Forthcoming in the *Journal of Personality and Social Psychology*

SPURIOUS? NAME SIMILARITY EFFECTS (IMPLICIT EGOTISM) IN MARRIAGE, JOB AND MOVING DECISIONS

Uri Simonsohn*
The Wharton School
University of Pennsylvania

Abstract.
Three papers published in this journal have shown that a disproportionate share of people choose spouses, places to live, and occupations with names similar to their own. These findings, interpreted as evidence of implicit egotism, are now included in most modern social psychology textbooks and many university courses. This paper successfully replicates the original findings but then shows that they are most likely caused by a combination of cohort, geographic and ethnic confounds, and reverse causality.

Keywords: Name-letter-effects, field studies, implicit egotism, statistics, spurious results.

* I would like to thank Jon Baron, Leif Nelson and Maurice Schweitzer for encouraging me to pursuit this project in its early stages, and for feedback and comments provided throughout. I would also like to thank Dan Benjamin, Eugene Caruso, Tom Gilovich, Emir Kamenica, Aparna Labroo, Katy Milkman, Mike Norton, Devin Pope, Deborah Small, Alison Wood, and George Wu for valuable comments. Undergraduate Wharton students Daniel Li and Brendan Forster provided outstanding research assistance. Alessandro Acquisti generously shared some of the data used in this paper. A special thanks to Marcello Gallucci who provided useful feedback despite being aware of my strong reservations with (Gallucci, 2003).
All remaining errors are the author's sole responsibility.

"*There are many hypotheses in science which are wrong. That's perfectly all right; they're the aperture to finding out what's right. Science is a self-correcting process*." Carl Sagan

The self-correcting aspect of science is arguably its strongest virtue compared to other methods of discovery. Science, however, is not literally *self*-correcting. Only if researchers are willing to exert the effort needed to revisit widely accepted findings, to leave aside for a moment the pursuit of new questions in order to revisit old ones, can science achieve its promise.

In line with this observation, this paper challenges a literature that is now broadly accepted in social psychology; one that is part of many textbooks, and both undergraduate and graduate courses. [1]

This paper re-examines the evidence interpreted as supporting the proposition that, because of 'implicit egotism' –the subconscious attraction to targets connected with the self - people disproportionately choose spouses, places to live and occupations with names similar to their own (Jones, Pelham, Carvallo, & Mirenberg, 2004; Pelham, Carvallo, DeHart, & Jones, 2003; Pelham, Mirenberg, & Jones, 2002), for a review see (Pelham, Carvallo, & Jones, 2005).

The evidence of implicit egotism from the laboratory is both abundant and convincing. It started with the demonstration of the Name Letter Effect (NLE), consisting of the observation that people like letters contained in their name more than other letters (Nuttin, 1985). The NLE has since been amply replicated (see e.g. Jones, Pelham, & Mirenberg, 2002; Kitayama & Karasawa, 1997; Koole, Dijksterhuis, & van Knippenberg, 2001; Nuttin, 1987) and extended to birthday number (Jones, et al., 2004;

[1] (see e.g. the textbooks by Baron, Branscombe, & Byrne, 2008; Baumeister & Bushman, 2010; Breckler, Olson, & Wiggins, 2005; Myers, 2009; Sanderson, 2009; Schacter, Gilbert, & Wegner, 2010)

Jones, et al., 2002; Kitayama & Karasawa, 1997), to the liking of products with brands that resemble people's names (Brendl, Chattopadhyay, Pelham, & Carvallo, 2005), and to the pursuit of consciously avoided outcomes (Nelson & Simmons, 2007).

The investigation of the NLE in decisions outside the lab began with the work of Pelham, et al. (2002), who showed a disproportionate number of people gravitating towards states, cities and occupations with names resembling their own (e.g., that Lauras are lawyers and Florences live in Florida). They refer to this generalization of the NLE as implicit egotism. Pelham, Carvallo, DeHart, & Jones (2003) extended these findings to streets and town names and Jones et al. (2004) showed a disproportionate fraction of people marrying others with similar names.

The current research sets out to re-evaluate these three papers, henceforth referred to as JPSP1, JPSP2 and JPSP3 respectively. Altogether these include sixteen field studies demonstrating eight main findings. These are, in the order discussed in this paper, that people are disproportionately likely to: (1) marry someone with a similar last name, and (2) first name, (3) choose occupations with names resembling their first name, (4) move to states with names resembling their first name, live (5) in states, (6) in towns, and (7) on streets, with names resembling their last name, and to (8) live in towns whose names contains their birthday numbers.

This paper is organized along these eight original findings. Each is first summarized, then potential problems with the causal interpretation of the results are put forward, and new studies that test for implicit egotism net of these problems are then presented.

*** Table 1 ***

While the challenge for the implicit egotism interpretation of the evidence is slightly different in each study, all such challenges are instantiations of the same perennial one for studies without random assignment: unobservable heterogeneity.

In particular, to examine the causal effect of people's names on the choices they make, one must take into account the impact of unobservable variables that may affect both the names people receive as newborns and the decisions they make as individuals.

Here I focus primarily on the impact of cohort, geographic and ethnic heterogeneity, as these greatly influence both baby naming and major life decisions. My general approach for taking these variables into account is to conduct analyses similar to those of JPSP1-3, but in subsamples for which there is a substantially reduced degree of heterogeneity in such variables. While implementation details vary across studies, the overarching approach and conclusion is the same: heterogeneous samples exhibit name-similarity-effects, those that are homogenous in ethnicity, age and geography do not.

The statistical tests themselves are just as simple and intuitive as those reported in JPSP1-3, typically consisting of $\chi^2(1)$ performed on 2x2 tables. The results are presented in figures that make the impact of eliminating the proposed confound evident to the naked eye.

In re-examining evidence for implicit-egotism, the current research is related to (Gallucci, 2003), an early critique of JPSP1.[2] Most relevant for the present paper, Gallucci showed that there is considerable variation in the point estimates of name-similarity-effects in at least one of the ten studies in JPSP1, and expressed the concern that overall estimates that aggregate across names may hence be driven by a small

[2] JPSP2 is a response to Gallucci's comment.

minority of them that exhibit large effects. He also examined the unreliability of studies based on just two names.

All studies here include several names, and individual name estimates are reported. Aggregate estimates across names are also reported, in part to provide a direct comparison to JPSP1-3, and in part to test the joint null hypothesis that *all* effects in a given study are equal to 0.

A fundamental difference between (Gallucci, 2003) and the present paper, is that the former concludes that JPSP1's findings are not statistically-significant, that there simply is no name-similarity-effect to be explained beyond sampling error, while the present paper, in contrast, does conclude there is a sizable, robust, widespread *and statistically significant* name-similarity-effect in JPSP1 (and JPSP2&3), but that implicit egotism is not a likely cause for such effect, while cohort, ethnic and geographic confounds are. [3]

Existing Finding 1 - Last name & marriage

JPSP3 analyze marriage licenses and birth-certificates from two counties and from several states. They find that the share of marriages where groom and bride share an initial or last name is greater than expected in all samples considered.

[3] It is beyond the scope of this paper to critically review Gallucci's attempts to properly estimate statistical significance.
He carried out four major sets of analyses to assess the reliability and/or statistical significance of the studies in JPSP1, concluding that "the hypothesis [of a name-similarity-effect] is not supported for the large majority of names considered by the [JPSP1] authors" (pp. 789).
His analyses included (i) computing the significance of individual name effects through a new test Gallucci himself created as an alternative to a standard $\chi^2(1)$ used to compare expected with observed frequencies, (ii) analyzing all possible subsets of two-name studies that could have been conducted from studies with more than two names, (iii) interchanging names and places in studies to estimate false positive rates, and (iv) comparing the number of significant individual name effects across studies with different numbers of names in them.

Potential problems with these findings

JPSP3 mention a few possible confounds for their finding, including a spurious effect arising because people tend to marry within their ethnic group, and the possibility that a groom's last name may by mistake be recorded as also being the bride's last name.

They address the ethnicity concern by arguing that the counties on which they run the analyses are homogenously white (the most homogenous one was 95% white), and by replicating their (same-initial) findings in a subsample where all grooms had a Hispanic last name. They address the archival error concern by showing the same-last-name effect has not gotten smaller over time. [4]

New Analyses

Studies 1-4 here address these concerns of ethnic sorting and reverse causality, concluding that they probably do account for JPSP3's findings. Study 1 suggests that the same-last-name-initials effect documented in JPSP1 is driven in part by a same-entire-last-name effect, and in part by an ethnic confound.

Lacking ethnicity information in the original datasets, Study 2 uses a different data source, for which ethnicity is easier to infer, to assess the potential impact of ethnic sorting, finding that a miniscule degree of ethnic heterogeneity (e.g., that present in a population that is 99.9% white) can lead to quite sizable same-last-name effects.

To control for such ethnic confounds, Study 3 employs an ethnically homogenous sample, focusing only on marriages among people with Hispanic last names. It replicates

[4] JPSP3 argue that their study looking at entire last names also addresses ethnic sorting because they focus on five last names (Smith, Johnson, Williams, Jones and Brown) which are "common European American Names" and are "extremely common Caucasian names" (p.668). According to the Census of 2000 (Word, Coleman, Nunziata, & Kominski, n.d.), however, the share of people with these last names that are white is just 61.7%.

the same-last-name effect, but finds absolutely no effect for even extremely similar last names (e.g., no greater tendency for Mora to marry Morales, or Gonzales to marry Gonzalez), suggesting reverse causality, rather than implicit egotism, is the source of the effect. Finally, Study 4 provides direct evidence of the reverse causality mechanism, documenting that a considerable share of brides change their last name to their to-be-husbands' before marriage.

Study 1 – The 'same initial' marriage effect is a 'same-last-name' effect

Method. The marriage archives for Walker (GA) and Liberty (FL) counties used in JPSP3 (Study 1) were obtained from the USGenWeb Archives (http://usgwarchives.net). The Walker county dataset spans 1882-1990 (N=11,855) and Liberty county 1823-1965 (N=3,063).

Seeking a dataset that was both larger in number of observations and more concentrated time-wise, the Texas' Marriage License Application Index for 2001 was also obtained; this dataset includes all marriages occurring in Texas that year (N=195,030). [5] JPSP3 allude to this latter dataset in their discussion of studies 1-3 and report some secondary analyses on it (p. 669). [6]

To address the issue of ethnic sorting, a subset of the Texas 2001 dataset was created with marriages where both groom and bride had a last name belonging to the 200 last names identified as most homogenously Hispanic by the 2000 Census (Word, et al., n.d.), N=24,645.

---

[5] For unknown reasons sample sizes are slightly different from those reported in JPSP3. The results from my replications, therefore, while qualitatively equivalent differ quantitatively by small amounts.
[6] All marriage records for Texas marriages since 1966 are available from http://www.dshs.state.tx.us/vs/marriagedivorce/mindex.shtm.

The key analysis in all samples is the comparison of the expected vs. actual proportion of marriages with same initial/last names. The expected rate is the product of the base rates of the respective initials/last names among grooms and brides in the sample. For example, if 5% of grooms had a last name starting with *A* and 4% of brides did, then the expected rate of marriages between a groom and a bride with an *A* last name is 5%*4%=0.2%. As done in JPSP3, and in line with Cochran (1954), these proportions are added across all letters and the sum is compared to the total numbers of such marriages observed in the sample through a $\chi^2(1)$.

Results.

In Walker county the expected proportion of same initial marriages is 6.67% and the actual proportion 7.65% ($\chi^2(1)=18.4$, $p<.0001$). In Liberty County these proportions are 6.85% and 8.65% respectively ($\chi^2(1) = 15.6$, $p<.0001$). These results are nearly identical to those reported in JPSP3 (see footnote 5).

Here and for all other studies in this paper, the ratio of actual over expected frequencies will be the key variable of interest. This actual/expected ratio ($R_{A/E}$) would equal 1 in the absence of a name-similarity effect, and would be greater than 1 in its presence. The rates reported above translate into $R_{A/E} = 1.15$ for Walker county and $R_{A/E}=1.26$ for Liberty county, (see light gray bars in Figure 1).

It is useful to assess the fraction of the initials-effect that is driven by people marrying someone with the exact same last name. The dark gray bars in Figure 1 show the $R_{A/E}$ for same-last-name marriages. In Walker county, same-last-name marriages were more than five times the expected rate, and in Liberty county over three times. For example, in Walker county 0.17% of grooms should have married a bride with their same last name, but 0.92% did ($\chi^2(1)=381$, p <.0001).

*** Figure 1 ***

Subtracting the frequencies of same-last-name marriages from the frequencies of same-initial ones, one obtains the frequencies of marriages in which spouses share an initial but not the full last name. The resulting $R_{A/E}$'s are shown with black bars in Figure 1. For both samples originally used in JPSP3, the $R_{A/E}$ drop markedly both in size and statistical significance, indicating that same-last-name marriages account for a substantial share of the same-initial marriages.[7]

Despite the drop in size, the $R_{A/E}$ for Liberty county is still significant (at the 5% level) and quite high, at $R_{A/E}$=1.16. A likely explanation for this is ethnic sorting. According to the Census of 1900, 50.6% of people in Liberty county were black, and interracial marriage has only been legal in Florida since 1967. Furthermore, consulting the Florida census of 1885 I found that blacks and whites had quite different distributions of last name initials in that county (N = 1,288, $\chi^2(21) = 86.6$, $p < .0001$). [8]

This suggests that ethnic sorting may be behind the same-initial effect (net of same-last-name effect) in Liberty. If so, there should be no such effect in the ethnically homogenous sample of Hispanic marriages from Texas 2001; there is not.

The results are shown in the last three bars of Figure 1. Before excluding same-last-name marriages, there is a significant same-initial marriage effect ($R_{A/E}$ =1.14,

[7] The raw frequencies on which these and all other figures are based in this paper are tabulated in the appendix.

[8] Walker county was 16% Black in 1900. JPSP3 report percent of population being White only for Walker county, and for the Census of 1990 rather than 1900. Considering that the marriage data span 1820-1990, the Census of 1900 is probably more representative.

$\chi^2(1)=49$, $p<.0001$) also in this sample. This effect, however, is entirely driven by a same-last-name effect ($R_{A/E}=1.77$, $\chi^2(1)=244$, $p<.0001$). Once these observations are excluded, no same-initial effect remains ($R_{A/E}=1.00$, $\chi^2(1) = .014$, $p = .90$).

In sum, Study 1 shows that the same-initial effect is a combination of an ethnic confound and of a same-last-name effect. The same-last-name effect could, of course, arise because of implicit egotism. In Studies 2-4 I examine this possibility.

