Spurious? Name Similarity Effects (Implicit Egotism) in Marriage, Job, and Moving Decisions

Uri Simonsohn
The Wharton School, University of Pennsylvania

Three articles published in the *Journal of Personality and Social Psychology* 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 included in most modern social psychology textbooks and many university courses. The current article successfully replicates the original findings but shows that they are most likely caused by a combination of cohort, geographic, and ethnic confounds as well as reverse causality.

Keywords: name–letter effects, field studies, implicit egotism, statistics, spurious results

> 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, *Cosmos*

The self-correcting aspect of science is arguably its strongest virtue compared with 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 article challenges a literature that is now broadly accepted in social psychology—one that is part of many textbooks, as well as both undergraduate and graduate courses.[1]

This article reexamines 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, consisting of the observation that people like letters contained in their name more than other people do (Nuttin, 1985). The name–letter effect 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., 2002, 2004; 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 name–letter effect in decisions outside the lab began with the work of Pelham et al. (2002), who showed a disproportionate number of people gravitating toward states, cities, and occupations with names resembling their own (e.g., that Lauras are lawyers and Florences live in Florida). Pelham et al. referred to this generalization of the name–letter effect as *implicit egotism.* Pelham, Carvallo, DeHart, and 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 reevaluate these three articles, published in the *Journal of Personality and Social Psychology*. Altogether these include 16 field studies demonstrating eight main

This article was published Online First February 7, 2011.

I would like to thank Jon Baron, Leif Nelson, and Maurice Schweitzer for encouraging me to pursue 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 School students Daniel Li and Brendan Forster provided outstanding research assistance. Alessandro Acquisti generously shared some of the data used in this article. I extend a special thanks to Marcello Gallucci, who provided useful feedback despite being aware of my strong reservations with Gallucci (2003).

The author accepts sole responsibility for all remaining errors.

Correspondence concerning this article should be addressed to Uri Simonsohn, 500 Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104. E-mail: uws@wharton.upenn.edu

[1] See, for example, the textbooks by Baron, Branscombe, and Byrne (2008); Baumeister and Bushman (2010); Breckler, Olson, and Wiggins (2005); Myers (2009); Sanderson (2009); Schacter, Gilbert, and Wegner (2010).

Journal of Personality and Social Psychology, 2011, Vol. 101, No. 1, 1–24

0022-3514/11/$12.00 DOI: 10.1037/a0021990

findings. These are, in the order discussed in this article, that people are disproportionately likely to (a) marry someone with a similar last name and (b) first name, (c) choose occupations with names resembling their first name, (d) move to states with names resembling their first name, live (e) in states, (f) in towns, and (g) on streets, with names resembling their last name, and (h) live in towns whose names contain their birthday numbers. The current article is organized according to these eight original findings. Each is first summarized, then potential problems with the causal interpretation of the results are put forward, and finally new studies that test for implicit egotism net of these problems are presented. For an overview of all new studies, see Table 1.

Although 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 focused 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 was to conduct analyses similar to those of Pelham et al. (2002), Pelham et al. (2003), and Jones et al. (2004) but in subsamples for which there was a substantially reduced degree of heterogeneity in such variables. Although implementation details varied across studies, the overarching approach and conclusion was 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 Pelham et al. (2002), Pelham et al. (2003), and Jones et al. (2004), typically consisting of chi-square tests (with $df = 1$) performed on 2×2 tables. The results are presented in figures that make the impact of eliminating the proposed confound evident to the naked eye.

In reexamining evidence for implicit egotism, the current research is related to (Gallucci, 2003), an early critique of Pelham et al. (2002).[2] Most relevant for the present article, Gallucci showed that there is considerable variation in the point estimates of name-similarity effects in at least one of the 10 studies in Pelham et al. (2002) and expressed the concern that overall estimates that aggregate across names hence may be driven by a small minority that exhibit large effects. Gallucci 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 with Pelham et al. (2002), Pelham et al. (2003), and Jones et al. (2004) 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 article is that the former concludes that the findings of Pelham et al. (2002) are not statistically significant, that there simply is no name-similarity effect to be explained beyond sampling error, whereas the present article, in contrast, concludes that there is a sizable, robust, widespread, and statistically significant name-similarity effect in Pelham et al. (2002) (and Pelham et al., 2003, and Jones et al., 2004) but that implicit egotism is not a likely cause for such effect, whereas cohort, ethnic, and geographic confounds are.[3]

Existing Finding 1: Last Name and Marriage

Jones et al. (2004) analyzed marriage licenses and birth certificates from two counties and several states. They found that the share of marriages where groom and bride share an initial or an entire last name was greater than expected in all samples considered.

Potential Problems With These Findings

Jones et al. (2004) mentioned 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 have been recorded, by mistake, as also being the bride's last name.

Jones et al. (2004) addressed 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-initials) findings in a subsample where all grooms had a Hispanic last name. The authors addressed the archival error concern by showing that 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 the findings of Jones et al. (2004). Study 1 suggests that the same-last-name-initials effect documented in Pelham et al. (2002) is driven in part by a same-entire-last-name effect and in part by an ethnic confound.

Lacking ethnicity information in the original data sets, Study 2 draws on 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 draws on an ethnically homogenous sample, focusing only on marriages among people with Hispanic last names. This study replicates 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

[2] Pelham et al. (2003) is a response to Gallucci's comment.

[3] Gallucci (2003) carried out four major sets of analyses to assess the reliability and/or statistical significance of the studies in Pelham et al. (2002), concluding that "the hypothesis [of a name-similarity effect] is not supported for the large majority of names considered by the [Pelham et al., 2002] authors" (p. 789). It is beyond the scope of this article to critically review Gallucci's attempts to properly estimate statistical significance.

[4] Jones et al. (2004) argued that their study examining entire last names also addressed ethnic sorting because they focused on five last names (Smith, Johnson, Williams, Jones, and Brown) that are "common European American Names" and are "extremely common Caucasian names" (p. 668). According to the Census of 2000 (Word, Coleman, Nunziata, & Kominski, 2009), however, the share of people with these last names that are White is just 61.7%.

Table 1
Overview of New Studies

Original finding	Between	Main problem with original finding	New study	New finding	Figure or table
I	Last name and 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 name to husband-to-be's before marriage	Study 1	The same-last-name-initial marriage effect is an exact-same-last-name effect	Figure 1
			Study 2	Few ethnic last names account for significant share of same-last-name marriages	Table 2
			Study 3	Very similar last names do not disproportionately marry each other	Figure 2
			Study 4	Remarriages to same previously divorced spouse are frequent	
II	First name and marriage	**Cohort confound** Popularity of similar male and female first names changes 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 3
III	First name and occupation	**Cohort (and probably socioeconomic status) confound** Original studies chose control names on the basis of overall 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 4
			Study 7	Georges no more likely to be in geology than any other science	None
IV	First name and 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 5
			Study 9	Names popular among state natives are popular among state immigrants	Figures 6 and 7
V	Last name and state	**Geographic/ethnic confound** More babies born in California with last name Cali__.	Study 10	People are not more likely to stay in state with a name resembling their last name	Figure 8
VI	Last name and town	**Reverse causality**	Study 11	Towns are named after their founders	None
VII	Last name and street	**Reverse 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 names.	Study 12	No effect of first names or last names with geographic designators on street choice	Figure 9
VIII	Birthday and 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 10

Morales or for Gonzale*s* to marry Gonzale*z*), suggesting that 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 that of their husband-to-be before marriage.

Study 1: The Same-Initials Marriage Effect Is a Same-Last-Name Effect

Method. The marriage archives for Walker (Georgia) and Liberty (Florida) Counties used in Study 1 of Jones et al. (2004) were obtained from the USGenWeb Archives (http://usgwarchives.net). The Walker County data set spans 1882–1990 (N = 11,855) and the Liberty County data spans 1823–1965 (N = 3,063). Seeking a data set that was both larger in number of observations and more concentrated timewise, I obtained the Texas Marriage License Application Index for 2001; this data set includes all marriages occurring in Texas that year (N = 195,030).[5] Jones et al. (2004) alluded to this latter data set in their discussion of Studies 1–3 and reported some secondary analyses on it (p. 669).[6]

To address the issue of ethnic sorting, I created a subset of the Texas 2001 data set 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, Coleman, Nunziata, & Kominski, 2009), N = 24,645. The key analysis in all

[5] For unknown reasons, sample sizes are slightly different from those reported in Jones et al. (2004). The results from my replications, therefore, though 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

samples was the comparison of the expected versus 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 would be $5\% \times 4\% = 0.2\%$. As was done in Jones et al. (2004) and in line with Cochran (1954), I added these proportions across all letters and, through a chi-square test (with $df = 1$), compared the sum with the total numbers of such marriages observed in the sample.

Results. In Walker County, the expected proportion of same-initials marriages was 6.67%, and the actual proportion was 7.65%, $\chi^2(1) = 18.4$, $p < .0001$. In Liberty County, these proportions were 6.85% and 8.65%, respectively, $\chi^2(1) = 15.6$, $p < .0001$. These results are nearly identical to those reported in Jones et al. (2004) (see Footnote 5).

Here and for all other studies in this article, the ratio of actual over expected frequencies was 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, they were over three times the expected rate. For example, in Walker County, by chance, 0.17% of grooms would have been expected to marry a bride with their same last name, but 0.92% actually did, $\chi^2(1) = 381$, $p < .0001$.

Subtracting the frequencies of same-last-name marriages from the frequencies of same-initials marriages, one obtains the frequencies of marriages in which spouses share an initial but not the full last name. The resulting $R_{A/E}$ values are shown with black bars in Figure 1. For both samples originally used in Jones et al. (2004), the $R_{A/E}$ values dropped markedly both in size and statistical significance, indicating that same-last-name marriages accounted for a substantial share of the same-initials marriages.[7]

Despite the drop in size, the $R_{A/E}$ for Liberty County was 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, a total of 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-initials effect (net of same-last-name effect) in Liberty County. If so, there should be no such effect in the ethnically homogenous sample of Hispanic marriages from Texas in 2001; this was indeed found to be the case.

