AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Personality and Social Psychology:
Attitudes and Social Cognition

ISSN: 0022-3514

https://doi.org/10.1037/pspa0000347

INNOVATIONS IN SOCIAL PSYCHOLOGY

Does the First Letter of One's Name Affect Life Decisions? A Natural Language Processing Examination of Nominative Determinism

Promothesh Chatterjee, Himanshu Mishra, and Arul Mishra
David Eccles School of Business, University of Utah

This research examines whether the phenomenon of nominative determinism (a name-driven outcome) exists in the real world. Nominative determinism manifests as a preference for a profession or city to live in that begins with the same letter as a person's own name. The literature presents opposing views on this phenomenon, with one stream of research documenting the influence and another stream questioning the existence and generalizability of the effect, as well as the proposed underlying process. To examine whether the effect occurs in the real world, we use large language models trained on Common Crawl, Twitter, Google News, and Google Books using two natural language processing word-embedding algorithms (word2vec and GloVe). After controlling for relevant variables, we find consistent evidence of the relationship between people's names and a preference for major life choices starting with the same letter as their first name. Our theoretical framework of identity expression builds on the implicit egotism explanation.

Keywords: implicit egotism, nominative determinism, text analysis, word embedding

Extant research suggests that people show a preference for letters in their own names (Nuttin, 1985). This preference has been documented in multiple countries using different evaluative methods (Koole & Pelham, 2003) and serves as a measure of implicit self-esteem (Hoorens, 2014). Subsequent research suggests that this preference for letters in one's name can affect several important life decisions, such as the choice of profession or the city in which to live (Anseel & Duyck, 2008; Jones et al., 2004; Pelham et al., 2002). We refer to this phenomenon as "nominative determinism"—literally "name-driven outcome" (Alter, 2013; Stekel, 1911) to distinguish it from the name–letter effect, or the tendency to evaluate alphabetical letters in one's name positively. In contrast with research on the downstream influence of names that focuses on both surnames and first names, our focus here is exclusively on first names to help address alternative explanations related to reverse causality (e.g., people with the last name Disney working at the Walt Disney Company). Our conceptualization of nominative determinism uses the first name because this name usually stays with people throughout life and can be perceived as conveying the sum total of who they are. Therefore, in this research, we examine the presence of nominative determinism as it indicates people's preference for profession or city names that begin with the same letter as the first letter of their first name. For example, nominative determinism would suggest that a person named Dennis is more likely to choose to be a dentist than, say, a lawyer, or that Dennis is more likely to choose to live in Denver than Cleveland. One feature of nominative determinism is that, when explicitly queried, people may not be aware that such a preference for name letters is influencing many of their life decisions (Pelham et al., 2002).

While prior research provides evidence for nominative determinism in some data sets (Anseel & Duyck, 2008, 2009; Pelham et al., 2002), researchers have also argued against it. For example, some have questioned the very existence of the effect, given issues such as large variation in point estimates, lack of publicly available data, small sample size, methodological issues, alternative explanations, and inadequate evidence of a process explanation (Dyjas et al., 2012; Gallucci, 2003; Simonsohn, 2011a). In addition, the reliability of the nominative determinism effect has been called into question for reasons such as reliance on limited names or specific data sets. Researchers have also questioned whether the effect would emerge for a larger set of names; for example, some work has evaluated only two professions, such as dentists and lawyers, and 16 names starting with the letter "D" or "L" (Pelham et al., 2002) or examined whether a small set of names such as "George" or "Geoffrey" have appeared disproportionately in scientific publications in geosciences (Gallucci, 2003; Pelham et al., 2002). Furthermore, research has argued that the effect should be tested across multiple, rather than specific, data sets to establish generalizability (Dyjas et al., 2012). Bayesian hierarchical modeling to address some of the methodological issues did not reach a conclusive finding (Dyjas et al., 2012). Thus, evidence of the nominative determinism effect on major life decisions is mixed.

A main objective of this research, therefore, is to algorithmically examine whether first name letters indeed affect important life

Promothesh Chatterjee https://orcid.org/0000-0002-0992-4336

Promothesh Chatterjee played a lead role in data curation, formal analysis, and writing–original draft and an equal role in conceptualization, methodology, and software. Himanshu Mishra played a supporting role in formal analysis and an equal role in conceptualization and writing–review and editing. Arul Mishra played an equal role in conceptualization, methodology, and writing–review and editing.

