8-2011

Spurious Also? Name-Similarity Effects (Implicit Egotism) in Employment Decisions

Uri Simonsohn
University of Pennsylvania

Spurious Also? Name-Similarity Effects (Implicit Egotism) in Employment Decisions

Abstract

Implicit egotism is the notion that major life decisions are influenced by name-similarity. This paper revisits the evidence for the most systematic test of this hypothesis. Anseel & Duyck (2008) analyzed data from 1/3 of all Belgian employees and found that a disproportionate fraction of them shared their initial with their employer. Using a dataset with American employees I replicate the finding, but new analyses strongly suggest they are due to reverse causality, whereby the documented effect seems to be driven by people naming companies they start after themselves rather than by employees seeking out companies they have a shared initial with. Walt Disney, for example, worked for a company starting with D (Disney World) not because of an unconscious attraction to such letter, but because the company was named after him.

Disciplines
Business | Cognition and Perception | Cognitive Psychology | Marketing

SPURIOUS ALSO? NAME SIMILARITY EFFECTS (IMPLICIT EGOTISM) IN EMPLOYER DECISIONS

November 2010

Uri Simonsohn
The Wharton School
University of Pennsylvania
uws@wharton.upenn.edu

Word count: 964

Abstract
Implicit egotism is the notion that major life decisions are influenced by name-similarity. This paper revisits the evidence for the most systematic test of this hypothesis. Anseel & Duyck (2008) analyzed data from 1/3 of all Belgian employees and found that a disproportionate fraction of them shared their initial with their employer. Using a dataset with American employees I replicate the finding, but new analyses strongly suggest they are due to reverse causality, whereby the documented effect seems to be driven by people naming companies they start after themselves rather than by employees seeking out companies they have a shared initial with. Walt Disney, for example, worked for a company starting with D (Disney World) not because of an unconscious attraction to such letter, but because the company was named after him.

(main document pp.1-7)

(supplemental materials 8-14)

Implicit egotism is the notion that name-similarity influences major life decisions (Pelham, Carvallo, & Jones, 2005). Anseel and Duyck (2008) put forward its most systematic test to date: Using data from a third of all Belgian full-time employees they examined if people disproportionately worked for companies matching their last name initial (e.g., Tom <u>D</u>ubois working for <u>D</u>istrigaz), finding this was the case for every letter in the alphabet. Overall, their sample of 582,007 employees contained 36,242 with a matching company initial, compared to the 31,952 that would be expected by chance. Extrapolating to the rest of the population they concluded that at least 12,000 Belgians chose their employer *because* they shared an initial with them.

One alternative explanation for these findings is reverse causality: Rather than employees seeking out companies with similar names, people starting new companies may name them after themselves. Walt <u>D</u>isney, for example, worked for a company starting with D (<u>D</u>isney World) not because of an unconscious attraction to such letter, but because the company was named after him.

Seeking to address this possibility Anseel and Duyck "exclud[ed] self-employed people" (p.1060), but many who work for companies they or their relatives named are not self-employed; e.g., neither Walt Disney nor Henry Ford were "self-employed." This confound gets exacerbated in small family firms that often hire founders' relatives.

A second alternative explanation is an ethnic/language confound: <u>Van</u>Boven works for <u>Van</u>Dyke Associates while <u>Le</u>Boeuf for <u>Le</u>Blanc Associates because the former lives in Dutch speaking Flanders and the latter in French speaking Walloon, and *Van* is a common Dutch preposition while *Le* a French one.

Unfortunately Anseel & Duyck are not allowed to share their data, I therefore employ a new (American) dataset (hence free of the language confound) to reassess the evidence for a name-similarity-effect in employer choice. In the new data I successfully replicate their findings but once reverse causality is controlled for the name-similarity-effect disappears *entirely*. [1]

METHOD AND RESULTS

In the U.S., donors to political campaigns must disclose their name and employer. Datasets with such information are freely available from sources such as the Center for Responsive Politics (http://www.opensecrets.org) from where I obtained data for the 2004 election cycle (N=2.52 million). Retaining one observation per donor, eliminating incomplete entries, and those of individuals self or not currently employed, led to N=438,111.

Three sets of analyses were conducted. All involved comparing expected with actual frequencies of employee-company pairs with a matching initial. Expected frequencies are calculated, as in (Anseel & Duyck, 2008), multiplying the proportions of employees and companies with that initial.