The results are shown in the last three bars of Figure 1. Before excluding same-last-name marriages, I note that also in this sample there was a significant same-initials marriage effect, $R_{A/E} = 1.14$, $\chi^2(1) = 49$, $p < .0001$. This effect, however, was entirely driven by a same-last-name effect, $R_{A/E} = 1.77$, $\chi^2(1) = 244$, $p < .0001$. Once these observations were excluded, no same-initials effect remained, $R_{A/E} = 1.00$, $\chi^2(1) = .014$, $p = .90$.

In sum, Study 1 suggests that the same-initials effect is a combination of an ethnic confound and a same-last-name effect. The same-last-name effect could, of course, have arisen because of implicit egotism. In Studies 2–4, I examined this possibility.

Study 2: A Miniscule Degree of Ethnic Heterogeneity Can Lead to Large (and Spurious) Same-Last-Name Effects

Given the considerable ethnic heterogeneity in Walker and Liberty Counties alluded to above, and the fact that the same-initials 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 might be obtained as a consequence of ethnic heterogeneity.

To examine 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. The fourth column shows the percentage of grooms marrying a bride with the same last name. These ratios ranged from extremely high (11% for Tran) to astronomically high (63% for Patel).[9] The sixth column shows the absurdly high $R_{A/E}$ values, hovering in the hundreds and thousands.[10] Most relevant for our purposes, the seventh column shows the percentage of the overall same-last-name effect in the entire Texas 2001 sample that was accounted for by each last name. Fully 17.6% of the entire excess of same-last-name marriages were accounted for by these four last names, which, combined, constituted 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 account for 217 such marriages, this means that if Texas was 99.89% homogenous (except 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 was performed on the Hispanic subsample only.

---

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

[8] Walker County was 16% Black in 1900. Jones et al. (2004) reported percentage of population being White only for Walker county and for the Census of 1990 rather than for 1900. Considering that the marriage data span was 1820–1990, the Census of 1900 is probably more representative.

[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.

[10] As remarked by a referee, although 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 data set where both groom and bride had one of Vietnam's 12 most common last names (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).

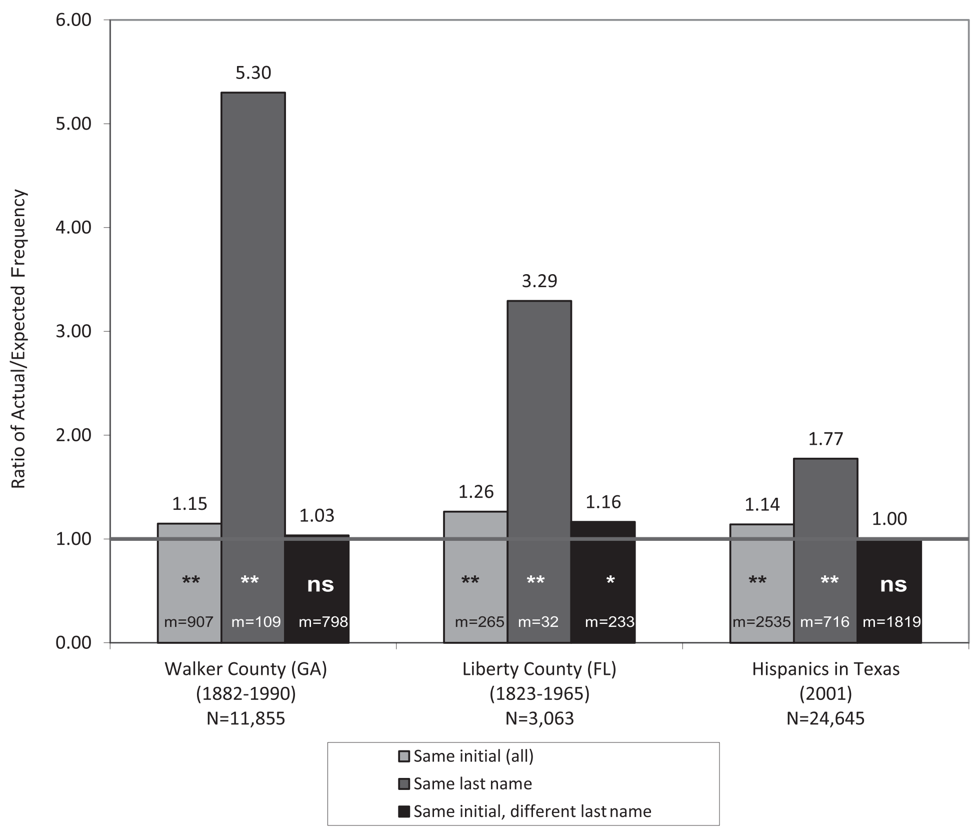

Figure 1. Same-last-name and same-initials marriages (Study 1). From marriage licenses from the respective locations. Expected marriage rates were obtained by multiplying the base rates of each initial/last name for grooms and brides and then adding up all the resulting products. Significance was assessed by chi-square tests (with $df = 1$) comparing aggregate actual versus expected rates. N = 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 contained 11,855 marriages, of which 907 were between people sharing a last name initial. * $p < .05$. ** $p < .01$. (*ns* indicates $p > .10$)

Study 3: People With Very Similar Last Names Do Not Disproportionately Marry Each Other

Method. Study 1 suggested that the same-initials effect actually consists of a same-last-name effect. The latter, of course, could still have arisen because of implicit egotism. It may very well be the case that initials are not sufficiently strong triggers of implicit egotism, whereas identical last names are. If a psychological mechanism is 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 grooms last named Morales to marry a disproportionate share not only of Morales brides but also of Mora and Moraga ones. To examine this question, while continuing to control for ethnicity, I extracted a subset of 10 pairs of very similar last names from the set of 200 previously mentioned Hispanic ones (see the x-axis in Figure 2 for the full list).

Table 2
Four (Ethnic) Last Names Accounting for 18% of All Same-Last-Name Marriages in Texas (Study 2)

Last name	Origin	No. observations	Marrying bride with same last name	% of sample	Ratio of actual over expected	% of total same last name 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. The last names presented are the four with the highest percentage of same-last-name marriage. These 217 marriages account for 17.6% of all same-last-name marriages in Texas in 2001 (N = 195,030).

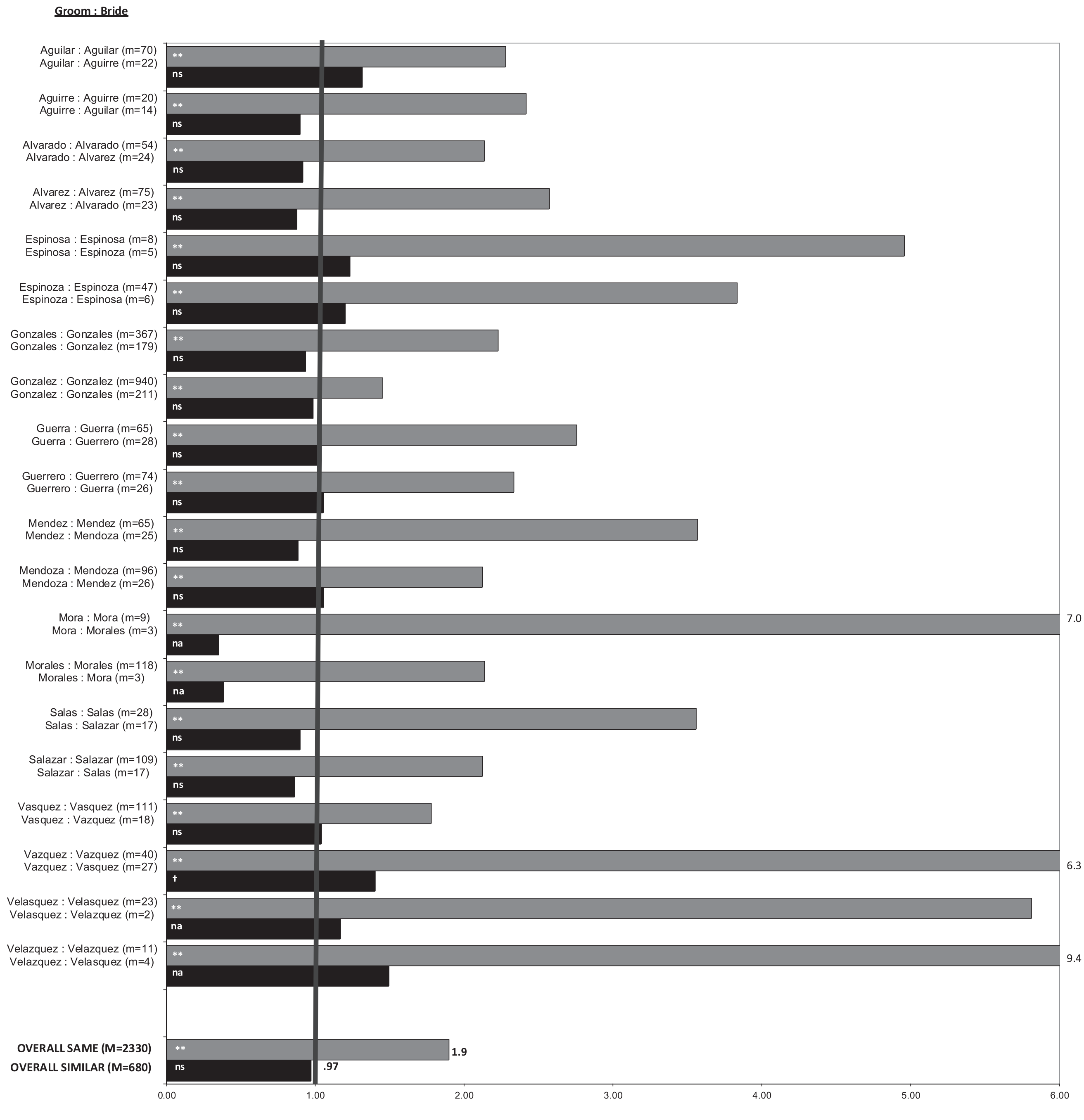

Figure 2. Same and similar-last-name marriages (Study 3). From database of Marriages in Texas (1966–2007) between a groom and a bride with one of the 20 last names presented in the figure (N = 13,335). Ratios of actual over expected frequency ($R_{A/E}$) values, and their significance, were obtained with 2 × 2 tables that categorize marriages as either including a target male name or not and as including a target female name or not. Each of the 20 same-last-name tables included the entire sample. Each of the 20 similar-last-name tables included 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. m = the number of matching marriages (e.g., there are 70 marriages between two people last named Aguilar); *na* = fewer than five expected observations. $^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. (*ns* = $p > .10$).