Study 2 –A Miniscule degree of ethnic heterogeneity can lead to large (& spurious) same-last-name effects

Given the considerable ethnic heterogeneity in Walker and Liberty counties alluded to above, and that the same-initial effect drops to 0 in the homogenous sample of Hispanic Texans, it seems important to have a sense of how much of a spurious same-last-name effect we might obtain as a consequence of ethnic homogeneity.

To look at this issue I computed, in the full Texas 2001 sample (N=195,030), the percentage of grooms that marry a bride with the same last name for each last name in the sample, leading to the identification of four obvious outlier last names: Patel, Nguyen, Kim and Tran. Table 2 summarizes some key facts about these four last names.

*** Table 2 ***

Column (4) shows the percentage of grooms marrying a bride with the same last names. These ratios ranged from extremely high (11% for Tran) to astronomically high (63% for Patel).[9] Column (6) shows the absurdly high $R_{A/E}$'s hovering in the hundreds

[9] Because I discovered these outliers through an exploratory process, I computed analogous calculations for the full Texas sample from 1966-2007 (N=7.6 million). The top four last names were also Patel, Nguyen, Kim and Tran in this larger sample, and had similarly high same-last-name marriage rates.

and thousands. [10] Most relevant for our purposes, column (7) shows the percentage of the overall same-last-name effect in the entire Texas 2001 sample that is accounted for by each last name. Fully 17.6% of the entire excess of same-last-name marriages is accounted for by these four last names, which combined are just 0.11% of the sample.

Considering that the expected frequency of same-last-name marriages in Texas for 2001 was just 251, and that these four last names combined have 217 such marriages, this means that if Texas was 99.89% homogenous (except only for these four last names), it would have an entirely ethnically confounded same-last-name ratio of actual over expected frequency of $R_{A/E} = 1.86$. In light of this, Study 3 is performed on the Hispanic subsample only.

Study 3 – Very similar last names do not disproportionately marry each other

Method. Study 1 suggests that the same-initials effect actually consists of a same-last-name effect. The latter, of course, could still arise because of implicit egotism. It may very well be the case that initials are not sufficiently strong triggers of implicit egotism, while identical last names are.

If a psychological mechanism was behind the same-last-name effect, be it implicit egotism or *any* other, then last names that are very similar should also have higher than expected marriage rates. For example, we would expect that grooms last named Morales

[10] As remarked by a referee, while these high ratios could arise if these ethnic groups married primarily among themselves (which they do), they could also arise if they exhibit strikingly high degrees of implicit egotism. To examine this rather unlikely possibility, I constructed a subsample of the Texas 2001 dataset where both groom and bride had one of Vietnam's 12 most common lastnames (Nguyen, Tran, Le, Pham, Huynh, Hoang, Phan, Vu, Vo, Dang, Bui and Do). Contrary to the "minorities have extremely high implicit egotism" story, in this sample (N=651), $R_{A/E}$=1.06, p=.42 (down from $R_{A/E}$=444 in the ethnically heterogeneous sample).

will marry not only a disproportionate share of Morales brides, but also of Mora and Moraga ones.

To examine this question, while continuing to control for ethnicity, a subset of ten pairs of very similar last names was extracted from the set of 200 previously mentioned Hispanic ones (see x-axis in Figure 2 for the full list).

I start from the full Texas marriage data (1966 to 2007) and create a subsample where both the groom and the bride had one of these 20 last names (N=13,335). For each last name, the actual vs. expected frequency of marriages are compared for people with identical last name, and for people with very similar last names.

For example, for grooms named Gonzalez (n=2,928), the same-last-name comparison contrasts the proportion of them marrying a Gonzalez bride (P=32.1%), to the proportion of grooms with the other 19 last names (n=10,407) doing so, P=19.3%, χ^2 (1) = 218.1, p <.0001. The second test compares the proportion of Gonzalez men marrying a Gonzales bride (P=10.6%) to the proportion of men not named Gonzalez doing so (P=10.8%), χ^2(1) = .05, p = .81. [11]

Results. The resulting ratios of actual over expected frequencies for each of these 20 last names are reported in Figure 2. The gray bars (depicting same-last-name rates) show $R_{A/E}$'s>1 for every single same-last-name considered (all *p's*<.01).

The black bars show the $R_{A/E}$ for similar-last names; all are quite close to $R_{A/E}$=1 and the only two statistically significantly different from it are below it. In other words,

[11] When computing the $R_{A/E}$ for similar names (e.g., marriage rates between Gonzalez and Gonzales) marriages between people with the exact same last name were excluded from the set of controls. This is important to do because if there is in fact an effect for very similar last names, but it is smaller than for exactly identical last names, then controlling for the identical last name may prevent us from observing the effect. Unsurprisingly, if this exclusion is not performed, the results show markedly *lower* than expected marriage rates between similarly last named individuals (intuitively, Gonzalez appears less likely than expected to marry a Gonzales because so many Gonzales do so, raising the average for everyone else).

there is no evidence that having an even extremely similar last name with another person increases the odds of marrying them by even a small amount.

The last two bars in Figure 2 report aggregate results. Here and in all other analyses in the paper that combine multiple 2x2 tables, the overall $R_{A/E}$ is the ratio of the sum of expected frequencies over the sum of actual frequencies. The significance of such aggregate $R_{A/E}$ is obtained with the method from Cochran (1954).[12]

The overall bars show that there are nearly twice as many same-last-name marriages as we would expect by chance ($R_{A/E}$=1.9, p <.001), but slightly *fewer* than expected very-similar-last name marriages ($R_{A/E}$=0.97, p = .438).

*** Figure 2 ***

Discussion. What could drive this dramatic disparity between identical and very similar last names? One possibility is that implicit egotism disappears with just one letter not matching (i.e., Gonzale<u>s</u> reminds Gonzale<u>s</u> of himself, but Gonzale<u>z</u> does not), but lab evidence for implicit egotism does find reliable effects when subjects share but a few letters with the target. For example, in JPSP3's (lab) study 6, they find that people are more attracted to individuals whose mock username shares just three letters with their own lastnames (as in Larry Murray being more attracted to STACEY_MUR than to STACEY_PEL, see pp. 675).

A similar alternative account, though more post-hoc, is that similar and identical names trigger implicit egotism, but almost identical last names trigger repulsion (e.g., a

---

[12] The method is as follows. Let the four cells of 2x2 table *i*, of a total of *n* such tables, be: a_i,b_i,c_i,d_i, and let DIFF be the difference between the sum of expected and actual frequencies, for the target cell, across the *n* tables. The statistic used to assess overall significance is $\chi^2(1) = \frac{DIFF}{\sum_{i=1}^{n}(a_i+b_i)(a_i+c_i)(c_i+d_i)(b_i+d_i)/N_i^2(N_i-1)}$.
See section 8 and appendix in Cochran, 1954 for an in-depth discussion. The formula itself was obtained from (Bickel, Hammel, & Oconnell, 1975).

person last named Gonzalez dislikes the name Gonzales). This explanation (first put forward by Brett Pelham in an interview discussing an earlier version of this paper, (Bialik, 2010)) seems unlikely for at least three reasons.

First, several last names in this study are *not* nearly identical (e.g., Mora:Morales or Aguilar:Aguirre) and we hence would expect a net effect of implicit egotism for them, but we do not see it. Second, it would be greatly coincidental that across different name pairs the negative effect of repulsion would always nearly exactly match the positive effect of implicit egotism, such that names with strong and weak same-last-name effects always have a very close to 0 very-similar-last-name effect. Third, as we shall see, people with very similar *first* names are disproportionately likely to marry each other (e.g., Andrew and Andrea), but the repulsion story would predict otherwise.

A third possibility for the robust same-last-name effect and entirely absent very-similar-last-name one is that grooms don't disproportionately marry brides with maiden names that match their last name, but rather, that some brides change their last name –to the groom's- *before* marriage. That is, that causality for marrying people with same last names is reversed.

There are multiple mechanisms that could lead to this: women may (i) marry a relative, (ii) upon divorcing or widowing marry a relative of their ex-husband, (iii) marry by the church/abroad/in-another-state/by-common-law-marriage, change their last name, then marry by the state, (iv) renew their vows through a new marriage, etc.

Most of these mechanisms are not easy to identify in the available data. One additional mechanism is: a couple marries, the wife changes her last name, the couple divorces, the couple then remarries (each other). Study 4 documents the surprising commonality with which this exact sequence occurs.

Study 4. Remarriages to same previously divorced spouse are frequent

Method. In 2001 there were 1,484 marriages in Texas between people with the same last name. A list was created containing the first name, middle name, and birth year of each bride in the list, and the full name and birth year of the groom, for each of these marriages. A computer script was then used to query Ancestry.com database of marriages and divorces for that combination of groom and bride characteristics. The script recorded how many records were found for that combination.

These 1,484 queries resulted in a set of 169 listings with more than 1 record. These 169 listings were given to a research assistant to hand-check if they appeared to correspond to a marriage-divorce-marriage sequence between the same two people.

Results. The number of marriages that seemed to correspond to remarriages was strikingly high: at least 68 of them. Figure 3 is a print-screen of the results from one such search on Ancestry.com. The second row shows that *Candi A. Hill* married *Stephen E. Nehring* in 1982. The first row shows that the couple divorced in 1997 (one needs to click on "See all information…" to see this), and the third that they married again in 2001, now both last named *Nehring*.

This high number of documented cases of reverse causality is particularly striking considering that we observe a rather small subset of all marriages having an artifactual origin for the matching last name (e.g., we can only document it for people who were *legally* married before, who married the first time and then divorced in some of the few states indexed by Ancestry.com, and who have relatively unusual combinations of names).

***** Figure 3 *****

Existing finding 2: First Names & Marriage

JPSP3 also examine whether people with similar *first* names are disproportionately likely to marry each other. They do so analyzing joint-listings in a nation-wide online phone directory (their Study 4).

They first identify twelve common female names that share at least four letters with a male name (e.g., Josephine with Joseph). Because name popularity changes over time, and people marry within their generation, JPSP3 group these twelve pairs into three sets of four, based on the popularity of the male name in the Censuses of 1960 & 1920.

They then obtain the number of phone listings for the 16 combinations of male-female names within each group, and compare expected vs. actual frequencies. JPSP3 find greater than expected frequencies, i.e., evidence of a first name similarity effect, for 10 of the 12 name-pairs considered.

Potential problems with this finding

The main potential problem with the first name and marriage finding is that the popularity of similar male and female names changes together over time. Using Social Security administration data on baby name popularity between 1930 and 1985, for example, I found that changes in the *rankings* of name popularity are positively correlated for ten of the twelve name pairs, five with $r > .9$.

On years when a greater share of baby girls were being named Erica, for example, more baby boys were being named Eric; twenty years or so later, the share of Erics marrying Ericas will naturally increase above expectations that assume name popularity is stable over time.

Does JPSP3 solution of grouping name-pairs based on the Census frequencies of male names fix this problem? Probably not. First, splitting data into of groups is a noisy way to control for a continuous variable; even if the frequencies are closer within a group than across groups (and we don't know if this is the case), they may still be quite different within groups. This problem is amplified by the fact that JPSP3 do not take into account changes in popularity of the female names, nor geographic or ethnic variation in changes in name popularity.

Another significant potential problem is the usage of telephone listings as a source of data, for at least two reasons. First, people listing their phone together need not be married to each other. They may be, for example, a parent and his or her child, and parents are disproportionately likely to name their children with names that resemble their own. [13] Second, phone listings are a nationwide sample, increasing the potential of a geographical confound.

To address these two concerns, Study 5 here analyzes marriage records rather than phone listings and employs control names explicitly chosen so as to minimize the role of cohort (and also ethnic and geographic) confounds when testing for implicit egotism.

[13] Using Ancestry.com data on birth certificates I computed the $R_{A/E}$ for children named similar to their parents, using the name combinations from JPSP1 (e.g., father: Eric, daughter: Erica). Every one of the twelve pairs had $R_{A/E}>1$, with an overall effect of $R_{A/E}= 1.93$.

Study 5 – The first-name effect on marriage disappears with proper controls.

Method. Identifying the *actual* frequency of marriages between people with similar names requires simply counting the observations in the data. The challenge is estimating a convincing *expected* frequency.

JPSP3 employ as expected-marriage-rates those of the other names in a given group. For example, Andrew's group also included Robert, Paul and Stephen, and hence JPSP3 tested whether the share of Andrews marrying Andreas (rather than Robertas, Paulas and Stephanies) is the same as the share of Roberts, Pauls and Stephens (combined) doing so.

As mentioned above, these calculations are likely to be biased due to cohort and also possibly ethnic and geographical confounds. An ideal control for Andrew would be a male name that if it weren't for an implicit egotism effect, we would strongly expect it to marry Andreas in the same rate as Andrews do.

One way to approximate this ideal control is to find male names that marry other women names in rates similar to Andrews'. For example, if 2% of Andrews marry Katys and 1% marry Schwandras, and Devins marry these women in those same rates, then we may use Devins as a control for Andrews.

To systematically find highly correlated control names for all 12 male and 12 female names used by JPSP3, I started from the full dataset of marriage licenses for Texas 1966-2007 (N = 7.58 million) and focused only on people's first names. After excluding those appearing less than 100 times, there were 2,226 male first names and 2,967 female ones. I then computed the frequency of marriages for every combination of male-female names, obtaining a table of 2,226 columns for males, and 2,967 rows for females, where each cell contains the actual frequency of marriages for that specific combination of names.

The correlation between any given two columns tells us how similar the distributions of brides' names are for two groom names, while the correlation between two rows how similar those two women names are in their choice of groom names.

In order to choose conservative controls from the perspective of implicit egotism the correlation calculations exclude names with the same initial of the target name (e.g., when computing it for Andrew, marriages to Agnes would not be considered) and control names must have a different initial (e.g., Amos is not used as a control for Andrew).

To test the first name similarity effect I hence created twelve name groups, one for each target name pair from JPSP3 (e.g., Stephen and Stephanie), and included in the group the top-3 most correlated names with each name.

For example, the group of Stephen and Stephanie includes the three most correlated names with Stephen (Michael, Douglas and Russell, *r*'s>.994), and with Stephanie (Michelle, Angela and Jennifer, r's>.992), leading to a total of 8 names (as in the original analyses).

The first name similarity effect on marriage is tested by comparing the proportion of Stephens marrying Stephanies (rather than Michelle, Angela or Jennifer, combined) to the proportion of Michael, Douglas and Russell (combined) doing so.

Results.

Figure 4 reports the results. The y-axis is the rate of actual over expected frequencies of marriages ($R_{A/E}$) for matching male/female names. The light gray bars report the original findings from JPSP3 (obtained directly from JPSP3 Table 2). Ten of the original twelve $R_{A/E}$'s are greater than 1, and nine of these significantly so ($p < .05$). One of the two $R_{A/E}<1$ is significant (Michael:Michelle).