Correspondence concerning this article should be addressed to Promothesh Chatterjee, David Eccles School of Business, University of Utah, SFEBB 7103, 1655 East Campus Center Drive, Salt Lake City, UT 84112, United States. Email: promothesh@eccles.utah.edu

decisions. We use language models trained on large text corpora that capture millions of occurrences of people's names and their profession names (and cities in which they live), not just cross-sectionally but over several decades as well. Such an investigation captures temporal associations of names with major life choices and is less likely to suffer from missing names. A second objective of this research is to examine the psychological processes underlying the nominative determinism effect of the choice of profession and cities in which to live, important life decisions that should not necessarily be influenced by factors such as the name of the profession or city. A key theoretical driver proposed for the effect is implicit egotism, or the unconscious tendency to prefer things associated with the self (Jones et al., 2004; Pelham et al., 2002). However, some research has questioned the impact of implicit egotism (McCullough & McWilliams, 2011; Simonsohn, 2011a, 2011b). For example, Simonsohn (2011a) accepts the nominative determinism effect but suggests alternative theoretical explanations such as geographic, ethnic, and socioeconomic confounds and reverse causality rather than implicit egotism. The geographic confound argument suggests that babies born in Georgia, for example, are named Georgia, not that babies named Georgia chose to live in Georgia. The ethnic confound argument suggests that areas with more minorities have more streets named after minority last names. Regarding reverse causality, towns are named after their founders, and streets are named after their residents. Thus, according to the reverse causality argument, Walt Disney working for Disney is an example of people working in family endeavors rather than implicit egotism. If implicit egotism were driving the nominative determinism effect, it would mean that people display such behavior because their name is signaling their identity and is reinforcing (albeit implicitly) their choice of career/city to be in line with their identity.

An implicit egotism account suggests that because people's names capture their identity, they make decisions in line with their identity when possible and when their name is the strongest indicator of their identity. However, if they have competing identities emanating, for instance, from the education received (e.g., in math, science, arts, architecture), their reliance on just their name is likely to be less, suggesting a reduced nominative determinism effect. One way to test for the role of competing identities is through the variable of education. Educational achievements can be an additional way people express or experience their identity. Therefore, we adopt the novel approach of using an identity-expressing variable to examine whether identity expression is a driver of the nominative determinism effect. Such a test also helps us understand the boundary conditions and drivers of nominative determinism using the exogenous variable of education.

Our third aim is methodological. Some work has criticized the nominative determinism effect, citing the lack of breadth and depth of data, including the use of proprietary data (Anseel & Duyck, 2008). As Pelham et al. (2002, p. 479) note, "the scarcity of public databases that include information about people's names and careers, makes archival studies of implicit egotism and career choice inherently difficult to conduct." To address this concern, we use language models trained on multiple large text corpora that are publicly available, including one spanning across a century; this allows us to examine nominative determinism both cross-sectionally and temporally. In today's world, with people producing vast amounts of digital text, these text data sets have become the holding body of people's cognitions, feelings, and actions across time and regions. Analysis of digitized text has provided insights into social, political, business, and even legal issues (Garg et al., 2018; Miller, 2013; Schrage, 2014; Stewart & Zhukov, 2009). Therefore, algorithms commonly use these text data sets to learn facts about humans as well as their preferences (Garg et al., 2018; Hamilton et al., 2016). Public availability encourages reproducibility, and the size of these language models trained on large text corpora in gigabytes helps address the issue of sample size. The language models trained on large corpora also reflect the voluntarily produced knowledge and opinions of millions of individuals. These language models allow examination of nominative determinism from thoughts produced by people voluntarily on a diverse set of topics, even when people are not being directly queried about them. Therefore, they provide a conservative testing ground for whether nominative determinism truly exists and can emerge even in the presence of noise from other associations and influences.