I started from the full Texas marriage data (1966 to 2007) and created 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 versus expected frequency of marriages was compared for people with identical last names and for people with very similar last names. For example, for grooms named Gonzale*z* ($n = 2{,}928$), the same-last-name comparison contrasted the proportion of them marrying a Gonzale*z* bride ($P = 32.1\%$) with the proportion of grooms who had the other 19 last names ($n = 10{,}407$) doing so ($P = 19.3\%$), $\chi^2(1) = 218.1$, $p < .0001$. The second test compared the proportion of Gonzale*z* men marrying a Gonzale*s* bride ($P = 10.6\%$) with the proportion of men not named Gonzale*z* doing so ($P = 10.8\%$), $\chi^2(1) = 0.05$, $p = .81$.[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 same last name considered (all ps $< .01$). The black bars show the $R_{A/E}$ values for similar last names; all were quite close to $R_{A/E} = 1$, and the only two statistically significantly different from it were below it. In other words, there is no evidence that having an even extremely similar last name with another person increases the odds of marrying that person by even a small amount.

The last two bars in Figure 2 report aggregate results. Here and in all other analyses in the article that combine multiple 2×2 tables, the overall $R_{A/E}$ was the ratio of the sum of expected frequencies over the sum of actual frequencies. The significance of such aggregate $R_{A/E}$ values was obtained with the method from Cochran (1954).[12] The overall bars show that there were nearly twice as many same-last-name marriages as would be expected by chance ($R_{A/E} = 1.9$, $p < .001$) but slightly fewer than expected very similar last-name marriages ($R_{A/E} = 0.97$, $p = .438$).

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*s* reminds Gonzale*s* of himself, but Gonzale*z* does not), but lab evidence for implicit egotism does find reliable effects when participants share but a few letters with the target. For example, in Study 6 (lab) of Jones et al. (2004), the authors found that people are more attracted to individuals whose mock username shares just three letters with their own last name (as in Larry Murray being more attracted to STACEY_MUR than to STACEY_PEL; p. 675).

A similar alternative account, though more post hoc, is that similar and identical names trigger implicit egotism but that almost identical last names trigger repulsion (e.g., a person last named Gonzale*z* dislikes the name Gonzale*s*). This explanation (first put forward by Brett Pelham in an interview discussing an earlier version of this article; 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, hence, a net effect of implicit egotism would be expected for these names, but this is not observed. 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 is shown below, people with very similar first names (e.g., Andrew and Andrea) are disproportionately likely to marry each other, but the repulsion story would predict otherwise.

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

There are multiple mechanisms that could lead to this occurrence: Women may (a) marry a relative; (b) upon divorcing or widowing, marry a relative of their ex-husband; (c) marry by the church/abroad/in another state/by common law marriage, change their last name, and then marry by the state; (d) renew their vows through a new marriage, and so forth.

Most of these mechanisms are not easy to identify in the available data. One additional mechanism is the following: A couple marries, the wife changes her last name, the couple divorces, and 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 the www.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 one record. These 169 listings were given to a research assistant to check by hand 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. For example, one of the same-last-name 2001 marriages consisted of Candi

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

[12] The method is as follows. Let the four cells of 2×2 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 the sum of 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 the Appendix in Cochran (1954) for an in-depth discussion. The formula itself was obtained from Bickel, Hammel, and O'Connell (1975).

A. Nehring (born in 1967) marrying Stephen E. Nehring (born in 1961). The search revealed that in 1982 a Candi A. Hill (born in 1967) married a Stephen E. Nehring (born in 1961) and that in 1997 a Candi A. Nehring (born in 1967) divorced a Stephen E. Nehring (born in 1961).

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

Existing Finding 2: First Names and Marriage

Jones et al. (2004) also examined whether people with similar first names are disproportionately likely to marry each other. They did this by analyzing joint listings in a nationwide online phone directory (their Study 4). They first identified 12 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, Jones et al. grouped these 12 pairs into three sets of four according to the popularity of the male name in the Censuses of 1960 and 1920. They then obtained the number of phone listings for the 16 combinations of male–female names within each group and compared expected and actual frequencies. Jones et al. found greater than expected frequencies, that is, 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 10 of the 12 name pairs, five with $r > .9$.

During years when a greater share of baby girls were being named Erica, for example, more baby boys were being named Eric; 20 years or so later, the share of Erics marrying Ericas will naturally increase above expectations that assume name popularity is stable over time.

Does the Jones et al. (2004) solution of grouping name pairs on the basis of the Census frequencies of male names fix this problem? Probably not. First, splitting data into 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 do not know if this is the case), they may still be quite different within groups. This problem is amplified by the fact that Jones et al. (2004) did 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 number 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, in Study 5 I analyzed marriage records rather than phone listings and used control names explicitly chosen so as to minimize the role of cohort (and also ethnic and geographic) confounds when testing for implicit egotism.

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 lies with estimating a convincing expected frequency. For the calculation of expected marriage rates, Jones et al. (2004) used rates of the other names in a given group. For example, the Andrew group also included Robert, Paul, and Stephen, and hence Jones et al. (2004) tested whether the share of Andrews marrying Andreas (rather than Robertas, Paulas, and Stephanies) was the same as the share of Roberts, Pauls, and Stephens (combined) doing so.

As mentioned above, these calculations are likely to be biased because of cohort and also possibly ethnic and geographical confounds. An ideal control for Andrew would be a male name that if it were not for an implicit egotism effect, we would strongly expect an individual with that name to marry Andreas at the same rate that those with the name Andrews do. One way to approximate this ideal control is to find male names of men that marry other women at rates similar to Andrews. For example, if 2% of men named Andrews marry women named Katy and 1% marry women name Schwandras, and if Devins marry these women at 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 Jones et al. (2004), I started from the full data set of marriage licenses for Texas 1966–2007 ($N = 7.58$ million) and focused only on people's first names. After excluding those appearing fewer than 100 times, I obtained 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 men and 2,967 rows for women, where each cell contained the actual frequency of marriages for that specific combination of names. The correlation between any given two columns indicates how similar the distributions of brides' names were for two groom names, whereas the correlation between two rows shows how similar those two female names were in their choice of groom names.

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

To test the first-name similarity effect, I created 12 name groups, one for each target name pair from Jones et al. (2004; e.g.,

[13] Using www.ancestry.com data on birth certificates, I computed the $R_{A/E}$ for children with names similar to those of their parents, with the name combinations from Pelham et al. (2002; e.g., father: Eric, daughter: Erica). Every one of the 12 pairs had $R_{A/E} > 1$, with an overall effect of $R_{A/E} = 1.93$.

Stephen and Stephanie), and included in the group the top-three most correlated names with each name. For example, the group of Stephen and Stephanie includes the three names most correlated with Stephen (Michael, Douglas, and Russell; rs $> .994$) and with Stephanie (Michelle, Angela, and Jennifer; rs $> .992$), leading to a total of eight names (as in the original analyses). The first-name similarity effect on marriage was tested by comparing the proportion of Stephens marrying Stephanies (rather than women with the names Michelle, Angela, and Jennifer, combined) with the proportion of men named Michael, Douglas, and Russell (combined) doing so.

Results. Figure 3 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 Jones et al. (2004; obtained directly from their Table 2). Ten of the original 12 $R_{A/E}$ values were greater than 1, nine of these significantly so ($p < .05$). One of the two $R_{A/E}$s < 1 was significant (Michael: Michelle).

Jones et al. (2004) aggregated the results by weighting observations equally within groups and then taking a simple average across groups (p. 670), leading to an overall $R_{A/E} = 1.08$. Figure 3 reports the overall effect computed with the same method, but statistical significance was not straightforward to compute and hence it was not reported in Jones et al. (2004) nor here. Adding up all the expected and actual frequencies, as elsewhere in this article, led to an overall $R_{A/E} = 1.03$, $p < .001$.

The darker gray bars show the results from the same analyses used by Jones et al. (2004) on the new data, with an overall effect dropping to $R_{A/E} = 1.03$ (from $R_{A/E} = 1.08$), suggesting that using phonebook 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 was significantly greater than 1.00 at the 10% level. All $R_{A/E}$s were noticeably closer to 1, including the three name pairs with the biggest $R_{A/E}$ in the original analyses. Note that Michael: Michelle also got 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 Jones et al. (2004), was $R_{A/E} = 1.00$. When computed as elsewhere in this article, it was $R_{A/E} = 1.01$, $p = .36$.

Existing Finding 3: First Names and Occupation

Study 7 in Pelham et al. (2002) compared 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] Pelham et al. (2002) obtained their data from online dentist and lawyer directories, querying them for the respective names in the eight most populous U.S. states. Pelham et al. (2002) found a higher than expected share of professionals with a name–occupation match, though the effect was not significant for men ($p = .14$), and it was small in absolute magnitude for women; there are 1,512 female La__ lawyers compared with the expected frequency of 1,503.4 ($p = .03$).