JPSP3 aggregate the results by weighting observations equally *within* groups, and then taking a simple average *across* groups (see JPSP3, p.670), leading to an overall $R_{A/E}$=1.08. Figure 4 reports the overall effect computed with the same method, but statistical significance is not straightforward to compute and hence it is not reported in JPSP3 nor here. Adding up all the expected and actual frequencies as done elsewhere in this paper leads to an overall $R_{A/E}$ =1.03, p<.001.

The darker gray bars show the results from the same analyses used by JPSP3 on the new data, with an overall effect dropping to $R_{A/E}$=1.03 (from $R_{A/E}$=1.08), suggesting that employing phone-book listings may have had an impact on the results.

The black bars show the new analyses on the marriage data. Not a single $R_{A/E}$ in the set of 12 is significantly greater than 1.00 at the 10% level. All $R_{A/E}$'s are now noticeably closer to 1, including the three name pairs with the biggest $R_{A/E}$ in the original analyses. Note that Michael:Michelle also gets closer to 1 (from $R_{A/E}$=.94, p <.001 in the original analyses to $R_{A/E}$ = 1.00, p = .97, in the new). The overall effect, computed as in JPSP3, is now $R_{A/E}$=1.00. When computed as is done elsewhere in this paper it is $R_{A/E}$=1.01, p = .36.

*** Figure 4 ***

Existing Finding 3: First Names and Occupation

Study 7 in JPSP1 compares the numbers of dentists and lawyers with eight names that start with Den_ (e.g., Denise, Dennis) and eight names that start with La_ (e.g., Laura, Lawrence).[14] JPSP1 obtained their data from online dentist and lawyer directories, querying them for the respective names in the eight most populous US states.

They find a higher than expected share of professionals with a name-occupation match, though the effect is not significant for men (p = .14), and it is small in absolute magnitude for women; there are 1512 female La_ lawyers compared to the expected frequency of 1503.4 (p = .03).

Concerned about the size of these effects, JPSP1 then obtained dentist data for all 50 states and compared the Den_ names to names of similar popularity in the 1990 Census. For instance, Dennis is the 40th most common first name in the 1990 Census, so they compare the number of Dennis dentists (n=482) to that of Jerry dentists, ranked 39th (n=257), and Walter, ranked 41st (n=270). This large difference in frequencies corresponds to a $R_{A/E}$=1.43. They obtain similar results for all other seven Den_ names, but the frequencies for other names are very low and not reported in their paper.

In their Study 8 they employ again the 1990 Census as a baseline and compare the numbers of <u>Geo</u>rges and <u>Geo</u>ffreys who work in the <u>geo</u>sciences (e.g., geology) to the number of geoscientists with names similarly frequent in the Census (Pete, Bennie, Donald, Randolph, Mark, Jonathon, Kenneth, and Daniel). They find more Georges and Geoffreys than expected based on such controls, $R_{A/E}$=1.42.

[14] The study focuses on the following 16 names: Dennis, Denis, Denver, Denny; Denise Dennis, Denna, Denice; Lawrence, Larry, Lance, Laurence; Laurie, Laverne, Lauren, Laura,

Potential Problems with these findings.

One of the implicit assumptions on the dentists/lawyer analyses is that the ratios of lawyers to dentists, and of La_ to Den_ names, are stable over time, or at least that their changes are not correlated; otherwise, if over time there has been an increase in the relative number of lawyers, and an increase in the relative number of La_ to Den_ names, say, a spurious association between these two variables may arise.

It turns out that the popularity of La_ relative to Den_ names has been increasing in recent decades, and so has the relative frequency of lawyers to dentists. In 1970, for example, there were three lawyers for every dentist in the United States, in 2000 there were six. Den_ names, in turn, peaked in the 1960s and have dropped in popularity precipitously since then, while La_ names remained stably popular from the 1920s till 1990s. [15]

Given how small the dentists vs. lawyers effects are, how large the cohort confound is, and that there is no date of birth information in the professional directories that would allow one to control for cohort effects, I am unable to assess the role that they play on the results. I focus instead on the results that use control names based on Census frequencies.

One problem with such frequencies is that they do not capture, of course, age distributions; there may be just as many Dennises as Walters in the population, for example, but if Walter is an older name, which it is, fewer of them may be working in any occupation, including dentistry, because a higher proportion of them will have

[15] The data on the frequency of lawyers and dentists was obtained from the Bureau of Labor Statistics. The names data from BabyNameWizard.com (their source is the Social Security administration). See following two graphs: http://www.babynamewizard.com/voyager#prefix=den&ms=false&exact=fal and http://www.babynamewizard.com/voyager#prefix=la&ms=false&exact=false

retired. Consistent with this logic, Study 6 here shows that Dennises are just as over-represented (compared to Walters and Jerrys) among lawyers as they are among dentists.

In addition to missing age variation, Census frequencies miss geographic, ethnic and socioeconomic variation across names. When assessing if there are 'too many' George geoscientists one should take such factors into account. To this end Study 7 here compares the number of geoscientists with the names used by JPSP1, to other scientists with those names. If Georges and Geoffreys are more likely to be geoscientists because of unobserved heterogeneity, then they might also be disproportionately likely to be scientists in general; they are.

Study 6 – More Dennis than Walter dentists, but also, more Dennis than Walter lawyers

Method and Results.

Figure 5 plots the yearly number of newborns named Dennis, Jerry and Walter in the United States, between 1880 and 2007, as recorded by the Social Security administration. The figure shows that while the raw number of Dennises, Jerrys and Walters might have been very similar in 1990, the share of them in working age, especially some 10 years later when JPSP1 collected the data, is unlikely to have been the same.

*** Figure 5 ***

If this age (or some other unobserved) discrepancy is behind the greater frequency of Dennises dentists, then we should find the same pattern in other professions. Furthermore, Dennis should be a more frequent name in any sample over-representing younger (or alive) individuals. To test this prediction in a parsimonious manner I employed the same lawyer directory used by JPSP1 for the La_ Den_ study (www.martindale.com) and obtained the total frequency of lawyers with these names.

Consistent with the cohort-confound explanation there were more lawyers named Dennis (n=2889), than Jerry (n=1411) or Walter (n=2000). These numbers imply a ratio of actual over expected frequency for Dennis-*lawyer* of $R_{A/E}$=1.38, quite similar to the $R_{A/E}$=1.43 for Dennis-*dentist* reported above (this difference is not significant, $\chi^2(1) = 1.28$, $p = .25$). For Denise lawyers a similar pattern arose: there were more lawyers named Denise (n=947) than Beverly (n=454) or Tammy (n=247), implying $R_{A/E} = 1.72$. [16]

Study 7 – Georges no more likely to be in Geology than any other science

Method and Results.

One way to address the possibility that the geoscientists' result is driven by unobserved heterogeneity that influences both the likelihood to be named George or Geoffrey *and* whether one becomes a geoscientist, is to assess whether there is a disproportionate share of people so named in other sciences.

Using ProQuest Dissertations & Theses Database I obtained the total number of dissertations written, and the subset of these written in geosciences by people with different names. To focus on a sample presumably similar to the author index used by JPSP1, I restricted the queries to dissertations written since 1950.

[16] The statistical significance of the difference of the Dennis $R_{A/E}$'s for dentists and lawyers is obtained with a difference of proportions tests of the share of Dennises (rather than Walter or Jerry) among dentists(P=48%) and lawyers (P=46%).

Consistent with the findings from JPSP1, the number of geosciences dissertations written by a George or Geoffrey, 682, was greater than what would be expected based on the eight control names, 550.2, $R_{A/E}$=1.24, p<.0001. This was also the case, however, for all *other* sciences combined. A total of 25,351 non-geosciences dissertations were written by a George or a Geoffrey, compared the expected frequency of 19,250.6, $R_{A/E}$=1.32, p<.0001. These results suggests that, if anything, Georges and Geoffreys are slightly *less* likely to be geoscientists rather than any other kind of scientist ($R_{A/E}$ = .94, *p* = .07)

Existing Finding 4 - First Names & States

JPSP1 examines whether people are disproportionately likely to live in, and move to, states with names resembling their first name. They analyze data from the Social Security Death Index (SSDI), a file containing information on deceased individuals who participated in the US social security system. The dataset includes information on the state in which individuals obtained their social security number (SSN), and the zip code to which benefits were last sent. [17]

JPSP1 proxy for living in a state by having received benefits there, and cleverly proxy for moving to a state by receiving benefits there but having obtained the SSN in a different state. Their analyses focus on eight different first names that are either identical to, or share a few letters with, a state name (George (GA), Louis (LA), Virgil (VA) and Kenneth (KY); Georgia (GA), Louise (LA), Virginia (VA), Florence (FL)).

[17] The index is searchable through various websites, including http://search.ancestry.com/search/db.aspx?dbid=3693. It can be purchased from http://www.ntis.gov/products/ssa-quarterly.aspx.

Potential problems with this finding

JPSP1 were rightly concerned about the possibility that systematic differences in the popularity of baby names across states may lead to a spurious association between state and people's names (e.g., that more Virginias live in Virginia because a higher share of babies born there are so christened). Focusing on people obtaining their SSN in a different state would seem to address this concern, but unfortunately it does not.

First, as JPSP1 acknowledge, SSNs used to be obtained several years after birth. This means that some people received their SSN in a state other than their birth state. Because people who leave their birth state are disproportionately likely to move (back) there compared to others not born there (Davanzo, 1983), the problem of reverse causality resurfaces: a high share of "movers" into a state were actually born there. Second, names popular in one state are often popular in nearby states and people are more likely to move to nearby rather than distant states.

In Study 8 here I assess whether there is evidence of implicit egotism after the confounding factors mentioned above are accounted for. I do so by assessing whether people are more likely to *stay* in a state, conditional on having obtained their SSN there, if their name is similar to that state's; they are not. In Study 9, in turn, I examine whether there is direct evidence that the two confounding factors mentioned above play a role in JPSP1 original finding; there is. In fact, they are large enough to fully account for those findings.

Study 8 – People are not more likely to stay in a state with a name resembling their first name.

Method. I obtained the full SSDI dataset covering deaths up to 12/31/2006 (N=80 million). Considering that the social security system was enacted in the US in 1935, the analyses are conducted on the subset of people born after that year, seeking to maximize the share of people obtaining the SSN in their birth state (N = 5.97 million). [18]

To asses name popularity across states, I compare the relative share of people obtaining their SSN (henceforth "born") in a state similar to their first name to the share of all people "born" there. For example, I compare the percent of all Georges "born" in Georgia, to the percent of all people not named George "born" there. For each target name and state, then, the full post-1935 SSDI dataset is categorized into a 2x2 table.

To study if people are disproportionately likely to stay in a state with a name resembling theirs, I create a new 2x2 table for each name/state pair. Each table includes only people "born" in the state of interest. The table categorizes people as having the matching name or not, and as receiving their last benefit ("staying") there or not.

Results and Discussion

The results from the analyses just described are depicted in Figure 6. The gray bars show the ratio of actual over expected frequencies ($R_{A/E}$) for being "born" in the matching state, and the black bars for "staying" there.

Seven of the eight names considered by JPSP1 are more popular among people "born" in their corresponding state than expected by chance (i.e., than in all other states),

[18] Because the SSDI only has deceased individuals, the sample post-1935 is a small percentage of the overall sample.

with an overall $R_{A/E}$ = 1.27 ($\chi^2(1)$=232.9, p <.0001). This suggests that systematic differences in baby naming across states is in fact an important confound.

One may worry that this is in fact the result of implicit egotism; people might simply be migrating to states with names resembling their name before they obtain a SSN. We can use more recent data to address this possibility. In particular, beginning in 1986 parents can only include children as dependents for tax purposes if they submit their SSNs; this means that since 1986, at the latest, the state where people obtain their SSN does coincide with birth state for the vast majority of Americans.

The Social Security administration releases top-100 baby names by state per year, which we can use to see if more babies named similarly to a state were born in those states. Consulting the rankings for 1986 I found that four of the eight first names from Figure 6 were top-100 in at least one state. All four of these (George, Louis, Kenneth and Virginia) were more highly ranked in their matching state than nationwide. In addition, these first names were also disproportionately popular in nearby states. For example, the first name George was highly ranked not only in Georgia, but also in neighboring North and South Carolina and Tennessee. It was not a top-100 name in any of the more distant Northwestern nor Mountain states.

Returning to Figure 6. If people liked states with similar names as their own, then they should be more likely than others to stay there (i.e., $R_{A/E}$>1 for “staying”). Five of the eight black bars in Figure 6, however, show an $R_{A/E}$<1, and the overall effect is $R_{A/E}$ = .99 ($\chi^2(1)$ = .56, p =.45).

*** Figure 6***

How can we reconcile the black bars in Figure 6 with JPSP1 finding that people are more likely to *move* to states with names resembling their name? Why may people who already live in a state that matches their name be no more likely to stay there, but those living elsewhere be more likely to move in?

If the reason was that the movers' finding was spurious, as is argued in the "problems with existing findings" section above, then we should find such pattern in 'placebo' states too. More specifically, other states where a given first name was popular among people obtaining the SSN there, should also see a disproportionate share of movers into that state with that name. Study 9 shows this is the case.

Study 9 – Names popular among state natives are popular among state immigrants

Method

Using the same post-1935 SSDI dataset from Study 8, I computed the proportion of people "born" in every state with each of the eight names of interest ("natives"), and the proportion of people with those names among movers to a state ("immigrants"). For example, I computed the percentage of people "born" in each state in the US who were named George, and the percentage of all immigrants to each state in the US who were named George.

Results and discussion

As predicted by the "the movers' finding is spurious" story, states with more natives of one name have more immigrants of that same name. A simple average of the correlation for all eight names is r = .87, ranging between r = .57 for Louis and r = .96 for Virgil.

Figures 7 and 8 plot the two variables for Virginia (r = .87) and George (r = .77) respectively (the two names with the most observations in target states). Each dot in the figures correspond to a state, the x-axis is the share of natives of those states named Virginia or George, and the y-axis is the share of immigrants so named. For example, in Figure 8 we see that about 0.83% of people receiving the SSN in Georgia are named George, and about 0.82% of those moving to Georgia are.

*** Figure 7 & 8 ***

If implicit egotism was behind the movers' finding in JPSP1 we would expect that a surprising number of movers into a state with a name similar to that state. Instead we find that such share is just what we would expect given the share of people "born" in that state receiving that name (i.e., both George in Georgia and Virginia in Virginia fall very close to the best fitted line across all states). The evidence, in sum, strongly suggests that the movers result from JPSP1 is spurious.

