excluding the own individual's observation. Calculated in this manner only for names that have at least two occurrences in the sample. For unique names, the BIND is set equal to the probability that the respondent is black, conditional on having a unique name.

SYLLABLES The number of syllables in a name. Coded by following the standard rules of syllablication. See http://english.glendale.cc.ca.us/syllables.html.

SCRABBLE A measure of linguistic complexity, measured by the simple word score the name would receive from the popular board game SCRABBLE.

UNUSUAL-SP A dummy variable equal to one if the name is spelled unusually, and zero otherwise. A name was determined to be unusually spelled if multiple spellings were pronounced the same way (e.g. Jeffrey and Geoffrey), the more popular name appeared at least four times, and the less popular spelling occurred half or fewer times as the most popular spelling.

NICKNAME A dummy variable equal to one if the name is denoted a nickname in Wallace (1992), and zero otherwise.

AH-END A dummy variable equal to one if the name ends with an 'ah' sound (e.g. Mariah), and zero otherwise.

0H-END A dummy variable equal to one if the name ends with an 'o' sound (e.g. Mario), and zero otherwise.

EE-END A dummy variable equal to one if the name ends with an 'ee' sound (e.g. Brittany), and zero otherwise.

VOW-BEG A dummy variable equal to one if the name begins with a vowel (e.g. Abigail), and zero otherwise.

CON-END A dummy variable equal to one if the name ends with a consonant

BIND The probability that a person with a given name is black.

OCCPREST The respondent's occupational prestige score based on 1980 occupational classifications and income data.

CHBF25 For men and women over 25, a dummy variable equal to 1 if they report having a child before 25, and zero otherwise. This question is not asked in all years.

HAPPY An index of happiness that is 1 if you are 'not too happy', 2 if you are 'pretty happy' and 3 if you are 'very happy'.

Additional Control Variables

AGE Age of Respondent.

BLACK Dummy variable 1 if race is black, and 0 otherwise.

ABROAD Dummy variable 1 if respondent or either of his or her parents was not born in the U.S., and 0 otherwise.

MALES Dummy variable 1 if male, and 0 if female.

LMEX Labor market experience, equal to $AGE - EDUC - 5$.

YRBORN The year in which the respondent was born.