Concerned about the size of these effects, Pelham et al. (2002) then obtained dentist data for all 50 states and compared the Den__ names with names of similar popularity in the 1990 Census. For instance, Dennis was the 40th most common first name in the 1990 Census, so Pelham et al. compared the number of dentists named Dennis ($n = 482$) with that of dentists named Jerry, ranked 39th ($n = 257$), and dentists named Walter, ranked 41st ($n = 270$). This large difference in frequencies corresponded to an $R_{A/E} = 1.43$. Pelham et al. obtained similar results for all seven other Den__ names, but the frequencies for other names were very low and not reported in their article.

In their Study 8, Pelham et al. (2002) again used the 1990 Census as a baseline and compared the numbers of Georges and Geoffreys who work in the geosciences (e.g., geology) with the number of geoscientists with names similarly frequent in the Census (Pete, Bennie, Donald, Randolph, Mark, Jonathon, Kenneth, and Daniel). The authors found more Georges and Geoffreys than expected based on such controls ($R_{A/E} = 1.42$).

Potential Problems With These Findings

One of the implicit assumptions on the dentists/lawyers 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, whereas La__ names remained stably popular from the 1920s until the 1990s.[15]

Given how small the dentists versus lawyers effects are, how large the cohort confound is, and the fact that there is no date of birth information in the professional directories that would allow one to control for cohort effects, I was unable to assess the role that these effects play in the results. I focused instead on the results that use control names based on Census frequencies. One problem with such frequencies is that they do not capture 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 retired. Consistent with this logic, Study 6 here shows that Dennises are just as overrepresented (compared with 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 whether there are "too many" George geoscientists, one should take such factors into account. To this end, Study 7 here compared the number of geoscientists having the names used by Pelham et al. (2002) with other scientists with those

[14] The 16 names were: Dennis, Denis, Denver, Denny; Denise, Dennis, Denna, Denice; Lawrence, Larry, Lance, Laurence; Laurie, Laverne, Lauren, Laura.

[15] The data on the frequency of lawyers and dentists were obtained from the U.S. Bureau of Labor Statistics. The names data are from www.babynamewizard.com (whose source is the Social Security Administration). See following two graphs: http://www.babynamewizard.com/name-voyager#prefix=den&ms=false&exact=false and http://www.babynamewizard.com/name-voyager#prefix=la&ms=false&exact=false

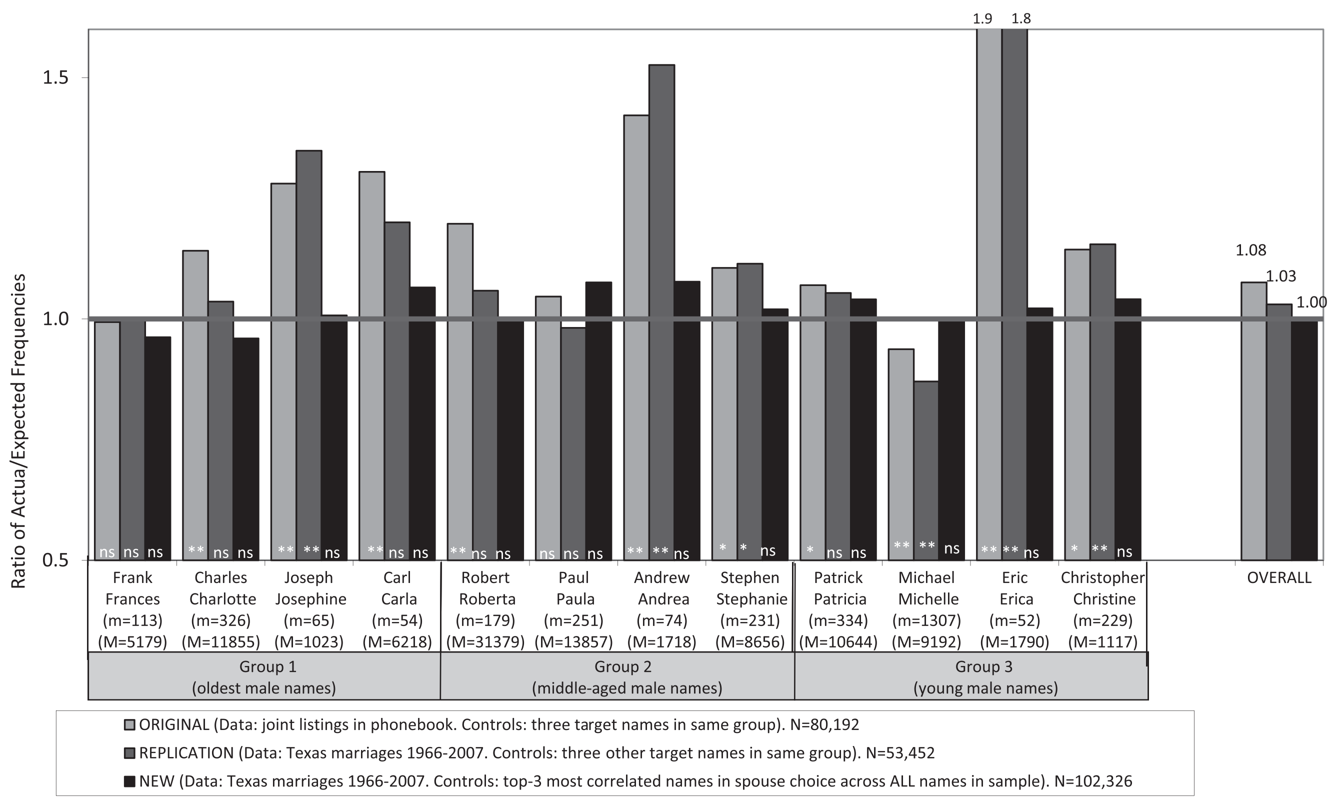

Figure 3. Similar first names and marriages (Study 5). Groups 1–3 are used for the original and replication analyses only. They combine, as in Jones et al. (2004), names on the basis of the difference in popularity between the 1960 and 1920 Census of the corresponding male name. Ratios of actual over expected frequencies ($R_{A/E}$), and their significance, were obtained with 2 × 2 tables that categorize marriages as including the target male name or not and as including the target female name or not. The 2 × 2 tables for the original and replication analyses included marriages among names within a group of names. The 2 × 2 tables for the new analyses included marriages between target names and their three new controls names. m = the number of matching-name marriages for the name pair in the new data set; M = the total number of marriages in the corresponding set of eight names (e.g., there are 5,179 marriages among Frank, Frances, and the top-three 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 Ms is greater than N because some control names were used more than once. * $p < .05$. ** $p < .01$. (ns = $p > .10$).

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, which they are.

Study 6: More Dennis Than Walter Dentists But Also More Dennis Than Walter Lawyers

Figure 4 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 although the raw number of Dennises, Jerrys, and Walters might have been very similar in 1990, the share of them at working age, especially some 10 years later when Pelham et al. (2002) collected the data, is unlikely to have been the same.

If this age (or some other unobserved) discrepancy is behind the greater frequency of dentists named Dennis, then we should find the same pattern in other professions. Furthermore, Dennis should be a more frequent name in any sample overrepresenting younger (or alive) individuals. To test this prediction in a parsimonious manner, I used the same lawyer directory used by Pelham et al. (2002) 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 = 2,889) than Jerry (n = 1,411) or Walter (n = 2,000). 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 was not significant, $\chi^2(1) = 1.28$, $p = .25$. For lawyers named Denise, a similar pattern arose: There were more lawyers

[16] The statistical significance of the difference of the $R_{A/E}$s for dentists and lawyers for the name Dennis 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%).

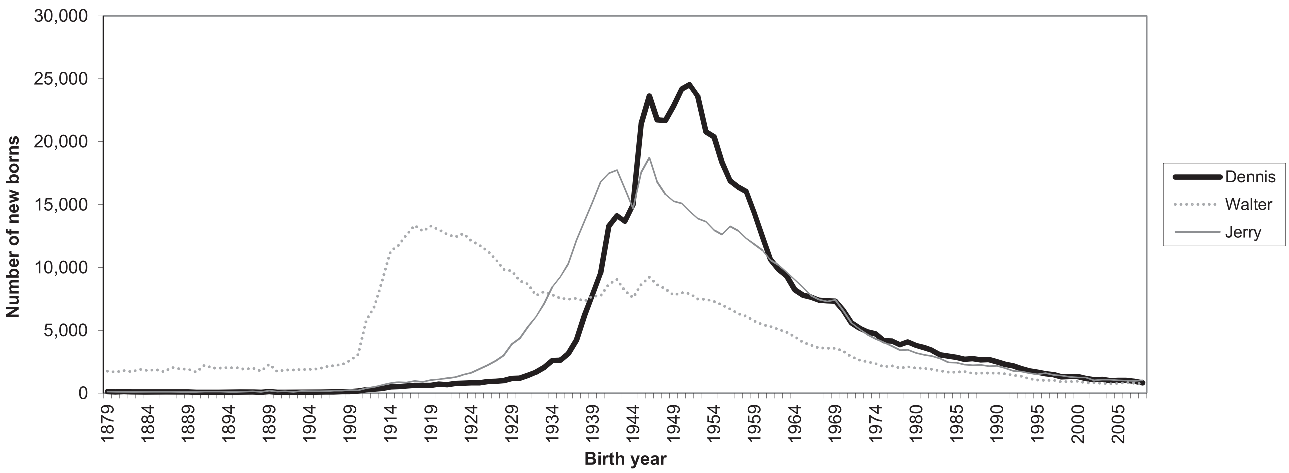

Figure 4. Popularity of first names used in dentists' study (Study 6). From the Social Security Administration.

named Denise ($n = 947$) than Beverly ($n = 454$) or Tammy ($n = 247$), implying $R_{A/E} = 1.72$.[16]

Study 7: Georges Are No More Likely to Be in Geology Than Any Other Science

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 the 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 Pelham et al. (2002), I restricted the queries to dissertations written since 1950.