Existing Finding 5: Last names & States

JPSP1 examine (Study 2) whether people are disproportionately likely to live in states with names resembling their last name, focusing on last names similar to the eight most populous states whose name is formed from a single word. They queried an online phonebook (the no longer available www.worldpages.com) for the frequency of listings of each last name across each of the eight states, obtaining an 8x8 table. Comparing expected and actual frequencies for this table they find that 19.9% of listings are found in states with names resembling the corresponding last name, compared to an expected rate of 16.6% ($R_{A/E}$=1.06, p <.001).

Potential problems with this finding

This finding has two major potential problems. The first is the source of the data. White pages often list businesses in them, and businesses are often named after the state in which they are located.

Because the phonebook used by JPSP1 is no longer available one cannot verify if this was a problem with their data, but as a proxy we can check currently available online white pages. I searched for "*last name*: *California*" "*State*: *California*" in three currently available ones (AnyWho.com, Whitepages.com, and Switchboard.com). In all three the first hit was a business rather than a person (Locksmith California, Bank California and Arthouse California respectively).

The second problem is that due to heterogeneity across states we don't have strong reasons to expect the distribution of last names to be the same across them. Study 10 here seeks to address both potential problems by comparing people "born" and "staying" across states, through analyses analogous to those of Study 8.

Study 10 – People are not more likely to <u>stay</u> in a state with a name resembling their last name.

Method and Results

The data source is the same post-1935 SSDI sample described above. I conducted analogous calculations to those described in Study 8, except that instead of focusing on eight first names similar to state names, here we focus on last names whose first three letters match the first three letters of the states used by JPSP1 (California, Texas, Florida, Illinois, Pennsylvania, Ohio, Michigan and Georgia). For each state here too we create a

2x2 table employing the full dataset for the "born" analyses, and using observations only of people "born" in a particular state for the "stay" analysis. These tables classify people as having the target last name or not (e.g., Cal_) and having obtained their SSN ("born") or receiving their last benefit ("staying") in the target state or not (e.g., California).

The results are shown in Figure 9. The gray bars indicate the ratio of actual over expected frequencies ($R_{A/E}$) for people "born" in a state, and the black bars for those "staying" in those states. For example, the post-1935 SSDI dataset includes 12,463 people with a last name beginning with Cal_, 9.7% of whom obtained their SSN in California. California as a whole has only 8.7% of the post-1935 SSDI sample leading to $R_{A/E}$ =1.12, p < .001. Of those Cal_ individuals "born" in California, 76.8% stayed there, compared to the baseline of 76.6% of all other individuals obtaining their SSN in California, leading to a $R_{A/E}$ =1.02, p =.79 for staying in California.

The figure is missing the bars for Texas and Ohio because there are too few observations to carry out the calculations. Of the remaining six states, four show a significantly greater than expected share of people with a similar last name obtaining the SSN ("born") there, and not a single one shows a significantly higher than expected share of people staying there. Illinois does have a high $R_{A/E}$ but it has just 21 people, if *two* fewer Ill_ last named individuals had stayed in Illinois, the $R_{A/E}$ would drop below 1.00. Overall, $R_{A/E}$ = 1.06, p <.001 for being "born" in a state, while $R_{A/E}$=1.00, p =.97 for staying there.

*** Figure 9 ***

Existing Finding 5: Last names & Towns

JPSP2 examine (Study 1) whether people are disproportionately likely to live in towns whose name contains their last name (e.g., Smiths living in Smithville). To test this they queried the SSDI dataset and compared, for the thirty most common last names in the United States, the overall fraction of individuals with that last name in the country, with the share of people with that last name in towns containing that last name in their name.

For example, they compare the proportion of Williams in the United States, P=0.57%, to the proportion of Williams living in towns that contain "Williams" in their name (e.g., Williamsburg, VA), P=0.64%. They find greater shares among the matching towns for 27 of the 30 last names considered (overall $R_{A/E}$=1.41).

Potential problems with this finding

This last-name-town finding has two potential problems: an ethnicity confound and reverse causality. The vast majority of towns considered are very small towns often with less than 1000 inhabitants. Ethnic minorities (e.g., Asians, Hispanics, Jews, etc.) tend to live in large cities and in a few coastal states, as a consequence, small towns in mid-America will have higher proportions of all non-ethnic last names than the country as a whole, independently of implicit egotism effects.

As serious as the ethnic confound problem may be, I do not address it empirically here because the problem of reverse causality swamps it. As Study 11 here will show, a staggering number of towns containing a last name in their name were founded by individuals with such last names.

Study 11 – Towns are named after their founders

Method and Results

To examine the extent to which the set of towns used by JPSP2 had the problem of reverse causality alluded to above, I computed for each individual town the share of people who had a matching last name in the SSDI dataset (e.g., the percentage of the people receiving their last social security benefit in Millersburg, PA whose last name was Miller). A research assistant then searched online for information on who founded the 100 towns with the highest percentage of people with the matching last name.

The information on towns' founders was not available for all towns and hence the RA went beyond the top-100 originally asked for. In all he searched for 143 towns and obtained the identity of the founder for 95 of them. A striking 72% of these 95 towns were founded by someone with the matching last name providing a plausible reverse causality explanation for this finding.

Existing Finding 6: Resident and Street Names

The second set of analyses in JPSP2 examine whether people are disproportionately likely to live on streets with names that include their last name. They studied this possibility for the six most common US last names (in 1990): Smith, Johnson, Williams, Jones, Brown and Davis.

Concerned about reverse causality, e.g., Mr. Smith living in Smith Rd. because he (or an ascendant of his) changed the name of the street to their last name, JPSP2 examine a second set of six last names for which they presume this reverse causality is less plausible. In particular, they consider two last names than are used in street names to describe nearby places: Hill & Park, two that describe the type of road involved: Lane & Street, and two that are of historic significance: Washington & Jefferson.

They obtain the frequency of people living on different streets are obtained from online phonebook listings across the entire United States and analyzed it in pairs. The analyses test, for example, whether a greater share of people last named Washington, rather than Jefferson, live on Washington Street (rather than Jefferson street). Actual frequencies proved greater than expected for all six pairs of last names, but their ratio was much greater for the first three pairs, $R_{A/E} = 1.59$, than for the next three, $R_{A/E}=1.07$.

Potential problems with this finding

The risk of reverse causality is quite real in the context of streets as it is not unusual for residents to be able to determine the name of their own street, especially in rural areas. In Lodi, NY, for example, most streets change name at each intersection, and the name of the street within each block corresponds, as was learned from phone conversations with the *Lodi Historical Society*, to the original owner of the property.

As mentioned above, JPSP2's concern about reverse causality lead them to their second set of six streets, which contained names that were *less* likely to be chosen by their residents than the first set. The fact that in the latter they obtain an effect about 1/8 the size of in the former suggests that reverse causality indeed plays a large role in their results.

While most streets containing the words Hill or Park in their names probably do so in reference to a nearby hill or park, there is no reason to suspect that residents last named Park, Hill, Street or Lane are less likely to change the name of the street they live on to their own name (in fact, it may be *easier* for them to do so).

Concerns about geographic confounds are also high in this setting because factors that influence variation in people's last name across areas also influence variation on streets' names across areas. For example, Cesar Chavez is a Hispanic historic figure who has a disproportionate share of streets named after him in states with large Hispanic populations (California, Texas, New Mexico and Arizona account for 80% of Cesar Chavez streets in the US). [19]

How ethnically diverse are the last name pairs used by JPSP2? From moderate to extreme. Census figures indicate that within any given pair, the relative frequency of African Americans differs between 30%-50% between the two last names (e.g., 22% of Smiths are black compared to 34% of Johnsons). The most diverse pair is Park (66% Asian) & Hill (<1% Asian).

In Study 12 I attempt to examine the impact of people's names on the streets they live in net of any reverse causality and ethnic confound effects by focusing on people's first names, and on last names which cannot possibly have influenced the name of the streets.

Study 12 – No name effect on streets from first names or geographic designators

Method.

To rule out reverse causality we could, as was done in Study 3, use similar rather than identical last names (e.g., looking at whether people last named Smith<u>ers</u> live on Smith St.). Because such last name variations are too infrequent, I use instead similar

[19] Source: http://www.melissadata.com/lookups/

first names, assessing for example, if men named William disproportionately live on Williams Ave.[20]

First names may address both concerns described above. On the one hand, they are not *identical* to the streets' names, alleviating reverse causality concerns. On the other, variation in ethnicity of first names should be different than that of the similar last names, reducing the ethnic link between peoples' and streets' names.

In addition to first names, I examine last names that begin with streets' orientation designators (West, East, North or South); such designator are not chosen on a street by street basis, but rather, determined by the street's actual orientation, further addressing concerns of reverse causality.

Seeking reliable address data I obtained the voter registration file for the entire state of New York. It contains the first and last name and the current mailing address of every registered voter in such state (N=12.8 million).

If two or more observations had identical last name and full address, only one of them was used in the analyses, reducing sample size to N = 8.9 million. Seeking to stay as close as possible to the original design, first names were chosen to closely resemble the last names used in JPSP2. These were John, William, Jon, David and Jeff, in lieu of the last names Johnson, Williams, Jones, Davis and Jefferson. For Washington and Jefferson I also employed the first name of the respective historic figures (George and

[20] For example, there are 56,035 people in the sample with a last name beginning with "Smit_"; 98.6% of them are last named Smith.

Thomas).[21] For West, East, North and South I include all last names that begin with those strings of letters (e.g., John Westfield, Oliver North, etc.).

Results.

I begin by attempting to replicate the results from JPSP2 on the New York data, that is, examining if the 12 last names used by JPSP2 are more commonly observed living in streets with those last names than would be expected by chance. Ten of the twelve last names have a ratio of actual over expected frequency ($R_{A/E}$) greater than 1 for the matching street, leading to an overall $R_{A/E}$=1.32. For the first set of six last names the estimated effect is $R_{A/E}$=2.3 (much *larger* than the $R_{A/E}$ = 1.59 from the original demonstration). For the second set of names the $R_{A/E}$ is dramatically lower also with the New York data, $R_{A/E}$=1.08. These results suggest that any differences between the conclusions from the new and old analyses do not stem from data differences.

The results for the new analyses are reported in Figure 10. Eight of the eleven name-street combinations have a $R_{A/E}$ *smaller* than 1. The largest $R_{A/E}$ is for last names beginning with North_, but the numbers are quite small (the expected frequency is 4.9 the actual is 8). Aggregating across all combinations $R_{A/E}$=.92, $p < .001$, indicating, if anything, a reversed pattern.

*** Figure 10***

[21] For Williams I include William, Will, Willie and Willy, for Jones, Jon and Jonathan. For Jefferson any first name whose first four letters are Jeff_ (there were many variants of Jeffery in the data). The heuristic to decide which names to use was based on the objective to obtain a high number of observations. I did not compare the results obtained including different subsets of first names.

Existing Finding VIII: Birth Day Numbers

JPSP1 examine (Study 6) whether people are more likely to live in a town whose name contains the number of their birthday. More specifically, they tested whether people born on the same numbered day and month (e.g., February 2^{nd}) were more likely to live in a town containing that number (e.g., Two Rivers) than in a town containing a different number (e.g., Four Corners).

There are few towns with the number one or nine on them, so they focus on birthdays and towns with names ranging between two and eight. They employ the SSDI dataset (described above) and conduct the analyses on a subsample that includes only people born on those seven dates and dying in towns containing a number between two and eight.

Overall they find that of the 485 people in their sample, 94 died in a town matching their birthday, compared to the 70 that would be expected if birthday was independent of town name; this is equivalent to a ratio of actual over expected frequency of $R_{A/E}$=1.33, $p < .001$.

Potential Problems with this finding

The main concern about this demonstration is reliability. The implicit egotism effect of birthday numbers is quite small and unreliable *in the lab*; its first demonstration, in fact, found that the effect was much smaller and not significant for numbers lower than twelve (Kitayama & Karasawa, 1997).

Starting from such a weak effect it would seem that a rather large dataset would be required for a field demonstration. Worried that the finding may be the consequence of sampling error, I set out to replicate it. I attempted four independent replications without success.

The simplest replication uses the same SSDI data and analyses employed in JPSP1 but asking whether day of birthday alone, rather than day and month combined, predicts the town in which people live (e.g., assessing the relationship between being born on the second day of *any* month and the likelihood living in a town like Two Rivers).

This operationalization of implicit egotism for birthdays is closer to the existing lab evidence which has used month and day number as *separate* predictors. Such operationalization is hence a natural starting point for any field study of implicit egotism with birthday numbers.

The other three replications used the New York voter registration dataset described above. They examined the relationship between (matching) birthday numbers on address number, street number, and apartment number; asking, for example, is someone born on February 2nd more likely to live on 2nd avenue or 2 Elm Street or Apartment 2?

One would expect, ex-ante, that the latter three replications would be a superior testing ground for implicit egotism over birthday numbers for three reasons. First, it is much more common to have the option to move to another apartment or street than to another town, and very few towns have numbers in them while all addresses do. Second, addresses have numerals in them, while town names have the number represented with words. Lastly, and related to the first point, the dataset for addresses is much larger and hence has more power that the one for town names.

Study 13 – No effect of just day of birthday on town

Method and Results. The entire SSDI dataset was used for this analyses, excluding only observations that lacked information on birthday or zip code of last benefit received (N=80.4 Million). Overall there were 967 individuals who died in a town that contained the day of their birthday (e.g., born on January 2nd, February 2nd or March 2nd, etc., died in Two Oaks), compared to the 954.5 that would be expected by chance, implying a ratio of actual over expected frequency of $R_{A/E} = 1.01$, $p = .46$. JPSP1 birthday result, in short, does not replicate using the birthday operationalization from lab studies.

Study 14 – No effect of birthday number on street, address, or apartment Number

Method and Results. The data consist of the New York voter registration file described in Study 12 (N=12.8 million). Three separate analyses are performed, one for street names, one for address numbers and one for apartment numbers. In each of them an address is considered as matching only if it contains a single digit and the digit matches the birthday day and month of the individual (e.g., for someone born on April 4th, the streets analyses would consider 4th Avenue a match, the address analyses 4 Broadway Ave, and the apartment analyses Apt. 4). I replicated the analyses considering also multi-digit matches obtaining identical results (e.g., 44th avenue, 4444 Broadway Ave, and Apt 444 would all be considered matches).

Each analysis creates eight 2x2 tables, where every individual is categorized as born on a given date or not, and as living in the matching address or not.

The results, are presented in Figure 11. *n* indicates the number of matching observations (e.g., there are 1827 people living at an address number that matches their birthday day and month), and *N* the number of total observations. Note that *n* is between 8 and 20 times larger than in the original studies, and *N* between 10,000 and 30,000 times larger.