Finally, to explore nominative determinism in major life decisions, we use word-embedding algorithms, a recent natural language processing (NLP) algorithm to create language models based on large text corpora. Unlike word counts that just count the presence or absence of a word, word-embedding algorithms convert unstructured text data to a numeric, multidimensional vector format, such that the dimensions of the vector preserve the meaning of each word uniquely (Mikolov et al., 2013; Pennington et al., 2014). Thus, word embedding captures the meaning of the word in a vector format in a manner similar to how humans infer meaning from reading a word. We first discuss the role of identity expression in major life decisions and then explain how word-embedding algorithms help us test for nominative determinism in large text data sets.

Identity Expression and Major Life Decisions

While names are simply arbitrary social labels used to differentiate one individual from another, in most cultures, namelessness signifies a lack of honor or identity (Frommer, 1982; Watson, 1986). One major reason many people have children is to perpetuate their own names (Arnold & Kuo, 1984; Callan & Kee, 1981; Ramu & Tavuchis, 1986), and early anthropologists knew of no existing culture that did not give people first names (Murdock, 1945). Names have tremendous psychological significance as carriers of identity (Dion, 1983). Consequently, it is not surprising to find that people prefer letters of their own names to others, a notion referred to as the name–letter effect (Hoorens, 2014; Koole & Pelham, 2003; Nuttin, 1985). As the self is a fundamental point of reference for cognition, emotion, motivation, and interpersonal behavior, one's name is strongly associated with one's identity—for example, through self-concept and self-esteem (Allport, 1937; Festinger, 1957; James, 1890). The name–letter effect suggests that the implicit favorable evaluation people have of themselves transfers to the letters in their name (Hoorens & Nuttin, 1993).

How do first names influence profession and city choices? Choices are a conduit through which people express themselves and validate their identity (Aaker & Schmitt, 2001; Kim & Drolet, 2003; Snibbe & Markus, 2005; Tafarodi et al., 2002). Identities are different aspects of the self that vary across time and context, helping people express who they are (Akerlof & Kranton, 2000; Oyserman, 2009). Extant research suggests that the degree to which choices are valued depends on whether they reaffirm people's principles and identities (Kleine et al., 1993; Levy, 1959; Solomon, 1983). For example, a person named

Ellen who is extremely concerned about environmental issues may choose to be an environmental advocate. Consistent with this argument, Wille et al. (2012) find that an employee's identity, shapes and is shaped by, vocational experiences, suggesting that choice of work can be an important source of identity expression. As the self and identity converge in social contexts, people are likely to define themselves in terms of what is relevant and possible in their context. Thus, we argue that the nominative determinism phenomenon will manifest to a greater degree for profession choices than city-to-live-in choices because it is far easier to change professions than to change cities (Karahan & Li, 2016; Molloy et al., 2011). Therefore, given the importance of work roles and where to live in modern society, choices of work and city are important sources of self-identity expression, though their manifestation may differ (Ashforth & Schinoff, 2016; Obodaru, 2016; Oyserman & Yoon, 2009). This is consistent with Stryker's notion of hierarchy of identities (Stryker, 1980), where salience of situational aspect decides which identity will take precedence in accounting for some self-relevant outcome. In other words, we suggest that identity emerging from belonging to a profession is higher in the hierarchy than identity emerging from belonging to a city. Next, we examine how names as carriers of identity and choice of profession/city as a source of self-identity expression are associated.

A core aspect of human identity is to evaluate oneself favorably and to maintain these favorable feelings about oneself. A wide variety of psychological phenomena (e.g., attitudes, social cognitions, behaviors) find their roots in this basic self-enhancement motive (Sedikides & Gregg, 2008; Sedikides & Strube, 1997). Extant research argues that much of this self-enhancing tendency occurs unconsciously (Greenwald & Banaji, 1995; Hetts et al., 1999; Pyszczynski et al., 1999). Furthermore, anything that is connected with one's identity is evaluated positively. Given that names are carriers of identity, it is no wonder that people evaluate their names positively (Hoorens & Nuttin, 1993). Similarly, and consistent with the implicit egotism explanation (Pelham et al., 2002), because choices of profession and city to live in are important sources of expressing identity, professions/cities congruent with profession seekers'/city dwellers' names (e.g., Dennis and dentist, Dennis and Denver) implicitly provide a source for identity expression and choice. Another implication of using choices related to one's own name (congruent profession choices) as a means of implicit identity expression is that alternative ways of expressing identity (e.g., choosing to acquire a college educational degree) can dampen the choices highlighting implicit preference for one's own name. In a mediation study, we discuss how alternative ways of expressing identity such as through education mediate and dampen the relationship between first names and profession choice. We next describe how word-embedding algorithms help us capture the meaning of a word quantitatively.