Consistent with the findings from Pelham et al. (2002), the number of geoscience dissertations written by a George or a Geoffrey (682) was greater than that which would be expected on the basis of 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 nongeoscience dissertations were written by a George or a Geoffrey compared with 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 and States

Pelham et al. (2002) examined whether people are disproportionately likely to live in, and move to, states with names resembling their first name. They analyzed data from the Social Security Death Index (SSDI), a file containing information on deceased individuals who participated in the U.S. Social Security system. The data set includes information on the state in which individuals obtained their Social Security number and the zip code to which benefits were last sent.[17]

Pelham et al. (2002) proxied for living in a state by having received benefits there and cleverly proxied for moving to a state by receiving benefits there but having obtained the Social Security number in a different state. Their analyses focused on eight different first names that are either identical to, or share a few letters with, a state name: George (Georgia), Louis (Louisiana), Virgil (Virginia), and Kenneth (Kentucky); Georgia (Georgia), Louise (Louisiana), Virginia (Virginia), and Florence (Florida).

Potential Problems With This Finding

Pelham et al. (2002) were rightly concerned about the possibility that systematic differences in the popularity of baby names across states may lead to spurious association between state names 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 Social Security number in a different state would seem to address this concern, but unfortunately it does not.

First, as Pelham et al. (2002) acknowledged, Social Security numbers used to be obtained several years after birth. This means that some people received their Social Security number in a state other than their birth state. Because people who leave their birth state are disproportionately likely to move (back) there compared with 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 assessed whether there was evidence of implicit egotism after the confounding factors mentioned above were accounted for. I did this by assessing whether people are more likely to stay in a state, conditional on having obtained their Social Security number there, if their name is similar to that of the state; I found that they are not. In Study 9, in turn, I examined whether there is direct evidence that the two confounding factors mentioned above played a role in Pelham et al.'s (2002) original

[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

finding; I found that there is such evidence. 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 data set covering deaths up to December 31, 2006 (N = 80 million). Considering that the Social Security system was enacted in the United States in 1935, the analyses were conducted on the subset of people born after that year, in order to maximize the share of people obtaining the Social Security number in their birth state (N = 5.97 million).[18]

To assess name popularity across states, I compared the relative share of people obtaining their Social Security number (henceforth, "born") in a state similar to their first name with the share of all people born there. For example, I compared the percentage of all Georges born in Georgia with the percentage of all people not named George born there. For each target name and state, then, the full post-1935 SSDI data set was categorized into a 2 × 2 table.

To examine whether people are disproportionately likely to stay in a state with a name resembling their own, I created a new 2 × 2 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 5. The gray bars show the ratio of actual over expected frequencies ($R_{A/E}$s) for being born in the matching state, and the black bars show the ratio for staying there.

Seven of the eight names considered by Pelham et al. (2002) were more popular among people born in their corresponding state than expected by chance (i.e., compared with all other states), 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 an important confound.

One may worry that this finding is, in fact, the result of implicit egotism; people might simply be migrating to states with names resembling their own name before they obtain a Social Security number. 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 Social Security numbers; this means that since 1986, at the latest, the state where people obtain their Social Security number does coincide with birth state for the vast majority of U.S.-born babies.

The Social Security Administration releases top-100 baby names by state per year, which can be used to determine whether 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 5 were in the top 100 listing 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 Carolina, South Carolina, and Tennessee. It was not a top-100 name in any of the more distant northwestern or mountain states.

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 5, however, showed an $R_{A/E} < 1$, and the overall effect was $R_{A/E} = .99$, $\chi^2(1) = 0.56$, $p = .45$. How can the black bars in Figure 5 be reconciled with the Pelham et al. (2002) 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 more likely to move in?

If the reason for this pattern is that the movers' finding was spurious, as is argued above, then such a pattern should be found in placebo states too. More specifically, other states where a given first name was popular among people obtaining the Social Security number there should also have a disproportionate share of people moving into that state with that name. Study 9 shows that this is the case.

Study 9: Names Popular Among State Natives Are Popular Among State Immigrants

Method. Using the same post-1935 SSDI data set 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 United States who were named George and the percentage of all immigrants to each state in the United States who were named George.

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

Figures 6 and 7 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 corresponds 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 7 it can be seen that about 0.83% of people receiving the Social Security number in Georgia are named George, and about 0.82% of those moving to Georgia are so named.

If implicit egotism were behind the movers' finding in Pelham et al. (2002), we would expect a surprising number of movers into a state with a name similar to that state. Instead we find that the proportion 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 Pelham et al. (2002) is spurious.

Existing Finding 5: Last Names and States

In their Study 2, Pelham et al. (2002) examined 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 (from the no longer available

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

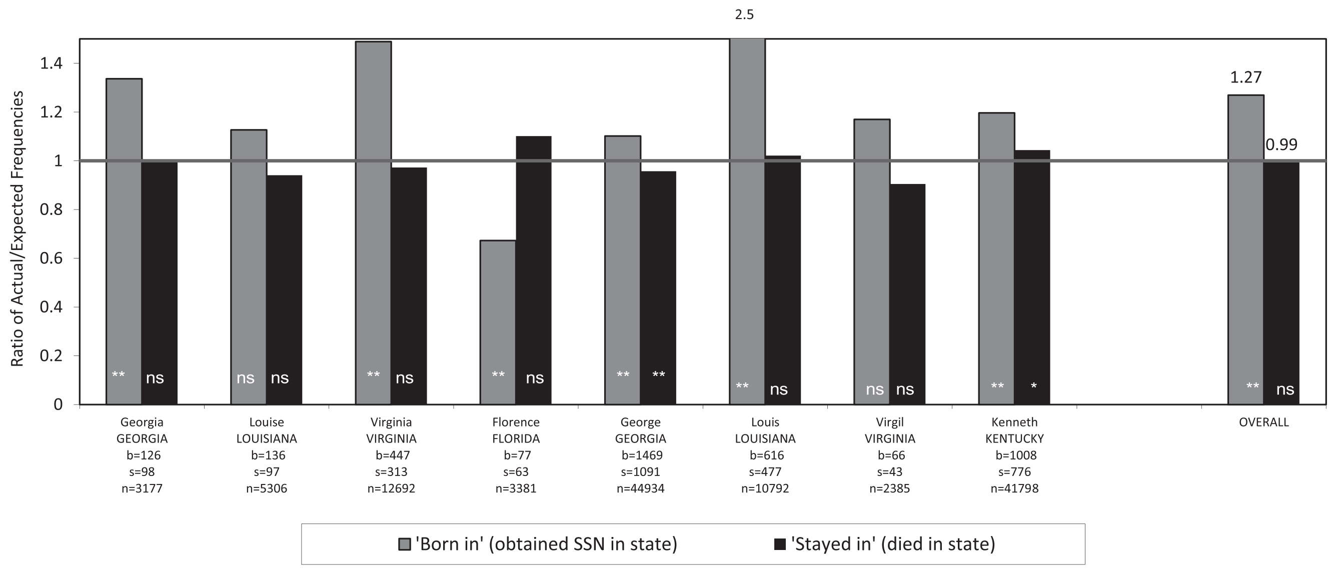

Figure 5. First names and states where people are "born" and "stay" (Study 8). From the Social Security Administration Death Index: subsample of people born after 1935 (N = 5.8 million). Because people obtain Social Security number (SSN) after birth, they may obtain it in a state other than their birth state, hence the use of the term "born." Ratio of actual over expected frequency ($R_{A/E}$) values and their significance are obtained from 2×2 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 was included in each table; for the black bars, the subset of people born in each state was included. b = the number of matching observations born in each state; s = the number of those staying in that state; n = the total number of people with that first name in the sample (e.g., there are 3,177 people with the name Georgia in the sample; 126 of them were born in the state of Georgia, and 98 of those stayed in Georgia). $^{*}p < .05$. $^{**}p < .01$. ($ns = p > .10$).

www.worldpages.com) for the frequency of listings of each last name across each of the eight states, obtaining an 8×8 table. Comparing expected and actual frequencies for this table, Pelham et al. found that 19.9% of listings are found in states with names resembling the corresponding last name, compared with 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 Pelham et al. (2002) is no longer available, one cannot verify whether 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 online versions (www.anywho.com, www.whitepages.com, and www.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 because of heterogeneity across states, there are no strong reasons to expect the distribution of last names to be the same across them. Study 10 here sought 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 Stay in a State With a Name Resembling Their Last Name

The data source for this study was 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 I focused on last names whose first three letters match the first three letters of the states used by Pelham et al. (2002; California, Texas, Florida, Illinois, Pennsylvania, Ohio, Michigan and Georgia). For each state here too I created a 2×2 table using the full data set for the "born" analyses, 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 as having obtained their Social Security number (born) or receiving their last benefit (staying) in the target state or not (e.g., California).