None of the replications show any evidence of implicit egotism. Not one of the 21 individual $\chi^2(1)$ tests is significant at the 5% level, two are significant at the 10% level, one is in the direction implied by implicit egotism (March 3rd and Third Avenue) and one is in the opposite direction (February 2nd, and Second Avenue). The overall $R_{A/E}$'s are .98, 1.01 and 1.01 for address, apartment and street, respectively, none of them significantly different from 1 (*p*'s .74, .73 and .50 respectively).

In sum, the only result that's free of confounds across the three implicit egotism papers considered, does not replicate in settings where ex-ante it could be argued that implicit egotism should be more likely to be observed and with samples orders of magnitude larger. While it is not possible to directly test if the original finding is due to sampling error, the evidence suggests there is more than reasonable doubt that it is.

Other Field Evidence of Implicit Egotism

The previous eight sections suggest that all existing evidence of implicit egotism in marriage, occupation and moving decisions is spurious. In this section I briefly discuss evidence of implicit egotism in other real-life decisions.

Last name Initial and employer (Anseel & Duyck, 2008)

Anseel and Duyck analyze a sample containing a third of all full time employees in Belgium (N=528,007). The dataset contains the first three letters of employees' last names and of the name of the company they work for. The authors assess whether people are disproportionately likely to work for companies with which they share their initial, finding evidence of this pattern for every single letter; the overall implied ratio of actual over expected frequency is $R_{A/E} = 1.13$.

There are several potential problems with this finding. Two particularly compelling ones are reverse causality and ethnic/language confounds. In terms of the former, family firms are often named with the family last name and employ family members. In addition, while the authors exclude self-employed individuals from their analyses, they would still include people who belong to a firm named after themselves if other people work there also, such as law firms with more than one partner.

In terms of the ethnic/language confound. Belgium has two main areas, Flanders in the north, where people speak Flemish, and Wallonia in the south, where they speak French. Companies and people in Flanders will tend to have Flemish names while people and companies in Wallonia will have French ones, leading to a spurious initial-matching effect because Flemish and French names have different distributions of initials. [22]

Through email communications I suggested they attempt replicating their findings excluding people who share all three letters with the company (avoiding reverse causality), and using as controls letter combinations which are highly correlated with each other in their choice of employers (as in Study 5 here with first-names). As of now, I do not believe such analyses have been conducted.

Initials and Performance (Nelson & Simmons, 2007)

[22] I obtained a list of the top-100 last names in both regions, and conducted difference of proportions tests for each initial across them, finding that the share of people with each initial differs significantly between regions (biggest p-value = 1.99×10^{-8}. Source http://www.eupedia.com/belgium/belgian_surnames.shtml

This paper includes four field studies. One shows that professional baseball players whose name starts with a K are more likely to strike out (a negative outcome often summarized in baseball statistics with a K).[23] The other three field studies show that people with first names starting with a C or D obtain worse academic outcomes than other students. Their fifth study is an experiment, alluded to in the introduction of this paper.

The academic performance studies do not share the concerns put forward elsewhere in this paper, as they include controls for age, country of origin and race. In addition, in one of their studies they find that only students with a C or D name who also like their own initial obtain a lower GPA than other students.

In terms of reverse causality. It is possible that high GPA parents are less likely to choose names starting with C or D for their children, and that high GPA parents have high GPA children (though it is unclear why this effect would be moderated by children's liking of their initials).

Donations and initials (Chandler, Griffin, & Sorensen, 2008)

This paper studies donations to the Red Cross. It examines whether people who share an initial with a hurricane name are more likely to donate to its victims than other people are. To test this prediction they compare the share of donors who share an initial with a hurricane just before and after the hurricane struck. For every hurricane considered

[23] McCullough & McWilliams (2010) find that if the analyses are run weighting by number of at-bats the effect is heavily attenuated. This is due to the fact that players that bat less per season are driving the effect.

they find such share increases post-hurricane. The overall effect is a ratio of actual over expected frequency of $R_{A/E}$=1.24, p =.008.

Given that the name a particular hurricane receives is random for our purposes (that is, uncorrelated with donor characteristics), none of the confounding factors discussed elsewhere, nor reverse causality, are plausible explanations for these findings.

General Discussion

When do psychological taste shifters influence real decisions?

This paper re-evaluates evidence that seemed to show that implicit egotism, the liking of things connected to the self, can influence marriage, occupation, and moving decisions, finding that all existing evidence appears to be spurious.

If implicit egotism is a real phenomenon, as the lab evidence suggests, why does it not (detectably) influence these decisions? One tempting explanation is that the stakes involved are large and that psychological influences impact only small stakes decisions. Many skeptics of laboratory research make this argument to broadly dismiss the relevance of experimental evidence.

We know, however, that big stakes decisions *are* influenced by small psychological factors. Housing and retirement decisions are the biggest financial decisions people make in their lifetimes, and prior research has shown that house buying (Simonsohn & Loewenstein, 2006) and selling (Genesove & Mayer, 2001) are influenced by contract effects and loss aversion respectively, and that participation in retirement savings program (Madrian & Shea, 2001) and the amount of money saved (Thaler & Benartzi, 2004) are influenced by defaults and simple commitment devices respectively.

Why may implicit egotism not matter for marriage, occupational and location decisions then? One explanation might be the effect size of implicit egotism. The field evidence of psychology affecting housing and investment decisions focused on psychological mechanisms that in the lab have very large and robust effects; it is possible that implicit egotism is simply too subtle an effect to be detectable in noisier environments.

Another explanation relies on what we might refer to as marginal vs. overall stakes. While home buying is a large stakes decision, which specific home to buy, e.g., whether the extra bedroom is worth $50,000, need not be. People may be nearly indifferent between options, or have difficulty determining which they ultimately prefer; a modest psychological effect can tip the balance one way or the other.

The decisions studied by JPSP1-3 do not, arguably, have small marginal stakes. Few people are nearly indifferent or ambivalent between the career, job and spouse they chose and their second most preferred option (and fewer still between a spouse, career and location which does vs. does not resemble their name). The number of Georges who are nearly indifferent or are ambivalent between living in Georgia and another state may be too small for the impact of implicit egotism in their choice to be detectable.

The paper with the most convincing evidence of implicit egotism in the field, in contrast, (Chandler, et al., 2008), studies a decision that while important (giving charity to needy victims), involves decisions over which probably people *are* nearly indifferent (e.g., between giving to Katrina victims or to victims of other disasters). Even if implicit egotism has a small effect on preferences, it could have a dramatic effect on aggregate charitable contributions.

Future work should strive to identify other domains where decisions in the field have small marginal stakes (though ideally large total stakes) and examine if implicit egotism plays a role in those decisions.

On the usage of quantity of evidence as evidence

The results from this paper also speak to a common argument employed to defend the validity of explanations put forward in multi-experiment papers. It is often proposed that since the authors' main thesis is the only parsimonious explanation for *all* studies in a given paper, one should favor it over multiple partial explanations that can explain subsets of studies only.

Implicit egotism is the only parsimonious explanation for all 16 studies conducted by JPSP1-3, and yet it is not directly supported by any of them. Sheer number of studies is not a sensible metric on which to judge the correctness of a thesis, it is merely a byproduct of the path chosen to investigate it.

Concluding remark. JPSP1-3 took the Name Letter Effect evidence and arrived at fascinating predictions which they tested on important domains. With great ingenuity they designed a large number of provocative studies. More often than not, they themselves identified the potential confounds which the evidence in this paper suggests are in fact behind their findings. Had Pelham et al. not so openly shared their thinking, data sources, and results, this paper would probably not have been written. For what is worth, I personally do believe in the psychological reality of implicit egotism, and also believe that it may influence real life decisions (over which people exhibit near indifference). Any study that convincingly demonstrates this, does so on the shoulders of Pelham, Jones, Carvallo, DeHart and Mirenberg,

Table 1. Overview of new studies

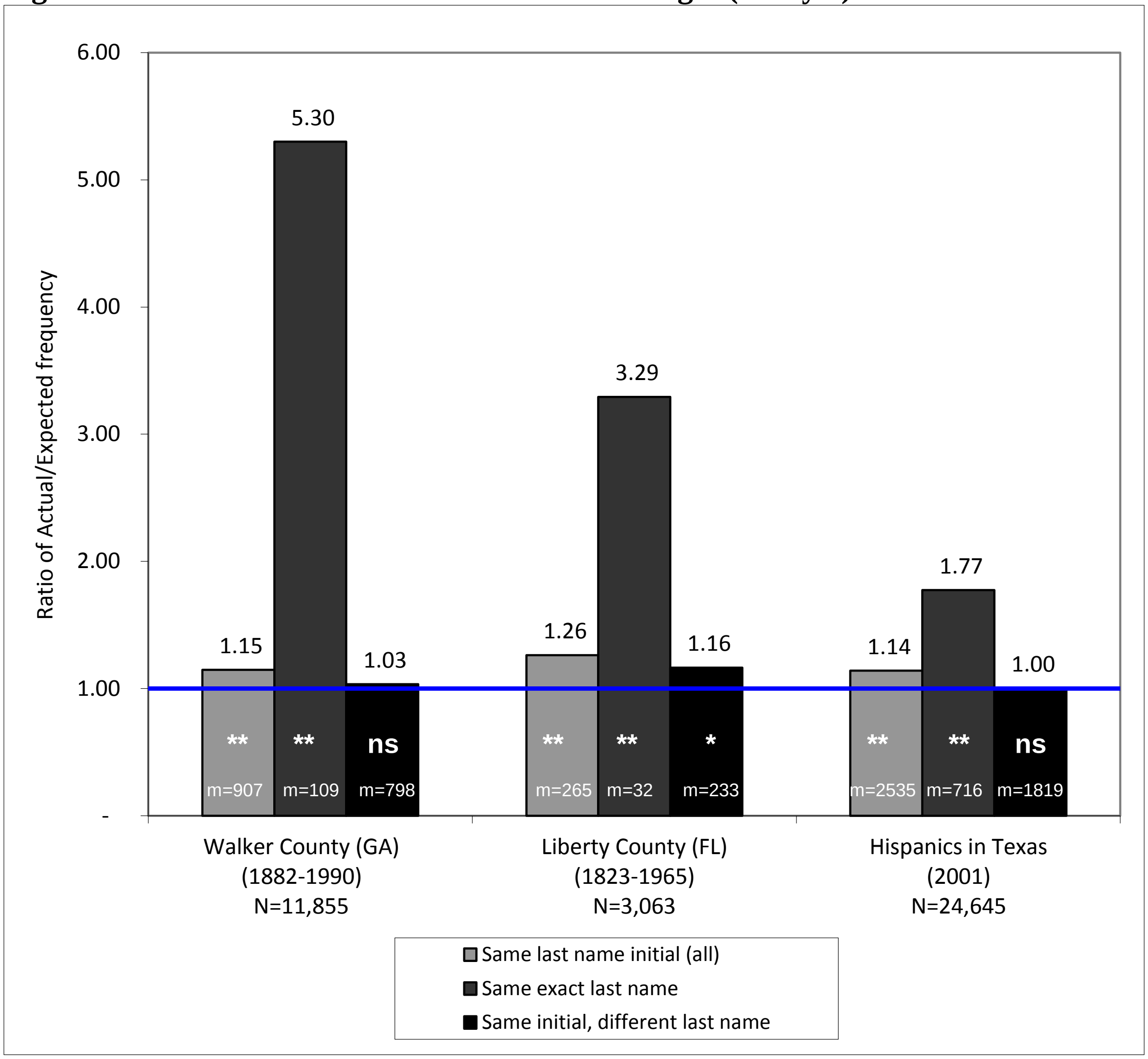
Original finding	Between		Main problem with original finding	New Study	New finding	Figures
I	last lame	Marriage	**Ethnic confound.** Ethnic groups more likely to marry within; ethnic groups have different distribution of last names and initials **Reverse causality** Some brides change last lame to husband-to-be's before marriage.	Study 1	The 'same-last-name-initial' marriage effect is an 'exact-same-last-lame' effect	Figure 1
				Study 2	Few ethnic last lames account for significant share of same-last-lame marriages	Table 2
				Study 3	Very similar last lames do not disproportionately marry each other	Figure 2
				Study 4	Remarriages to same previously divorced spouse are frequent	Figure 3
II	First name	Marriage	**Cohort confound** Popularity of similar male & female first names changes together over time; people marry people of similar age.	Study 5	The first-name-effect on marriage disappears with control names that make similar spouse selections.	Figure 4
III	First name	Occupation	**Cohort (& probably socioeconomic status) confound** Original studies chose control names based on overal Census frequency, ignoring age distribution and socioeconomic background of people with such names.	Study 6	More Dennis than Walter dentists, more Dennis than Walter lawyers	Figure 5
				Study 7	Georges no more likely to be in Geology than any other science	None
IV	First name	State	**Geographic confound, reverse causality** More babies born in Georgia, *and in surrounding states,* named Georgia.	Study 8	People are not more likely to stay in a state with a name resembling their first name.	Figure 6
				Study 9	Names popular among state natives are popular among state immigrants	Figures 7 & 8
V	last lame	State	**Geographic/ethnic confound** More babies born in California with last lame Cali_.	Study 10	People are not more likely to stay in state with a name resembling their last lame	Figure 9
VI	last lame	Town	**Reverse causality**	Study 11	Towns are named after their founders	None
VII	last lame	Street	**Rerverse causality** Streets are often named after their residents (especially in rural areas). **Ethnic confound** Areas with more minorities have more streets named after minority last lames.	Study 12	No effect of first names or last lames with geographic designators on street choice	Figure 10
VIII	Birthday	Town	**Sampling error** Fails to replicate in four much larger samples where ex-ante effect is more likely.	Study 13	No effect of just day of birthday on town	None
				Study 14	No effect of birthday number on street, address, or apartment number	Figure 11

Table 2. Just four (ethnic) last names account for 18% of all same-last-name marriages in Texas (Study 2)

(1) Lastname	(2) Origin	(3) # obs.	(4) Marrying same last name bride	(5) Percent of sample	(6) Ratio of actual over expected	(7) Percentage of total same lastname effect
Patel	India	52	63%	.03%	2,378	4.2%
Nguyen	Vietnam	137	31%	.07%	444	11.1%
Kim	Korea	11	17%	.01%	3,047	0.9%
Tran	Vietnam	17	11%	.01%	1,258	1.4%
OVERALL		217				17.60%