Overall effects are obtained by adding actual and expected frequencies across all initials, leading to a simple $\chi^2(1)$, (Cochran, 1954). Results from all analyses are reported in Figure 1, expressed as the ratio of actual over expected frequencies ($R_{A/E}$).

*** Fig 1 ***

---

[1]See supplemental materials for an assessment of the role of the language confound in the Belgian data and for a suggested analysis that could control for it.

The black bars show that, for every initial, actual frequencies were higher than expected: people do disproportionately share initials with their employers. [2] Overall, more than two-and-a-half as many people work for an employer with a matching initial as would be expected by chance, $R_{A/E}=2.60$, $\chi^2(1)=56,240.8$, $p<.0001$. The effect is notably *larger* than that reported in Anseel & Duyck, possibly because the U.S. sample is likely to over-represent wealthier individuals for whom reverse causality is more pronounced.

If the matching-initial effect is due to reverse causality, then employees should tend to share more than just an initial with their employers; while implicit egotism would predict that both Tom Peeters and Pieters are disproportionately likely to work for *Peeters and Associates*, reverse causality could only account for the former. The second set of analyses, therefore, focus on employees matching an initial but not all first three letters with their employee; see light gray bars. The difference with the black bars is stark: All $R_{A/E}$'s are close to 1, and the overall ratio is $R_{A/E}=0.99$. There is absolutely no initials-effect.

If people sharing at least three letters entirely account for the black bars, then there must be an astonishingly greater than expected number of them. Overall 61,025 people match their company's initial in the sample, compared to the 23,491 expected by chance; a surplus of 37,534 individuals. We only expect 596 employees to share at least three letters with their employers, if instead 37,534 do, then the name-similarity-effect is a bewildering $R_{A/E}=37,534/596=64.4$.

It seems implausible that implicit egotism would lead to a nil increase for sharing one or two letters, but to a sixty-four fold one for sharing three or more. The pattern is consistent, however, with reverse causality.

The third set of analyses are analogous to the first, but are performed on the subset of companies with at least 50 observations (N=54,710); see darker gray bars. The premise being

---

[2] The letters Q,X,Y&Z combined have an expected frequency of just 17 and are hence excluded from the figure.

that if the effect is driven by reverse causality it should be present primarily among smaller firms where such mechanism is more likely and consequential (as it affects a greater proportion of employees). Also here $R_{A/E}$'s are dramatically closer to 1 and the overall effect is $R_{A/E}=1.01$, $\chi^2(1)=.30$, $p=.583$.

CONCLUSIONS

The results strongly suggest the existing evidence for implicit egotism in employer choice is driven by reverse causality. In questioning the implicit egotism interpretation of name-similarity-effects this research is related to (Simonsohn, in press), which did so for earlier findings on marriage, occupation and moving decisions.

These findings do not mean that implicit egotism is not a real psychological phenomenon, but given that the effect is of moderate size in the laboratory, settings where people are closer to indifference among options are more likely to lead to detectable effects outside of it.

Figure 1. Ratio of actual/expected ($R_{A/E}$) frequencies for matching employee-employer initial
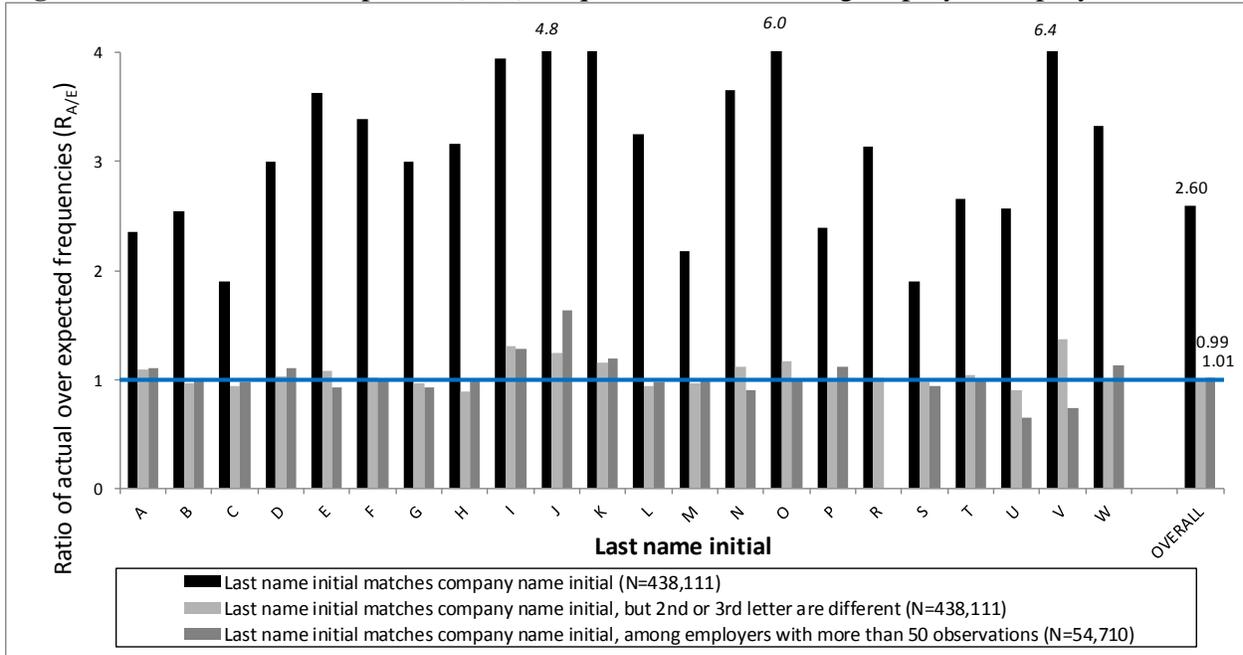
Ratios censored at $R_{A/E}=4$, censored value printed above bar.
Source: Author's calculations on Individual Political Contributions Data File, for 2004, from the Center for Responsive Politics.