The results are shown in Figure 8. The gray bars indicate the ratio of actual over expected frequencies ($R_{A/E}$) for people born in a state, and the black bars show the ratios for those staying in those states. For example, the post-1935 SSDI data set includes 12,463 people with a last name beginning with Cal__, 9.7% of whom obtained their Social Security number 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 with the baseline of 76.6% of all other individuals obtaining their Social Security

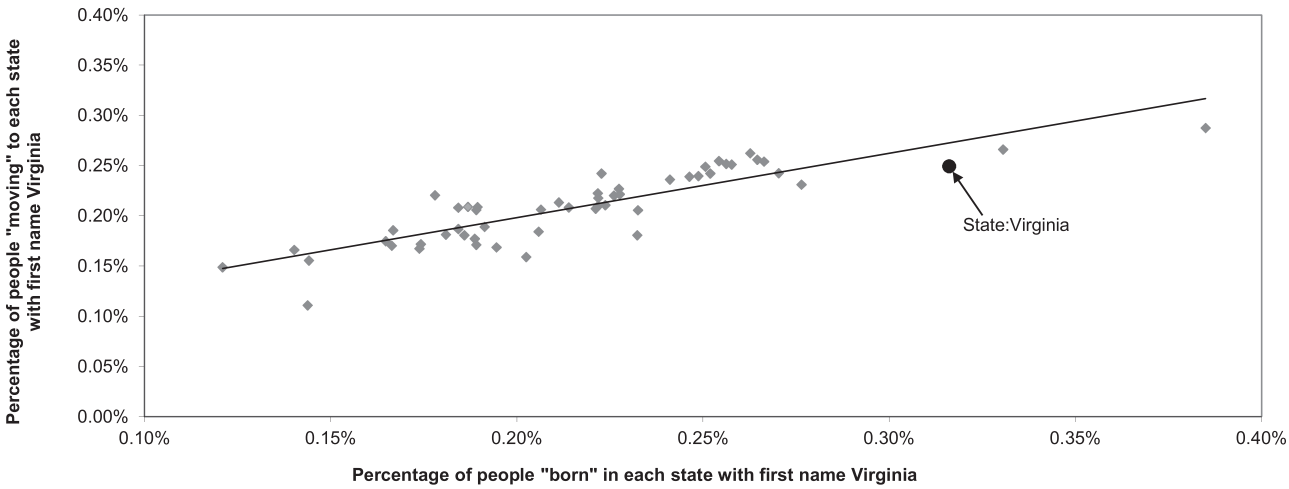

Figure 6. Share of people named Virginia "born" in and "moving" to each state (Study 9). From Social Security Administration Death Index: subsample of people born after 1935 (N = 5.97 million). Each dot in the figure represents a state. The x-axis indicates the percentage of people obtaining the Social Security number ("born") in that state whose first name is Virginia. The y-axis represents the percentage of people whose first name is Virginia who were born elsewhere but are receiving their last benefit (moving to) that state.

number 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 were 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 Social Security number (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 includes 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, whereas $R_{A/E} = 1.00$, $p = .97$, for staying there.

Existing Finding 5: Last Names and Towns

In their Study 1, Pelham et al. (2003) examined whether people are disproportionately likely to live in towns whose name contains their last name (e.g., Smiths living in Smithville). To test this, Pelham et al. queried the SSDI data set and compared, for the 30 most common last names in the United States, the overall fraction of individuals with that last name in the country, with

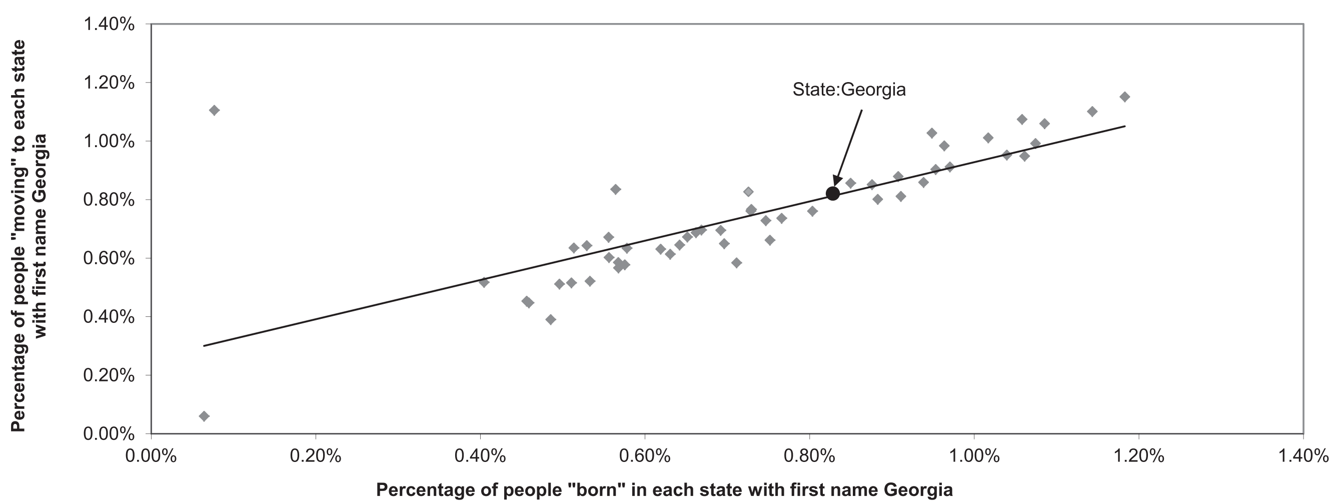

Figure 7. Share of people named George "born" and "moving" to each state George (Study 9). From the Social Security Administration Death Index: subsample of people born after 1935 (N = 5.97 million). Each dot in the figure is a state. The x-axis indicates the percentage of people obtaining the Social Security number ("born") in that state whose first name is George. The y-axis indicates the percentage of people whose first name is George who were born elsewhere but receiving their last benefit (moving to) that state.

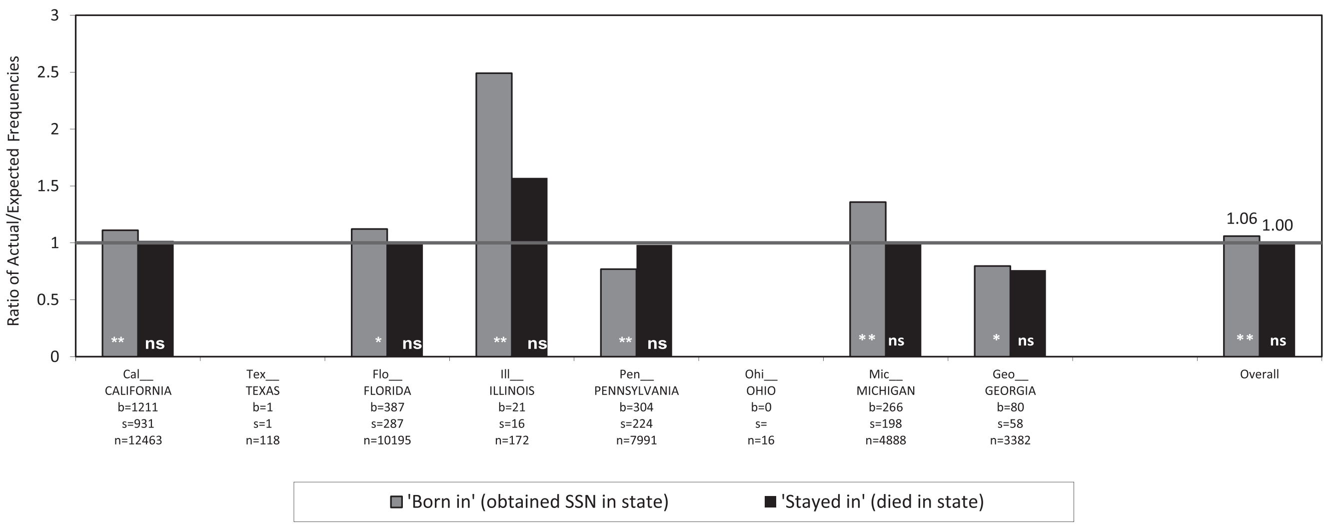

Figure 8. Last names and states where people are "born" and "stay" (Study 10). From the Social Security Administration Death Index of people born after 1935 ($N = 5.8$ million). Because people obtain Social Security number (SSN) after birth, they may obtain it in a state other than their birth state, hence the use of the term "born." Ratio of actual over expected frequency ($R_{A/E}$) values and their significance are obtained from 2×2 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 included in each table; for the black bars the subset of people born in each state is included. b = 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 last name in the sample (e.g., there are 12,463 with a Cal__ last name in the sample, 1,211 of them were born in California, and 931 of those stayed there). All significance tests were computed with respect to the null value of $R_{A/E} = 1$. $^{*} p < .05$. $^{**} p < .01$. ($ns = p > .10$).

the share of people with that last name in towns containing that last name in their name. For example, Pelham et al. (2003) compared the proportion of people with the last name Williams in the United States (.57%) with the proportion of Williams living in towns that contain "Williams" in their name (e.g., Williamsburg, VA; .64%). They found 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 were very small towns often having fewer than 1,000 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 have higher proportions of all nonethnic last names than the country as a whole does, 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 shows, 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

To examine the extent to which the set of towns used by Pelham et al. (2003) 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 data set (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 the individuals 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 research assistant 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 Pelham et al. (2003) examined 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 U.S. last names (in 1990): Smith, Johnson, Williams, Jones, Brown, and Davis. Concerned about reverse causality—for example, Mr. Smith living on Smith Road because he (or an ascendant of his) changed the name of the street to his last name—Pelham et al. (2003) examined a second set of six last names for which they presumed that this reverse causality was less plausible. In particular, they considered two last names that are used in street names to describe nearby places: Hill and Park; two last names that describe the type of road involved: Lane and Street: and two that are of historic significance: Washington and Jefferson.

Pelham et al. (2003) obtained the frequencies of people living on different streets from online phonebook listings across the entire United States and analyzed these in pairs. The analyses tested, 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, New York, for example, most streets change name at each intersection, and as was learned from phone conversations with the Lodi Historical Society, the name of the street within each block corresponds to the original owner of the property.

As mentioned above, the concern about reverse causality expressed by Pelham et al. (2003) led 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 obtained an effect about one eighth the size of that in the former suggests that reverse causality indeed played a large role in their results. However, whereas 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 United States).[19]

How ethnically diverse are the last-name pairs used by Pelham et al. (2003)? Their diversity ranges from moderate to extreme. Census figures indicate that within any given pair, the relative frequency of African Americans differs between 30% and 50% between the two last names (e.g., 22% of Smiths are Black compared with 34% of Johnsons). The most diverse pair is Park (66% Asian) and Hill (<1% Asian). In Study 12 I attempted to examine the impact of people's names on the streets on which they live in net of any reverse causality and ethnic confound effects by focusing on people's first names and on last names that cannot possibly have influenced the name of the streets.