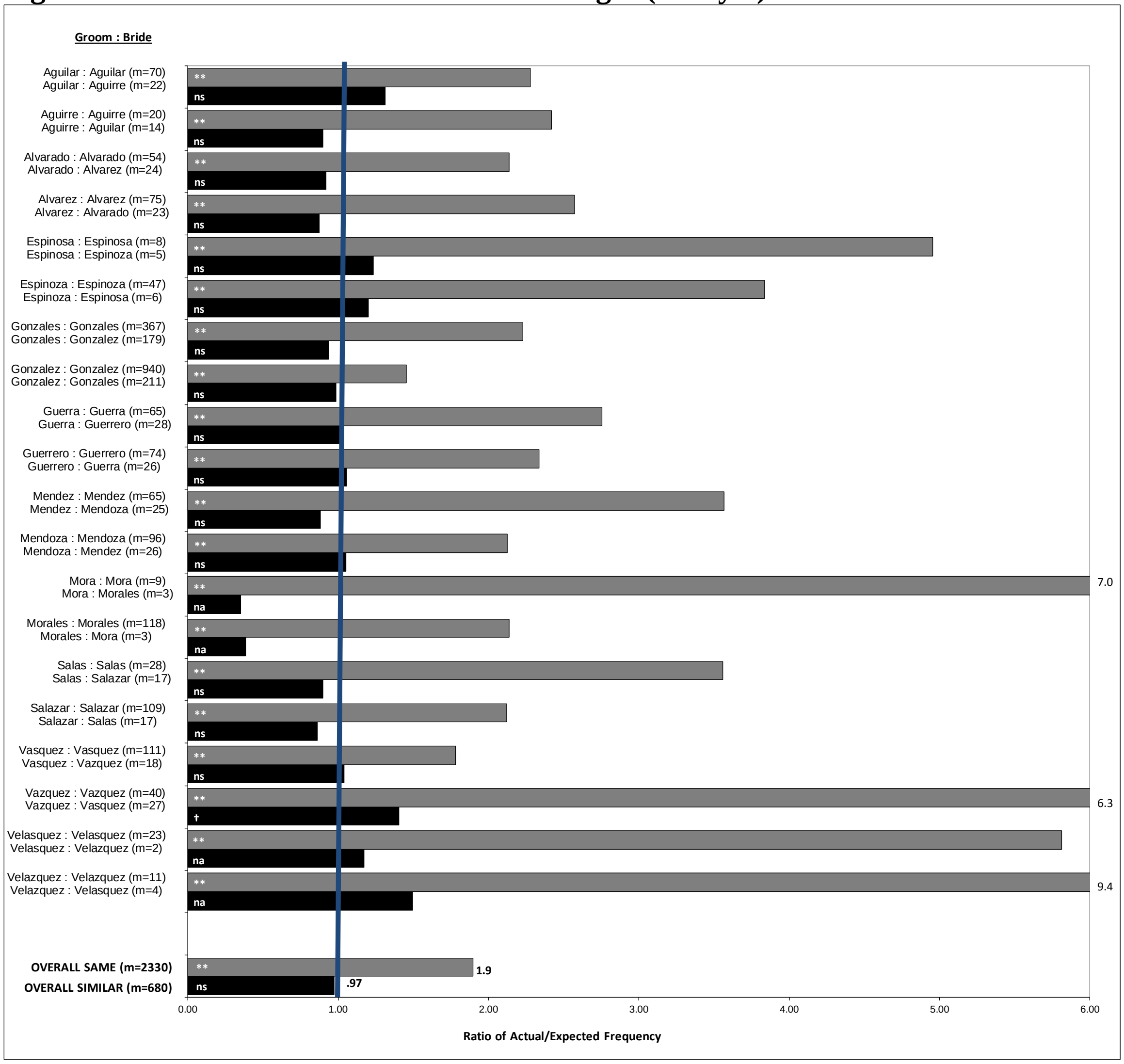
Note: These are the four last names with the highest percentage of same-last-name marriage.
Their 217 marriages account for 17.6% of all same-last-name marriages in Texas 2001 (N=195,030)

Figure 1. Same-last-name and same-initial marriages (Study 1)



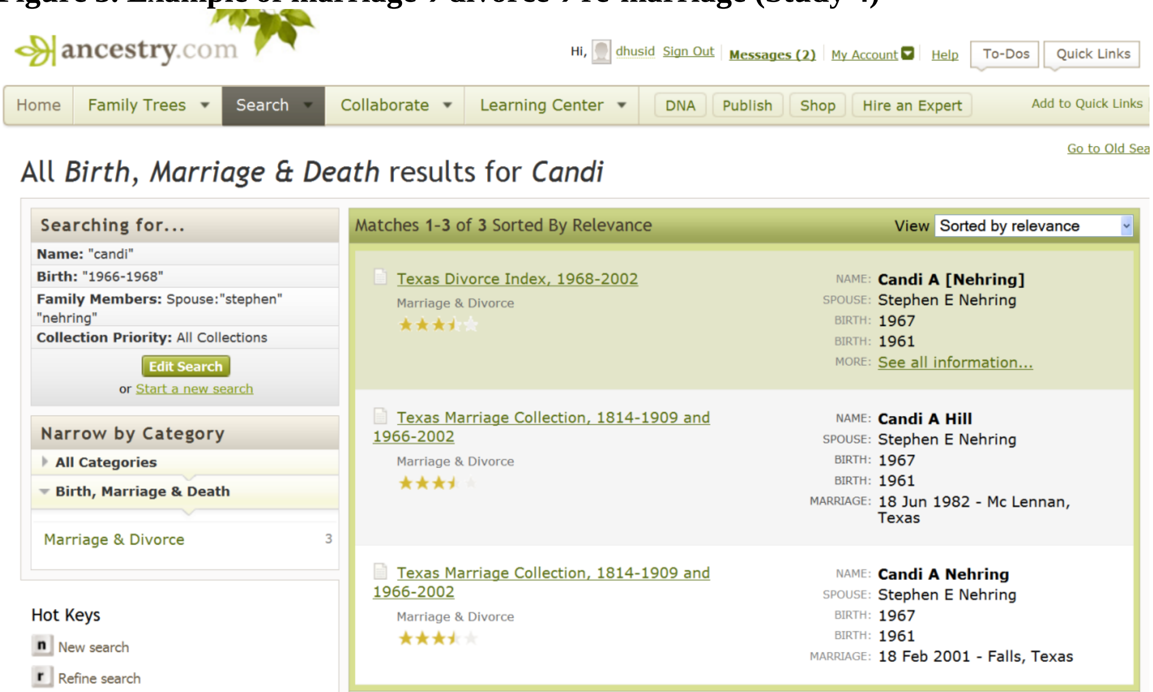
Source: marriage licenses from the respective locations.
Note. Expected marriage rates are obtained by multiplying the base rates of each initial/last name for grooms & brides, and then adding up all the resulting products. Significance is assessed with $\chi^2(1)$ comparing aggregate actual vs. expected rates.
**: significant at the 1% level. ns indicates p >.10.
N is the total number of marriages in the sample, *m* of those exhibiting the name-similarity-effect in each bar. For example, the Walker country sample contains 11,855 marriages, of which 907 are between people sharing a last name initial.

Figure 2. Same and similar last name marriages (Study 3)



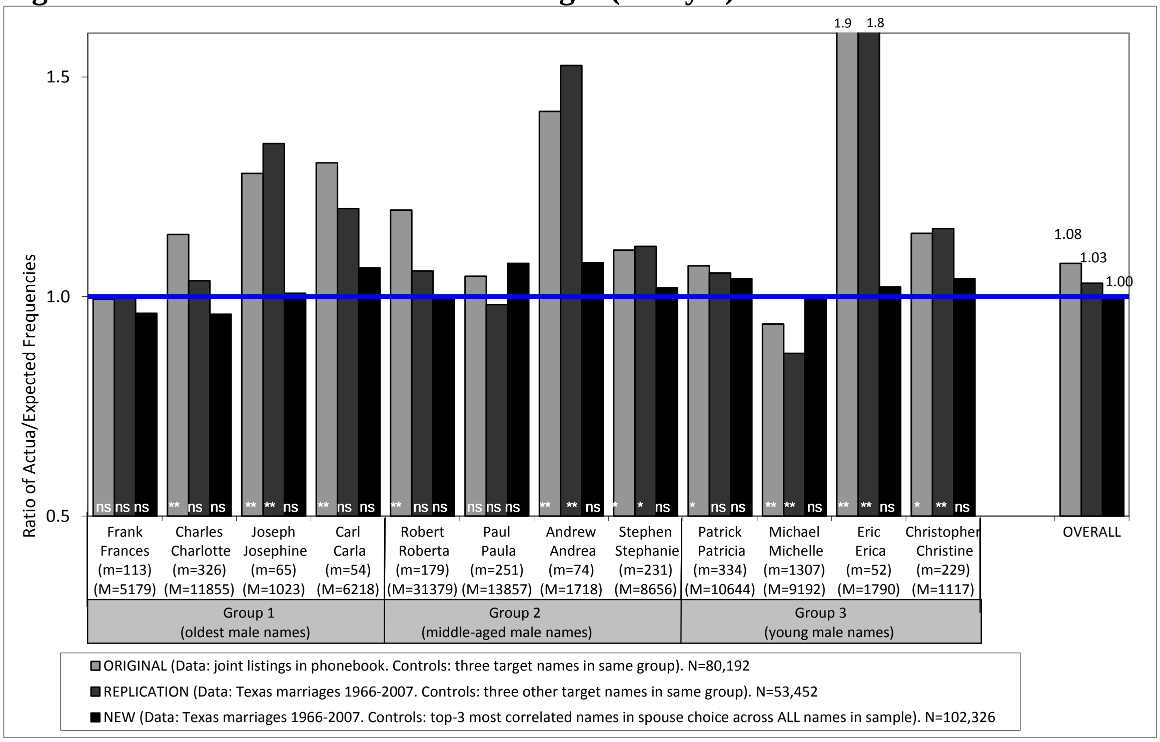
Source: Marriages in Texas (1966-2007) between a groom and a bride with one of the 20 last names in the figure (N=13,335).
Notes: Ratios of actual over expected frequencies ($R_{A/E}$), and their significance, are obtained with 2x2 tables that categorize marriages as including a target male name or not, and as including a target female name or not. Each of the 20 same-last-name tables includes the entire sample. Each of the 20 similar-last name tables includes the entire sample except for same-last-name marriages (see footnote 11). $R_{A/E}$'s > 6 are indicated with a number to the right of censored bars.
**, *, † significant at the 1%,5% and 10% level respectively. *ns*: p-value >.1. *na*: less than 5 expected observations. *m* is the number of matching marriages (e.g., there are 70 marriages between two people last named Aguilar).

Figure 3. Example of marriage→divorce→re-marriage (Study 4)



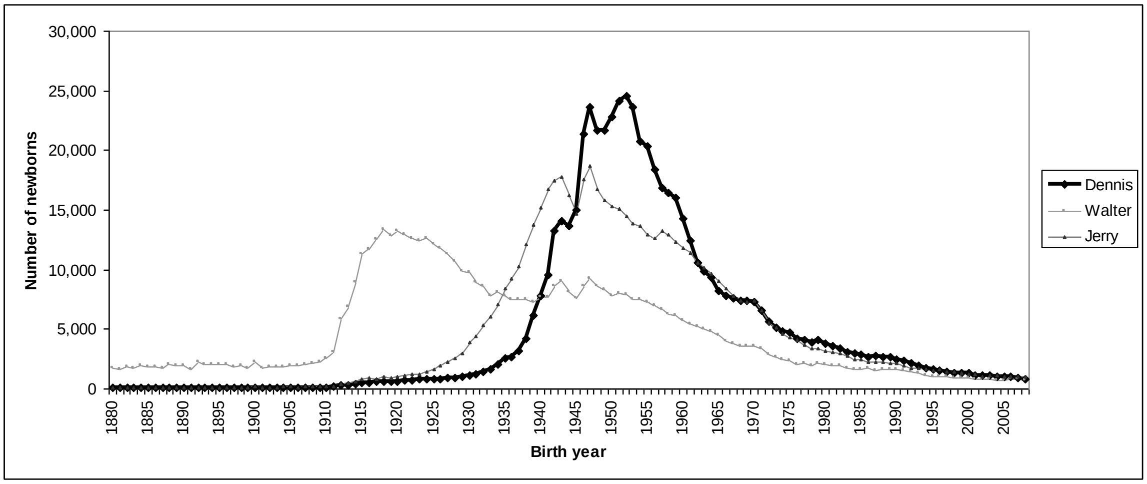
Printed-screen from Ancestry.com. It exemplifies one of the 68 cases of multiple marriage/divorce records for a bride with a given first name, middle name and birth year, and a man with the same name last name and birth year as those from a same-last-name couple marrying in 2001.

Figure 4. Similar first names and marriages (Study 5)



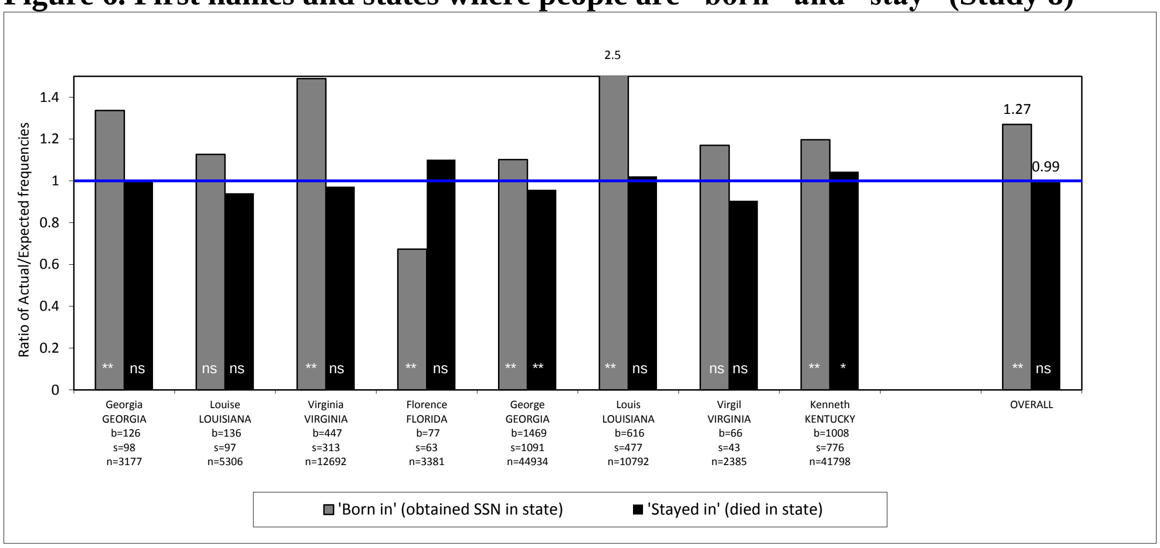
Notes: Groups (1-3) are used for the Original and Replication analyses only. They combine, as in JPSP3, names based on the difference in popularity between the 1960 and 1920 Census of the corresponding male name.
Ratios of actual over expected frequencies ($R_{A/E}$), and their significance, are obtained with 2x2 tables that categorize marriages as including the target male name or not and as including the target female name or not.
The 2x2 tables for the Original and Replication analyses include marriages among names within a group of names.
The 2x2 tables for the New analyses include marriages between target names and their three new controls names.
m the number of matching-name marriages for the name pair in the New dataset, and *M* the total number of marriages in the corresponding set of eight names (e.g., there are 5179 marriages among Franks, Frances, and the top-3 most correlated names with them, of which 113 are between a Frank and a Frances). See text for discussion of overall effect.
The sum of all *M*s is greater than *N* because some control names are used more than once.
**, * significant at the 1% and 5% level respectively. *ns*: p-value >.1.