Word-Embedding Language Models

Words and the context of their occurrence in human language indicate human thoughts and processing (Liu & Karahanna, 2017; Stubbs, 1996). Analyzing text is important because people produce much of their thoughts in the form of language, written or spoken, rather than in the form of numbers. However, traditional statistical models did not have the ability to analyze text, which is an unstructured form of data. Until recently, analysis of text relied on either subjective assessment by linguistic experts (which was time intensive and cost prohibitive) or counting words associated with a certain topic. Recent developments in NLP word-embedding algorithms allow the mining of large text corpora for human beliefs, attitudes, and sentiments using language models. These sophisticated language models trained on large text corpora move beyond count-based methods and are capable of detecting meanings, extracting information, and uncovering relationships between variables (Domingos, 2015). The ability to capture the meaning of words quantitatively is the reason word-embedding algorithms are installed in voice-assisted chatbots such as Siri or Alexa that help complete sentences meaningfully and provide guidance. Different fields of business, computational social sciences, and natural sciences are also using NLP algorithms to examine both substantive and conceptual issues (Bailey et al., 2019; Berger et al., 2020; Bhatia, 2017, 2019; Boghrati & Berger, 2022; DeFranza et al., 2020; Jaidka et al., 2020; Kozlowski et al., 2019; Mooijman et al., 2018; Padarian & Fuentes, 2019; Tataru & David, 2020; Tshitoyan et al., 2019). The influential article by Tshitoyan et al. (2019) looks at how complex material sciences concepts such as structure-property relationships and the underlying structure of the periodic table are captured by word embeddings that make it possible for scientists to distinguish patterns in element properties, including electronegativity, ionization energy, and atomic radius. Word embeddings have been used to decode microbiome-level properties to quantify taxon co-occurrence patterns in gut bacteria to understand inflammatory bowel disease (Tataru & David, 2020). In the geosciences, Padarian and Fuentes (2019) used embeddings for taxonomic analysis of soil profiles that would not be possible otherwise. In our research, we use two specific NLP methods that create word embeddings and thus enable a study of semantic associations among words.

One can get an intuitive idea of word-embedding algorithms following Firth's (1957, p. 11) hypothesis, "You shall know a word by the company it keeps." The key idea here is that the meaning of a word is not innate in that word but is derived from the context in which it appears. By extension, words that frequently appear in the same context share semantic association. Consider the following sentences: "After intense training that lasted the whole day, my legs are wappered out" and "If the horse is not rested properly, he is going to be wappered out." While the meaning of the word "wappered" from medieval English might be unknown to many, the context in which the word appears shows that it relates to exhaustion. Word-embedding algorithms use a similar logic to infer the meaning of words from the context in which they appear in language. They rely on the distribution hypothesis, a principle put forth by linguists that helps understand how words acquire meaning. Linguists (de Saussure, 2011) suggest that "language is a system of interdependent terms in which the value of each term results solely from the simultaneous presence of others." That is, the context in which a word appears informs receivers of its meaning. For example, the meaning of "spring" as a season or as coiled metal can only be ascertained by the context in which it appears in language. The distributional hypothesis proposes a link between how words are distributed in language and the similarity in their meaning (Jurafsky & Martin, 2020). Important to note is that such a representation is not based on word or co-occurrence frequency but rather an estimation of the probability of co-occurrence. This results in a much more expressive representation than, for example, methods dependent on a term-frequency matrix (Baroni & Lenci, 2010; Jaidka et al., 2020;

Turney & Pantel, 2010). For example, a comparison of word embeddings is capable of estimating associations of words that occur infrequently in the training text and words that are logically associated but co-occur infrequently; it is also capable of collapsing the meaning of various forms of a word (e.g., plural, progressive) or similar words (e.g., synonyms, related concepts) into a single representation, as necessary (Garten et al., 2018).