Study 12: No Name Effect on Streets from First Names or Geographic Designators

Method. In order to rule out reverse causality, similar rather than identical last names could be used, as was done in Study 3 (e.g., looking at whether people last-named Smithers live on Smith Street). Because such last-name variations are too infrequent, I used similar first names instead, assessing, for example, whether men named William disproportionately live on Williams Avenue.[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 hand, 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 examined last names that begin with streets' orientation designators (West, East, North, or South); such designators are not chosen on a street by street basis but rather are 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. This file 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, I chose first names that closely resemble the last names used in Pelham et al. (2003). 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 used the first name of the respective historic figures (George and Thomas).[21] For West, East, North, and South, I included all last names that begin with those strings of letters (e.g., John Westfield, Oliver North).

Results. I began by attempting to replicate the results from Pelham et al. (2003) on the New York data, that is, examining whether the 12 last names used by Pelham et al. (2003) are more commonly observed living on streets with those last names than would be expected by chance. Ten of the 12 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). As was the case in the data analyzed by Pelham et al. (2003), the second set of names has a much lower $R_{A/E}(1.08)$ in the New York data. These results suggest that any differences between the conclusions from the new and the old analyses do not stem from data differences.

The results for the new analyses are reported in Figure 9. Eight of the 11 name–street combinations have an $R_{A/E}$ 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 frequency is 8). Aggregating across all combinations, $R_{A/E} = .92$, $p < .001$, indicating, if anything, a reversed pattern.

[19] Data are from http://www.melissadata.com/lookups/

[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.

[21] For Williams, I included William, Will, Willie, and Willy; for Jones, I include Jon and Jonathan. For Jefferson, I included 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.

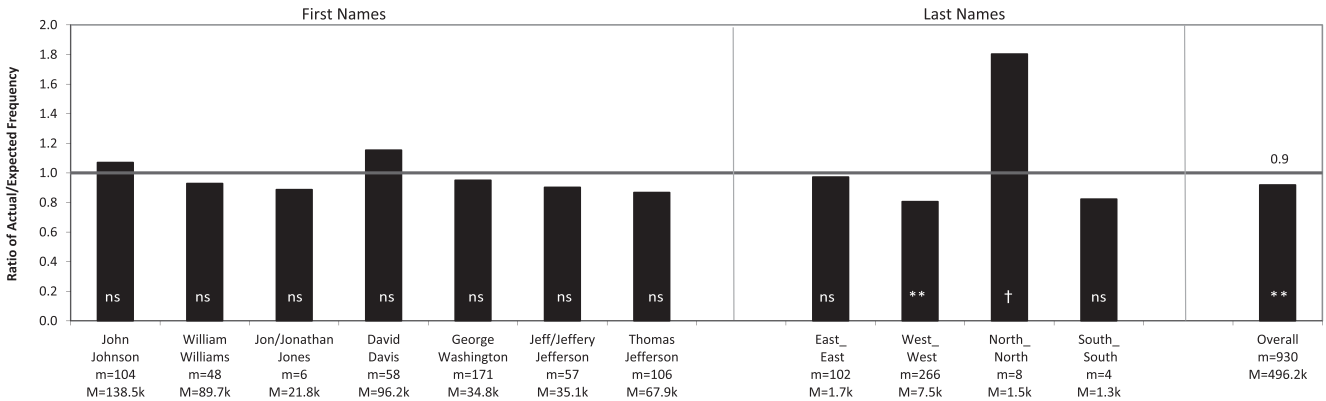

Figure 9. No effect of name on street choice (Study 12). Ratios of actual over expected frequency ($R_{A/E}$) values and their significance are obtained from 2 × 2 tables that classify people as living in the target street or not and as having the target name or not. The entire sample (New York voter registration file eliminating repeated addresses, $N = 8.9$ million) is included in each table. m = the number of matching observations (e.g., 104 men name John live on a Johnson street); M = the total number of people with that name in the sample (e.g., there are a total of 138.5 thousand Johns in the New York sample). $^{*} p < .05$. $^{**} p < .01$. ($ns = p > .10$).

Existing Finding 8: Birthday Numbers

In their Study 6, Pelham et al. (2002) examined 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 2nd) 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 in their names, so Pelham et al. (2002) focused on birthdays and towns with names ranging between two and eight. They used the SSDI data set (described above) and conducted 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 Pelham et al. (2002) found that of the 485 people in their sample, 94 died in a town matching their birthday, compared with 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 12 (Kitayama & Karasawa, 1997). Starting from such a weak effect, it would seem that a rather large data set 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 used the same SSDI data and analyses used in Pelham et al. (2002) but asked 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 of 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 data set described above. The replications 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 at 2 Elm Street or in Apartment2?

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 the option to move to another town, and very few towns have numbers in them, whereas all addresses do. Second, addresses have numerals in them, whereas town names have the number represented with words. Lastly, and related to the first point, the data set 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

The entire SSDI data set was used for this analysis, 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; died in Two Oaks) compared with the 954.5 individuals that would be expected by chance, implying a ratio of actual over expected frequency of $R_{A/E} = 1.01$, $p = .46$. Pelham et al.'s (2002) 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

The data consisted of the New York voter registration file described in Study 12 ($N = 12.8$ million). I performed three separate analyses: one for street names, one for address numbers, and one for apartment numbers. In each of these, an address was 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 would consider 4 Broadway Avenue a match, and the apartment analyses would denote Apartment 4). I replicated the analyses considering also multidigit matches obtaining identical results (e.g., 44th avenue, 4444 Broadway Avenue, and Apartment 444 would all be considered matches). Each analysis created eight 2×2 tables, where every individual was categorized as born on a given date or not and as living at the matching address or not.

The results are presented in Figure 10. The *n* values indicate the number of matching observations (e.g., there are 1,827 people living at an address number that matches their birthday day and month), and *N* indicates the number of total observations. Note that *n* is between 8 and 20 times larger than in the original studies, and *N* is 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-square tests (with $df = 1$) was significant at the 5% level, two were significant at the 10% level, one was in the direction implied by implicit egotism (March 3rd and Third Avenue), and one was in the opposite direction (February 2nd and Second Avenue). The overall $R_{A/E}$s were .98, 1.01, and 1.01 for address, apartment, and street, respectively, none of them being significantly different from 1 (ps = .74, .73 and .50, respectively).

In sum, the only result that was free of confounds across the three implicit egotism articles considered did 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. Although it is not possible to directly test whether the original finding is due to sampling error, the evidence suggests that 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 (2008) analyzed a sample containing a third of all full- time employees in Belgium ($N = 528{,}007$). The data set contains the first three letters of employees' last names and of the name of the company for which they work. The authors assessed 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 was $R_{A/E} = 1.13$.

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, although the authors excluded self-employed individuals from their analyses, they still included people who belonged to a firm named after themselves if other people worked there also, such as law firms with more than one partner.

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

Through e-mail communications I suggested to one of the authors (Anseel) that they attempt to replicate their findings excluding people who share all three letters with the company (avoiding reverse causality) and using as controls letter combinations that are highly correlated with each other in their choice of employers (as in Study 5 here with first names). As of the publication of this article, I do not believe they have conducted such analyses. A few weeks following the final acceptance of this article, I obtained data from American (rather than Belgian) employees that allowed me to assess the role of reverse causality on the last name initial and employer results. After replicating their main findings, new analyses suggested they were indeed likely to be spuriously caused by reverse causality (see Simonsohn, in press).

Initials and Performance (Nelson & Simmons, 2007)

Nelson and Simmons (2007) included four field studies in their investigation. One showed 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 showed that people with first names starting with a C or a D obtained worse academic outcomes than other students. The authors' fifth study was an experiment, alluded to in the introduction of this article.

The academic performance studies do not share the concerns put forward elsewhere in this article, as they include controls for age, country of origin, and race. In addition, in one of their studies, Nelson and Simmons (2007) found that only students with a name beginning with a C or a D who also like their own initial obtained a lower grade point average (GPA) than other students.

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

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

Chandler et al. (2008) studied donations to the Red Cross. They examined whether people who share an initial with a hurricane name are more likely to donate to its victims than are other people. To test this prediction, Chandler et al. compared the share of donors who

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

[23] McCullough and McWilliams (2010) found 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.