Figure 5. Popularity of first names used in dentists' study (Study 6)



Source: Social Security administration

Figure 6. First names and states where people are "born" and "stay" (Study 8)



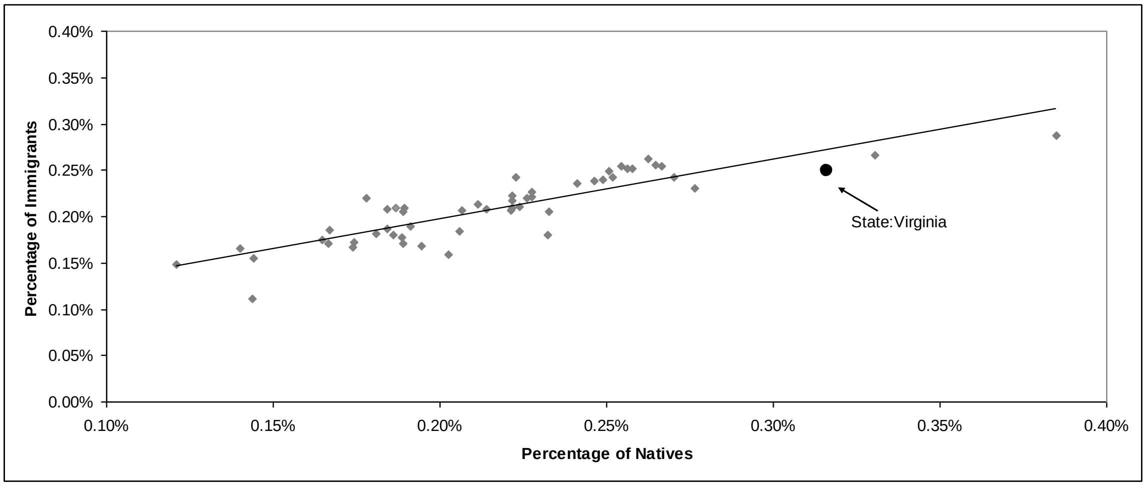
Source: Social Security Death Index; subsample of people born after 1935 (N=5.8 million).

Notes: SSN stands for Social Security Number. Because people obtain SSN after birth, they may obtain it in a state other than their birth state, hence "born". Ratio of actual over expected ($R_{A/E}$) and their significance are obtained from 2x2 tables that classify people as in the target state or not, and as having the target name or not. For the gray bars the entire sample is used in each table, for the black bars the subset of people "born" in each state is.

**, * significant at the 1% and 5% level respectively. *ns*: p-value >.1.

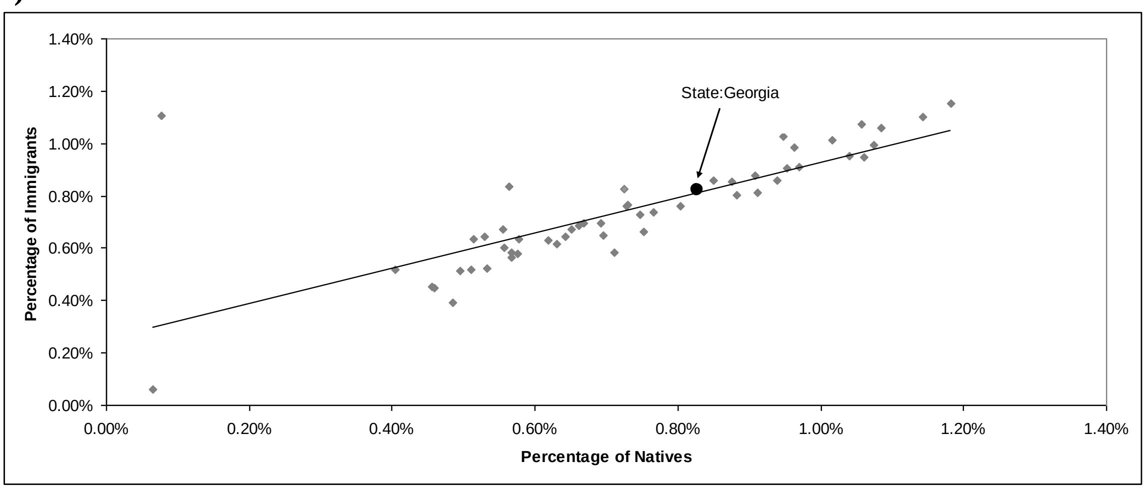
b is the number of matching observations "born" in each state, *s* the number of those "staying" there and *n* the total number of people with that first name in the sample (e.g., there are 3,177 Georgias in the sample, 126 of them were born in the state of Georgia, and 98 of those stayed in Georgia).

Figure 7. Share of people "born" and "moving" to each state, named Virginia (Study 9)



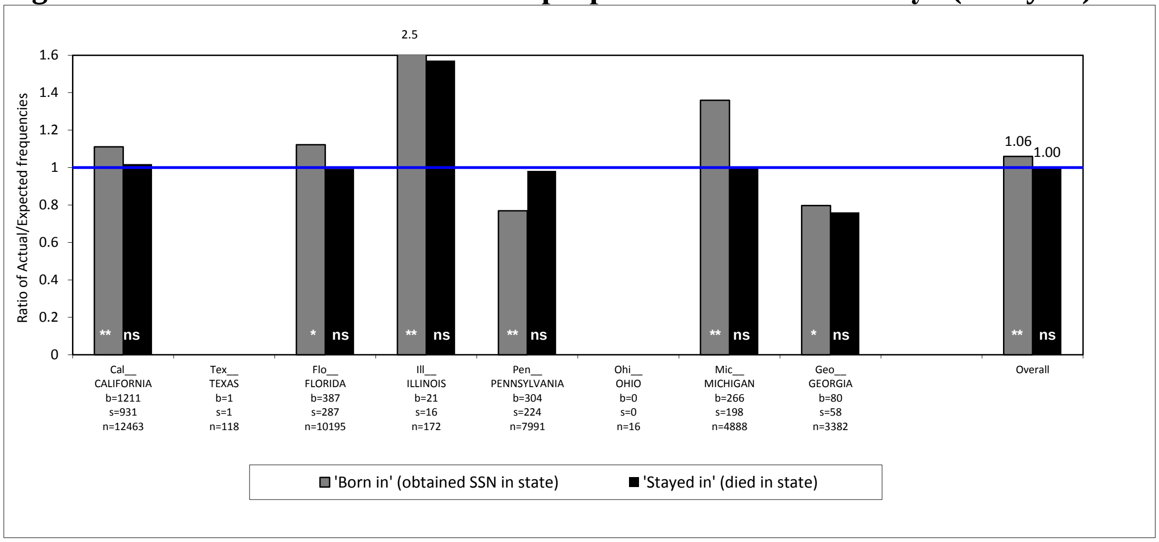
Source: Social Security Death Index; subsample of people born after 1935 (N=5.97 million).
Notes: Each dot in the figure is a state. The x-axis indicates the percentage of people obtaining the Social Security Number ("born") in that state whose first name is Virginia. The y-axis the percentage of people "born" elsewhere but receiving their last benefit ("moving to") that state, whose first name is Virginia.

Figure 8. Share of people "born" and "moving" to each state, named George (Study 9)



Source: Social Security Death Index; subsample of people born after 1935 (N=5.97 million).
Notes: Each dot in the figure is a state. The x-axis indicates the percentage of people obtaining the Social Security Number ("born") in that state whose first name is George. The y-axis the percentage of people "born" elsewhere but receiving their last benefit ("moving to") that state, whose first name is George.

Figure 9. Last names and states where people are "born" and "stay" (Study 10)



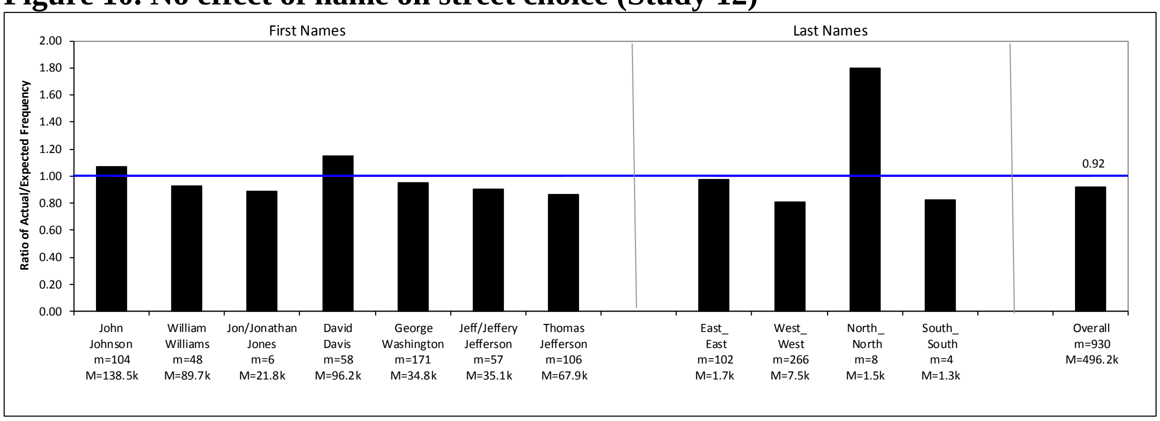
Notes: SSN stands for Social Security Number. Because people obtain SSN after birth, they may obtain it in a state other than their birth state, hence "born". Ratio of actual over expected frequency ($R_{A/E}$) and their significance are obtained from 2x2 tables that classify people as in the target state or not, and as having the target name or not. For the gray bars the entire sample is used in each table, for the black bars the subset of people "born" in each state is.

**, * significant at the 1% and 5% level respectively. *ns*: p-value >.1. All significance tests computed with respect to null of $R_{A/E}$=1

b is the number of matching observations "born" in each state, *s* the number of those "staying" there and *N* the total number of people with that lastname in the sample (e.g., there are 12,463 with a Cal_ lastname in the sample, 1211 of them were born in California, and 931 of those stayed there).

Sample: Social Security Death Index of people born after 1935 (N=5.8 million).

Figure 10. No effect of name on street choice (Study 12)



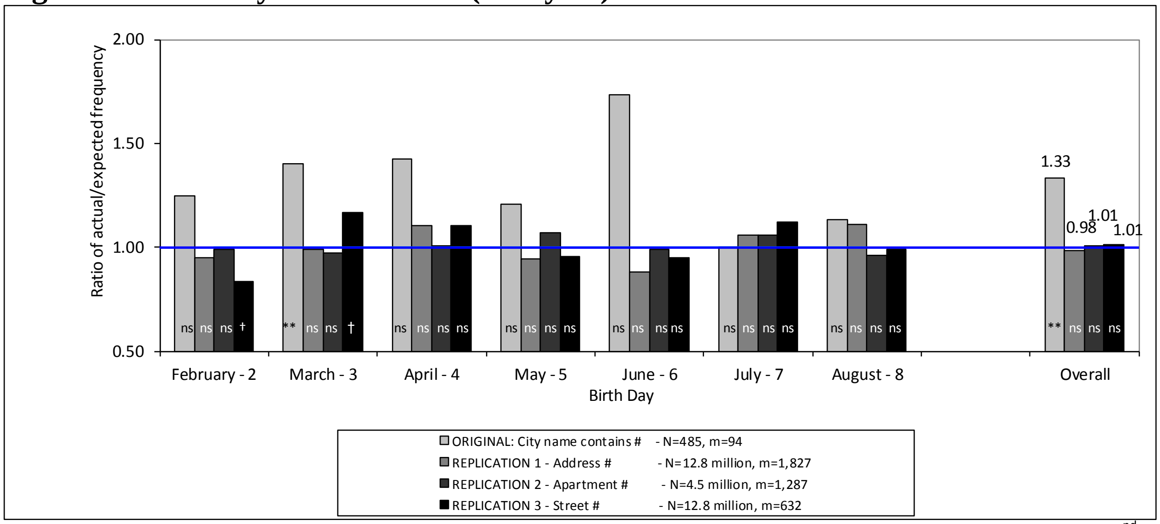
Notes: Ratios of actual over expected frequency ($R_{A/E}$) and their significance are obtained from 2x2 tables that classify people as living in the target street or not, and as having the target name or not. The entire sample (NY voter registration file eliminating repeated addresses, N=8.9 million) is used in each table.

**, * significant at the 1% and 5% level respectively. *ns*: p-value >.1.

m is the number of matching observations (e.g., 104 males name John live on a Johnson street).

M s the total number of people with that name in the sample (e.g., there is a total of 138.5 thousand Johns in the NY sample).

Figure 11. Birthday and Location (Study 14)



Source for original results. Social Security Death Index, subset of people born on a matching day-month birthday (e.g., February 2[nd]) and dying in a town with a number between 2 and 8 in its name.
Source for replication results: New York Voter Registration file (N=12.8 million)
Notes: Ratios of actual over expected frequency ($R_{A/E}$) and their significance, for the three replications, are obtained from 2x2 tables that classify people as having been born on the corresponding date or not, and living in the corresponding location or not. Each of these tables includes the entire sample (*N*). *m* is the number of matching observations (e.g., 1,827 people live on an address that matches their birthday month and day).
**,† indicates significant at the 1% and 10% level respectively. *ns*: p-value >.1.

REFERENCES

Anseel, F., & Duyck, W. (2008). Unconscious Applicants: A Systematic Test of the Name-Letter Effect. *Psychological Science, 19*(10), 1059-1061. doi: 10.1111/j.1467-9280.2008.02199.x

Baron, R. A., Branscombe, N. R., & Byrne, D. R. (2008). *Social Psychology* (1st ed.): Allyn & Bacon.

Baumeister, R. F., & Bushman, B. J. (2010). *Social Psychology and Human Nature, Comprehensive Edition*: Wadsworth Publishing

Bialik, C. (2010, 15 May 2010). Are You W.S.J.? Is That Why You're Reading This?, *The Wall Street Journal*. Retrieved from http://online.wsj.com/article/SB10001424052748704635204575242704020485696.html

Bickel, P. J., Hammel, E. A., & Oconnell, J. W. (1975). Sex Bias in Graduate Admissions - Data from Berkeley. *Science, 187*(4175), 398-404.