References

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

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

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

Simonsohn, U. (in press). Spurious? Name Similarity Effects (Implicit Egotism) in Marriage, Job, and Moving Decisions. *Journal of Personality and Social Psychology*.

"ASSESSING AND FIXING THE LANGAUGE CONFOUND IN THE BELGIAN DATA"

SUPPLEMENTAL MATERIAL FOR:

SPURIOUS ALSO? NAME-SIMILARITY-EFFECTS (IMPLICIT EGOTISM)
IN EMPLOYER DECISIONS

Uri Simonsohn
The Wharton School
University of Pennsylvania
uws@wharton.upenn.edu

This supplemental material (i) discusses in some detail the likely role ethnic/language sorting is playing in the analyses of (Anseel & Duyck, 2008), and (ii) proposes new analyses that could eliminate such confound.

(i) Diagnosing the Ethnic/language confound in Belgian Data.

The data from (Anseel & Duyck, 2008) is from Belgium. Belgium contains two main geographical regions, Flanders, where people speak Dutch, and Walloon, where they speak French. This geographical segregation gives rise to a very likely confound that could lead to employers sharing an initial with employees in the absence of implicit egotism.

To gain an intuition for the problem let's begin with an extreme scenario; suppose every person in Flanders had a last name beginning with "Van" (e.g., VanBoven, VanDyke, etc.) and every person in Wallonia one with "Le" (e.g., LeBoeuf, LeBlanc, etc.).
'If' even a small number of people named companies after individuals, and 'if' even a small number of people preferred to work in the region of Belgium in which they live rather than in the other, then we would see a disproportionate share of "Van" individuals working in "Van" companies and "Le" individuals working in "Le" ones.

How different are Dutch and French last names in Belgium? Employing the online phonebook www.infobel.com/en/belgium I searched the number of listings with last names beginning with each letter in the alphabet in two cities of in Belgium, one in Flanders (Ghent) the other in Wallonia (Charleroi).

The results are tabulated in Table A1. The overall correlation in frequency across initials is $r = .84$. While high, it is not close to $r = 1$, the implicit assumption in analyses that treat Belgium as a single uniform country. In fact, difference of proportions tests contrasting the

shares of last names starting with each initial across the two cities reject the null that they are the same, at the 5% level, for 21 of the 26 letters. An overall $\chi^2(26)$ that the distributions are identical is comfortably rejected (p<.00001).

How big a bias can we expect from these dissimilar distributions? In other words, if there was no implicit egotism and one estimated expected initial matchings from overall Belgian data without controlling for ethnicity/language, how big a spurious ratio of actual over 'expected', $R_{A/E}$, might one get?

Assume the distribution of business initials was identical to that of last names within each city and that people only work in the city they live. This would mean that within each city we would see the percentage of each letter squared with a matching initial. For example, in Ghent, 2.29% of last names start with A, so 2.29%*2.29% of all people in Ghent should have a last name and an employer starting with the letter A. In Charleroi 3.4% of last names start with A so that number is 3.4%*3.4%. If we repeat this exercise for each letter and add the numbers we will get the percentage of people that will have a matching initial in the absence of implicit egotism.