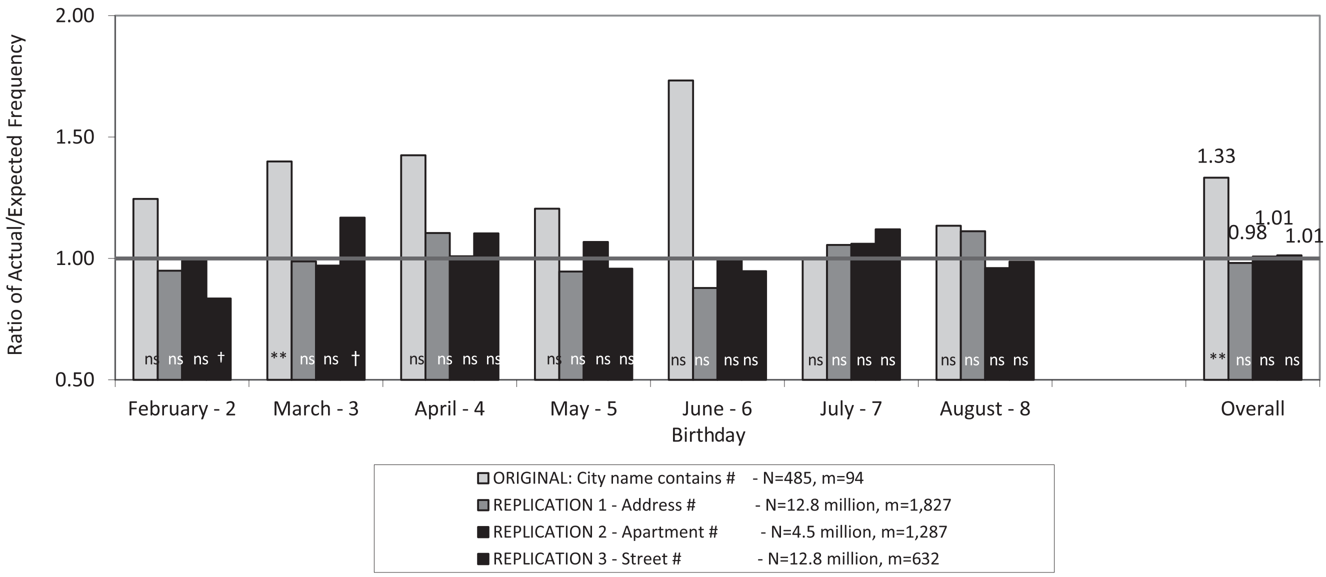

Figure 10. Birthday and location (Study 14). Original data are from the Social Security Administration Death Index: subset of people born on a matching day–month birthday (e.g., February 2nd) and dying in a town with a number between 2 and 8 in its name. Replication results are from the New York Voter Registration file (N = 12.8 million). Ratios of actual over expected frequency ($R_{A/E}$) values and their significance, for the three replications, are obtained from 2 × 2 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 = the number of matching observations (e.g., 1,827 people live at an address that matches their birthday month and day). $^{*}p < .05$. $^{**}p < .01$. ($ns = p > .10$).

share an initial with a hurricane name just before and after the hurricane struck. For every hurricane considered, the authors found such share increased posthurricane. 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 (i.e., 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?

In the present article I reevaluated evidence that seemed to show that implicit egotism, the liking of things connected to the self, can influence marriage, occupation, and moving decisions, and found 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 influence 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 among 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, then, may implicit egotism not matter for marriage, occupational, and location decisions? 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 versus overall stakes. Although home buying is a large stakes decision, the decision of which specific home to buy, for example, 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 Pelham et al. (2002), Pelham et al. (2003), and Jones et al. (2004) do not, arguably, have small marginal stakes. Few people are nearly indifferent or ambivalent about the career, job, and spouse they chose and their second most preferred option (and fewer still are indifferent regarding a spouse, career, and location that does versus does not resemble their name). The number of Georges who are nearly indifferent or are ambivalent with regard to living in Georgia versus another state may be too small for the impact of implicit egotism in their choice to be detectable.

The article with the most convincing evidence of implicit egotism in the field (Chandler et al., 2008), in contrast, studied a decision that although important (giving charity to needy victims), involved choices over which people are likely to be 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. In future work, researchers should strive to identify other domains where decisions in the field have small marginal stakes (though ideally large total stakes) and examine whether implicit egotism plays a role in those decisions.

On the Usage of Quantity of Evidence as Evidence

The results from the present article also speak to a common argument used to defend the validity of explanations put forward in multiexperiment articles. It is often proposed that because the authors' main thesis is the only parsimonious explanation for all studies in a given article, 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 Pelham et al. (2002), Pelham et al. (2003), and Jones et al. (2004) 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 by-product of the path chosen to investigate it.

Concluding Remark

Pelham et al. (2002), Pelham et al. (2003), and Jones et al. (2004) 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 article suggests are, in fact, behind their findings. Had Pelham and colleagues not so openly shared their thinking, data sources, and results, this article probably would not have been written. For what it 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.

References

Anseel, F., & Duyck, W. (2008). Unconscious applicants: A systematic test of the name–letter effect. *Psychological Science, 19,* 1059–1061. doi:10.1111/j.1467-9280.2008.02199.x

Baron, R. A., Branscombe, N. R., & Byrne, D. R. (2008). *Social psychology*. Boston: Allyn & Bacon.

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

Bialik, C. (2010, May 15). 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., & O'Connell, J. W. (1975). Sex bias in graduate admissions: Data from Berkeley. *Science, 187,* 398–404. doi:10.1126/science.187.4175.398

Breckler, S. J., Olson, J., & Wiggins, E. (2005). *Social psychology alive.* Belmont, CA: Wadsworth.

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,* 405–415. doi:10.1086/497552

Chandler, J., Griffin, T. M., & Sorensen, N. (2008). In the "I" of the storm: Shared initials increase disaster donations. *Judgment and Decision Making, 3,* 404–410.

Cochran, W. G. (1954). Some methods for strengthening the common ×2 tests. *Biometrics, 10,* 417–451. doi:10.2307/3001616

DaVanzo, J. (1983). Repeat migration in the United States: Who moves back and who moves on. *Review of Economics and Statistics, 65,* 552–559. doi:10.2307/1935923

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,* 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,* 1233–1260. doi:10.1162/003355301753265561

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,* 665–683. doi:10.1037/0022-3514.87.5.665

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,* 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,* 736–742. doi:10.1177/0146167297237006

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,* 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 savings behavior. *Quarterly Journal of Economics, 116,* 1149–1187. doi:10.1162/003355301753265543

McCullough, B. D., & McWilliams, T. P. (2010). Baseball players with the initial "K" do not strike out more often. *Journal of Applied Statistics, 37,* 881–891. doi:10.1080/02664760902889965

Myers, D. (2009). *Social psychology* (10th ed.). Boston, MA: McGraw-Hill.

Nelson, L. D., & Simmons, J. P. (2007). Moniker maladies: When names sabotage success. *Psychological Science, 18,* 1106–1112. doi:10.1111/j.1467-9280.2007.02032.x

Nuttin, J. M., Jr. (1985). Narcissism beyond Gestalt and awareness: The name letter effect. *European Journal of Social Psychology, 15,* 353–361. doi:10.1002/ejsp.2420150309

Nuttin, J. M., Jr. (1987). Affective consequences of mere ownership: The name letter effect in 12 European languages. *European Journal of Social Psychology, 17,* 381–402. doi:10.1002/ejsp.2420170402

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,* 800–807. doi:10.1037/0022-3514.85.5.800

Pelham, B. W., Carvallo, A., & Jones, J. T. (2005). Implicit egotism. *Current Directions in Psychological Science, 14,* 106–110. doi:10.1111/j.0963-7214.2005.00344.x

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,* 469–487. doi:10.1037/0022-3514.82.4.469

Sanderson, C. A. (2009). *Social psychology*: Hoboken, NJ: Wiley.

Schacter, D. L., Gilbert, D. T., & Wegner, D. M. (2010). *Introducing psychology*: New York: Worth.

Simonsohn, U. (in press). Spurious also? Name similarity effects (implicit egotism) in employment decisions. *Psychological Science*.

Simonsohn, U., & Loewenstein, G. (2006). Mistake #37: The impact of previously observed prices on housing demand. *Economic Journal, 116,* 175–199. doi:10.1111/j.1468-0297.2006.01052.x

Thaler, R. H., & Benartzi, S. (2004). Save More Tomorrow(TM): Using behavioral economics to increase employee saving. *Journal of Political Economy, 112,* S164–S187. doi:10.1086/380085

Word, D. L., Coleman, C. D., Nunziata, R., & Kominski, R. (2009). *Demographic aspects of surnames from Census 2000.* Retrieved from http://www.census.gov/genealogy/www/data/2000surnames/surnames.pdf

Appendix

Frequencies Used for Creating Figures in the Text

Table A1
Tabulations for Figure 1

County/variable	Actual	Expected	Ratio
Walker County			
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

Table A2
Tabulations for Figure 2

	Exact same last name				Similar last name			
	Control		Target		Control		Target	
Groom name	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
Total			2,330	11,005			680	10,325

(*Appendix continues*)

Table A3
Tabulations for Figure 3: 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

Table A4
Tabulations for Figure 3: 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

Table A5
Tabulations for Figure 5

	"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	12,245	101,497	39,469	313	134
Florence	202,228	5,771,762	77	3,304	150,242	51,986	63	14
George	175,909	5,756,528	1,469	43,465	136,566	39,343	1,091	378
Louis	135,375	5,831,204	616	10,176	102,650	32,725	477	139
Virgil	141,347	5,833,639	66	2,319	101,767	39,580	43	23
Kenneth	119,491	5,816,082	1,008	40,790	88,105	31,386	776	232

(*Appendix continues*)

Table A6
Tabulations for Figure 8

	"Born"				"Stayed"			
	Controls		Target first name		Controls		Target first name	
Last names	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	137,599	39699	58	22

Table A7
Tabulations for Figure 9

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

(*Appendix continues*)

Table A8
Tabulations for Figure 10

	Everyone else in New York state		People with target birthday	
Date	Matching apartment no.	Other apartment no.	Matching apartment no.	Other apartment no.
	Apartment numbers (e.g., Apartment 2)			
February	13,042	4,275,816	605	208,875
March	12,828	4,387,950	281	97,279
April	12,500	4,439,395	144	46,299
May	13,108	4,457,474	77	27,679
June	13,170	4,461,578	61	23,529
July	13,238	4,465,232	62	19,806
August	13,359	4,467,778	57	17,144
	Address (e.g., 2 Elm Street)			
February	37,274	12,730,578	258	88,915
March	36,776	12,733,317	243	86,689
April	35,932	12,731,658	254	89,181
May	36,629	12,726,729	287	93,380
June	36,848	12,730,289	257	89,631
July	37,418	12,728,570	283	90,754
August	37,673	12,732,897	245	86,210
	Street numbers (e.g., 2nd avenue)			
February	37,513	12,811,635	93	38,028
March	36,975	12,810,720	133	39,441
April	36,188	12,825,951	78	25,052
May	36,898	12,808,936	114	41,321
June	37,146	12,830,017	55	20,051
July	37,707	12,823,971	84	25,507
August	37,940	12,823,564	75	25,690

Received April 30, 2010
Revision received September 19, 2010
Accepted September 24, 2010 ■