Breckler, S. J., Olson, J., & Wiggins, E. (2005). *Social Psychology Alive*: Wadsworth Publishing

Brendl, C. M., Chattopadhyay, A., Pelham, B. W., & Carvallo, M. (2005). Name Letter Branding: Valence Transfers When Product Specific Needs Are Active. *Journal of Consumer Research, 32*(3), 405-415.

Chandler, J., Griffin, T. M., & Sorensen, N. (2008). In the "I" of the Storm: Shared Initials Increase Disaster Donations. *Judgment and Decision Making, 3*(5), 404-410.

Cochran, W. G. (1954). Some Methods for Strengthening the Common X2 Tests. *Biometrics, 10*(4), 417-451.

Davanzo, J. (1983). Repeat Migration in the United-States - Who Moves Back and Who Moves On. *Review of Economics and Statistics, 65*(4), 552-559.

Gallucci, M. (2003). I Sell Seashells by the Seashore and My Name Is Jack: Comment on Pelham, Mirenberg, and Jones (2002). *Journal of Personality and Social Psychology, 85*(5), 789-799. doi: 10.1037/0022-3514.85.5.789

Genesove, D., & Mayer, C. (2001). Loss Aversion and Seller Behavior: Evidence from the Housing Market. *The Quarterly Journal of Economics, 116*(4), 1233-1260.

Jones, J. T., Pelham, B. W., Carvallo, M., & Mirenberg, M. C. (2004). How Do I Love Thee? Let Me Count the Js: Implicit Egotism and Interpersonal Attraction. *Journal of Personality and Social Psychology, 87*(5), 665-683.

Jones, J. T., Pelham, B. W., & Mirenberg, M. C. (2002). Name Letter Preferences Are Not Merely Mere Exposure: Implicit Egotism as Self-Regulation. *Journal of Experimental Social Psychology, 38*(2), 170-177. doi: 10.1006/jesp.2001.1497

Kitayama, S., & Karasawa, M. (1997). Implicit Self-Esteem in Japan: Name Letters and Birthday Numbers. *Personality and Social Psychology Bulletin, 23*(7), 736-742.

Koole, S. L., Dijksterhuis, A., & van Knippenberg, A. (2001). What's in a Name: Implicit Self-Esteem and the Automatic Self. *Journal of Personality and Social Psychology, 80*(4), 669-685. doi: 10.1037//0022-3514.80.4.669

Madrian, B., C., & Shea, D., F. (2001). The Power of Suggestion: Inertia in 401(K) Participation and Saving Behavior. *Quarterly Journal of Economics, CXVI*(4), 1149-1187.

McCullough, B. D., & McWilliams, T. P. (2010). Baseball Players with the Initial "K" Do Not Strike out More Often. *Journal of Applied Statistics, 37*(6), 881-891.

Myers, D. (2009). *Social Psychology*: McGraw-Hill Humanities/Social Sciences/Languages.

Nelson, L. D., & Simmons, J. P. (2007). Moniker Maladies - When Names Sabotage Success. *Psychological Science, 18*(12), 1106-1112.

Nuttin, J. M. (1985). Narcissism Beyond Gestalt and Awareness - the Name Letter Effect. *European Journal of Social Psychology, 15*(3), 353-361.

Nuttin, J. M. (1987). Affective Consequences of Mere Ownership - the Name Letter Effect in 12 European-Languages. *European Journal of Social Psychology, 17*(4), 381-402.

Pelham, B. W., Carvallo, A., & Jones, J. T. (2005). Implicit Egotism. *Current Directions in Psychological Science, 14*(2), 106-110.

Pelham, B. W., Carvallo, M., DeHart, T., & Jones, J. T. (2003). Assessing the Validity of Implicit Egotism: A Reply to Gallucci (2003). *Journal of Personality and Social Psychology, 85*(5), 800-807. doi: 10.1037/0022-3514.85.5.800

Pelham, B. W., Mirenberg, M. C., & Jones, J. T. (2002). Why Susie Sells Seashells by the Seashore: Implicit Egotism and Major Life Decisions. *Journal of Personality and Social Psychology, 82*(4), 469-487. doi: 10.1037//0022-3514.82.4.469

Sanderson, C. A. (2009). *Social Psychology*: Wiley.

Schacter, D. L., Gilbert, D. T., & Wegner, D. M. (2010). *Introducing Psychology*: Worth Publishers

Simonsohn, U., & Loewenstein, G. (2006). Mistake #37: The Impact of Previously Observed Prices on Housing Demand. *Economic Journal, 116*(1), 175-199.

Thaler, R. H., & Benartzi, S. (2004). Save More Tomorrow (Tm): Using Behavioral Economics to Increase Employee Saving. *Journal of Political Economy, 112*(1), S164-S187.

Word, D. L., Coleman, C. D., Nunziata, R., & Kominski, R. (n.d.). *Demographic Aspects of Surnames from Census 2000*. US Census Unpublished Manuscript, (http://www.census.gov/genealogy/www/).

Appendix – Frequencies used for creating figures in the text.

Tabulations for Figure 1.			
Walker County	**Actual**	**Expected**	**Ratio**
Same last name initial	907	792.4	1.15
Same entire last name	109	20.6	5.30
Net	798	771.8	1.03
Liberty County			
Same last name initial	265	209.8	1.26
Same entire last name	32	9.6	3.29
Net	233	200.2	1.16
Hispanics in Texas, 2001			
Same last name initial	2,535	2,221.9	1.14
Same entire last name	716	402.6	1.77
Net	1,819	1,819.3	1.00

Tabulations for Figure 2								
	Exact same last name				**Similar last name**			
	control		**target**		**control**		**target**	
Groom	**match**	**no match**	**match**	**no match**	**match**	**no match**	**match**	**no match**
Aguilar	578	12,125	70	562	306	10,137	22	540
Aguirre	328	12,690	20	297	564	10,144	14	283
Alvarado	546	12,227	54	508	543	9,954	24	484
Alvarez	567	12,162	75	531	523	9,951	23	508
Espinosa	148	13,049	8	130	338	10,537	5	125
Espinoza	343	12,573	47	372	142	10,491	6	366
Gonzales	1184	10,734	367	1,050	1,828	8,127	179	871
Gonzalez	2007	8,400	940	1,988	973	8,044	211	1,777
Guerra	490	12,278	65	502	571	9,932	28	474
Guerrero	599	12,108	74	554	464	9,987	26	528
Mendez	404	12,413	65	453	663	9,889	25	428
Mendoza	688	11,878	96	673	378	9,954	26	647
Mora	111	13,081	9	134	701	10,170	3	131
Morales	704	11,735	118	778	108	10,119	3	775
Salas	287	12,715	28	305	665	10,035	17	288
Salazar	682	11,787	109	757	270	9,978	17	740
Vasquez	790	11,622	111	812	217	9,976	18	794
Vazquez	235	12,792	40	268	763	9,974	27	241
Velasquez	218	12,898	23	196	94	10,715	2	194
Velazquez	96	13,093	11	135	214	10,656	4	131
			2330	11005			680	10325

Tabulations for Figure 4												
Frequencies	*ORIGINAL*				*REPLICATION*				*NEW*			
	Control grooms		**Target groom**		**Control grooms**		**Target groom**		**Control grooms**		**Target groom**	
Groom	*match*	*no match*	*match*	*no match*	*match*	*no match*	*match*	*no match*	*match*	*no match*	*match*	*no match*
Frank	907	1,125	251	316	750	1,080	113	163	319	3,452	113	1,295
Charles	386	1,197	311	705	260	715	326	805	1,197	8,014	326	2,318
Joseph	214	1,618	130	637	143	1,475	65	423	83	494	65	381
Carl	350	2,000	50	199	395	1,500	54	157	746	5,078	54	340
Robert	306	2,411	596	2,751	103	1,407	179	2,083	470	22,276	179	8,454
Paul	1,572	2,996	547	949	1,180	1,918	251	423	2,015	10,415	251	1,176
Andrew	1,169	4,483	125	287	611	2,894	74	193	245	1,103	74	296
Stephen	1,491	3,764	258	551	1,143	2,060	231	338	1,327	6,071	231	1,027
Patrick	3,847	6,336	490	712	3,396	3,453	334	302	3,371	6,351	334	588
Michael	1,502	2,023	2,754	5,106	964	1,573	1,307	3,641	1,671	3,489	1,307	2,725
Eric	207	9,859	57	1,262	335	6,602	52	496	206	1,231	52	301
Christopher	2,273	8,108	255	749	868	5,264	229	1,124	152	320	229	416

Control names				
Groom	**Original and replication**		**New**	
Frank	Frank Charles Joseph Carl	Frances Charlotte Josephine Carla	Joe Louis Ernest	Mary Alice Betty
Charles	Frank Charles Joseph Carl	Frances Charlotte Josephine Carla	James William Gerald	Beverly Janice Sharon
Joseph	Frank Charles Joseph Carl	Frances Charlotte Josephine Carla	Paul Patrick Peter	Beatrice Eva Gloria
Carl	Frank Charles Joseph Carl	Frances Charlotte Josephine Carla	James Kenneth Lawrence	Pamela Sheila Tina
Robert	Robert Paul Andrew Stephen	Roberta Paula Andrea Stephanie	John Thomas James	Margaret Mary Nancy
Paul	Robert Paul Andrew Stephen	Roberta Paula Andrea Stephanie	John Robert Joseph	Cynthia Cindy Nancy
Andrew	Robert Paul Andrew Stephen	Roberta Paula Andrea Stephanie	Joseph Patrick Peter	Christina Rachel Michelle
Stephen	Robert Paul Andrew Stephen	Roberta Paula Andrea Stephanie	Michael Douglas Russell	Angela Michelle Jennifer
Patrick	Patrick Michael Eric Christopher	Patricia Michelle Erica Christine	Michael Gregory Keith	Sandra Nancy Brenda
Michael	Patrick Michael Eric Christopher	Patricia Michelle Erica Christine	Steven David Gregory	Angela Stephanie Tanya
Eric	Patrick Michael Eric Christopher	Patricia Michelle Erica Christine	Christopher Jonathan Aaron	Jessica Monica Vanessa
Christopher	Patrick Michael Eric Christopher	Patricia Michelle Erica Christine	Jonathan Sean Eric	Theresa Denise Valerie

Tabulations for Figure 6	"BORN"				"STAYED"			
	Controls		Target First name		Controls		Target First name	
Name	*Matching state*	*Other State*	*Matching state*	*Other State*	*Matching state*	*Other State*	*Matching state*	*Other State*
Georgia	177,252	5,796,942	126	3051	137,559	39,693	98	28
Louise	135,855	5,836,210	136	5170	103,030	32,825	97	39
Virginia	140,966	5,823,713	447	12245	101,497	39,469	313	134
Florence	202,228	5,771,762	77	3304	150,242	51,986	63	14
George	175,909	5,756,528	1469	43465	136,566	39,343	1,091	378
Louis	135,375	5,831,204	616	10176	102,650	32,725	477	139
Virgil	141,347	5,833,639	66	2319	101,767	39,580	43	23
Kenneth	119,491	5,816,082	1008	40790	88,105	31,386	776	232

Tabulations for Figure 9	"BORN"				"STAYED"			
	Controls		Target First name		Controls		Target First name	
Lastnames	*Matching state*	*Other State*	*Matching state*	*Other State*	*Matching state*	*Other State*	*Matching state*	*Other State*
Cal_	521,493	5,443,415	1211	11252	399,196	122,297	931	280
Tex_	382,414	5,594,839	1	117	319,453	62,961	1	0
Flo_	201,918	5,765,258	387	9808	150,018	51,900	287	100
Ill_	293,063	5,684,136	21	151	196,522	96,541	16	5
Pen_	295,205	5,674,175	304	7687	218,546	76,659	224	80
Ohi_	279,770	5,697,585	0	16	204,887	74,883	0	0
Mic_	239,065	5,733,418	266	4622	177,717	61,348	198	68
Geo_	177,298	5,796,691	80	3302	137599	39699	58	22

Tabulations for Figure 10	Rest of NY		Targets	
	Matching street	*Other street*	*Matching street*	*Other street*
First:John Street:Johnson	6,136	8,747,843	104	138,530
First:William Street:Williams	5,069	8,797,766	48	89,730
First:Jon/Jonathan Street:Jones	2,749	8,868,038	6	21,820
First:David Street:Davis	4,593	8,791,738	58	96,224
First:George Street:Washington	45,507	8,812,087	171	34,848
First:Jeff/Jeffery Street:Jefferson	15,909	8,841,546	57	35,101
First:Thomas/Tom Street:Jefferson	15,860	8,808,770	106	67,877
Last:East_ Street: East	506,577	8,384,198	102	1,736
Last:West_ Street: West	373,496	8,511,333	266	7,518
Last:North_ Street: North	25,945	8,865,144	8	1,516
Last:South_ Street: South	33,328	8,857,989	4	1,292

Tabulations for Figure 11

Apartment #s (e.g., Apt 2)

	Everyone else in NY State		**People with target birthday**	
	Matching apt #	*Other apt #*	*Matching apt #*	*Other apt #*
2-Feb	13,042	4,275,816	605	208,875
3-Mar	12,828	4,387,950	281	97,279
4-Apr	12,500	4,439,395	144	46,299
5-May	13,108	4,457,474	77	27,679
6-Jun	13,170	4,461,578	61	23,529
7-Jul	13,238	4,465,232	62	19,806
8-Aug	13,359	4,467,778	57	17,144
		m	1,287	
			4,498,338	

Address (e.g. 2 Elm Street)

	Everyone else in NY State		**People with target birthday**	
	Matching address	*Other addresses*	*Matching address*	*Other addresses*
2-Feb	37,274	12,730,578	258	88,915
3-Mar	36,776	12,733,317	243	86,689
4-Apr	35,932	12,731,658	254	89,181
5-May	36,629	12,726,729	287	93,380
6-Jun	36,848	12,730,289	257	89,631
7-Jul	37,418	12,728,570	283	90,754
8-Aug	37,673	12,732,897	245	86,210
		m	1,827	
			12,857,025	

Street # (e.g. 2nd avenue)

	Everyone else in NY State		**People with target birthday**	
	Matching Street	*Other Streets*	*Matching Street*	*Other Streets*
2-Feb	37,513	12,811,635	93	38,028
3-Mar	36,975	12,810,720	133	39,441
4-Apr	36,188	12,825,951	78	25,052
5-May	36,898	12,808,936	114	41,321
6-Jun	37,146	12,830,017	55	20,051
7-Jul	37,707	12,823,971	84	25,507
8-Aug	37,940	12,823,564	75	25,690
		m	632	
			12,887,269	