We can then contrast this number with what one would 'expect' if one neglected geography and computed overall frequencies for the two cities combined rather than separately, as in (Anseel & Duyck, 2008). For example 2.6% of all last names in the sample of the two cities start with A, so 2.6%*2.6=.000676 is the expected rate (neglecting geography). Overall there are 156,184 people in the sample so that would mean that we would expect .000676*156184=108.2 A-A matchings. But once we take into account geography we would

expect to actually see 2.3%*2.3%* 106140 =55.6 in Ghent, 3.4%*3.4%*50044=56.4, a total of 112, leading to a $R_{A/E}$ for A's of 112/108.2=1.04.

That is to say, if we compute expected frequencies for As ignoring language differences and if people do randomly choose employers but within their city, we will find a completely spurious $R_{A/E}$=1.04 for the letter A. If we repeat this exercise for every letter we obtain an overall $R_{A/E}$ = 1.05, which is roughly half the effect size reported in (Anseel & Duyck, 2008).

So if controlling for reverse causality in the Belgian data reduced the effect size in half (it does much more than that in the current American data) we may expect the other half to disappear once we take into account the language confound.

(ii) Fixing the ethnic confound in the Belgian data.

One tempting fix to the confound is to replicate the analyses by (Anseel & Duyck, 2008) on a city by city basis, but the problem is likely to persist because presumably even in Dutch speaking cities there will be areas with more and less French speaking Belgians.

One alternative is to use a different dataset where the language confound is likely to be minimal. The American data used in the present paper is one such case and of course the result is that there is no name-letter-effect there.

Another alternative is to use the technique used in Study 3 in (Simonsohn, in press). I suggested this analysis to Frederik Anseel, via email communications, in June of 2010.

In a nutshell one identifies initials (or rather two-first-letter strings) that are highly correlated in their matchings of employers and use those as controls. In particular, one can do the following 6 steps:

11

1) Eliminate all observations where employee and employer share all three letters (this is conservative but eliminates reverse causality).

2) Create a variable that includes the first two letters of a last name, and another with the first of the employer name (each variable can take 26*26=676 values: "AA", "AB", "AC"… "ZZ")

3) Create a table of 676x676 reporting the co-occurrence of each combination of letters (e.g., how many AA_ work for an AA_ employer, etc)

4) Compute correlations for the columns and rows in that matrix so that we know how correlated each combination of letters is with which. For example, we would know the correlation in employer names of employees with a last name starting with VA_ and LE_.

5) Select the top-3, say, most correlated two-letter-strings from each two-letter string (e.g., the top-3 most correlated strings for VA_ names, and the top-3 most correlated strings for VA_ employers).

6) Now test if a given letter combination is more likely to work for the matching string than the controls. This could be done for the top-20 most common last name two-letters to avoid a large number of computations.

For example, let's think of Van_ last names. We first eliminate from the dataset all Van_ individuals working for Van_ companies to eliminate reverse causality. We will find the top-3, say, most correlated two-letter initials of last names starting with Va_.

Let's say these are Pe_ Kr_ and Vo_. We then find the most correlated employer two-letter initials for Va_, let's say those are Ab_ De_ and Po_. We then test, if Va_ more likely to work for Va_, rather than for Ab_ De_ and Po_, than Pe_ Kr_ and Vo_ combined are.

This test is more valid, the higher the correlation between two letter strings and their controls are. In Study 3 of (Simonsohn, in press) the correlations were above $r = .99$.

Table A1. Number of phone listings with each initials in two cities in Belgium.

	Gent	Charleroi
A*	2430	1680
B*	9132	4736
C*	6375	3789
D*	22181	7443
E*	1250	593
F*	1070	1656
G*	3931	2883
H*	5169	2030
I*	486	326
J*	985	746
K*	2063	834
L*	4031	3623
M*	6204	4204
N*	1337	868
O*	1031	423
P*	3606	2866
Q*	112	146
R*	3352	2018
S*	6641	2974
T*	2216	1495
U*	253	127
V*	18396	3187
W*	3303	938
X*	8	8
Y*	248	156
Z*	330	295
Total	106140	50044

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

Simonsohn, U. (in press). Spurious? Name Similarity Effects (Implicit Egotism) in Marriage, Job, and Moving Decisions. *Journal of Personality and Social Psychology*.