With large quantities of text, the goal is to use algorithms to learn semantic relationships among words, just like a human can learn about the word "wappered." Therefore, word-embedding algorithms use a context window to determine what other words occur in the context of a specific word across the numerous occurrences of that word in the data set and how those words are semantically related. This process results in each word being mapped onto a semantically relevant, high-dimensional space; that is, word embeddings convert a word into a high-dimensional vector. A word would be converted into a vector uniquely identified by, for example, 200 or 300 numerical values. While each of the 200 dimensions of the large language model would not have any meaning on their own, the numerical values on these 200 dimensions capture the relationship of various words in the corpus, which can be semantic (e.g., rose–pleasant), syntactic (car–cars), or rhythmic (doom–gloom; Rezaei, 2022).

Once a word has numerical representation in the form of a vector, numerous mathematical operations can be performed. For instance, it is easy to find similarity between words in this high-dimensional space using cosine distance (instead of Euclidean distance, which is used in a two- or three-dimensional space). Thus, computing the cosine of the angle between the word vectors can be used to measure similarity between words (Caliskan et al., 2017; Garg et al., 2018).

$$\text{Cosine Similarity} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}. \tag{1}$$

For similar words, cosine distance will be close to 1 (i.e., cos (0)). As the dissimilarity between words increases, cosine distance will decrease and be close to 0 (i.e., cos (90)). If words carry opposite meanings, cosine similarity will be nearly −1 (i.e., cos (180)).

Using Large Language Models to Test for the Nominative Determinism Effect

As discussed previously, word-embedding algorithms in the large language models use a context window to infer the meaning of a word from the context in which it appears in language. Not only is the co-occurrence of words in a context window considered across thousands of occurrences of a word in the corpus, but the words in the window also allow the algorithm to infer meaning. Thus, the vector created captures the meaning of the word quantitatively, which in turn allows meaning-based analysis. Applied to the nominative determinism effect, the word-embedding algorithm captures the semantic association of one's name with a profession or city name (e.g., Dennis is a successful dentist), allowing the algorithm to infer from the language context whether or not Dennis and dentist are associated. Similarly, if the name Dennis does not co-occur much with the profession of lawyer, the algorithm will infer that Dennis is not associated much with the profession of lawyer, and the cosine similarity of the word vectors of Dennis and lawyer will reflect this lesser association. When the large language models reflect such a pattern of association, the analysis will reveal that the angle between the word vectors of Dennis and dentist is smaller (reflecting more similarity) than the word vectors of Dennis and lawyer (reflecting less similarity). In summary, the word vectors allow us to determine the level of association of people's names with professions or cities.

We utilized large language models created with two word-embedding algorithms: word2vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014). In the Appendix A, we describe their mathematical formulation and how they help us identify semantic associations. Next, we validate the ability of the algorithms to make accurate and relevant predictions in our context and then report two studies examining the nominative determinism effect in the domain of profession and choice of city in which to live, using different data sets and word-embedding algorithms.

Empirical Overview

We first conducted validation tests across three text data sets to test whether word-embedding algorithms used in large language models can be used to test for association between name and city/profession. For validation, we used several famous people who were uniquely associated with a profession or a city (e.g., Phelps–swimmer, Buffett–Omaha). Moreover, these large language models capture millions of occurrences of people's names, cities, and careers and are less likely to suffer from sample size issues. After having established that word embeddings can be used to detect the association between name and city/profession, we then tested for the nominative determinism effect across three large language models for both profession name and city name after controlling for potential confounding variables. Finally, in a decade-by-decade analysis, we further examined whether implicit egotism might be at work in driving the nominative determinism effect by considering whether the presence of alternative identity-expressing variables through which people signal their identity (e.g., education) can reduce their reliance on their name and thereby dampen the nominative determinism effect.

Data Sets

We used multiple word-embedding language models based on large text corpora to examine whether nominative determinism exists in the real world. First, Common Crawl is a web-crawl repository, several 100 TB in size, that contains a compilation of billions of different web pages scraped from the public-facing internet since 2013 (Caliskan et al., 2017; DeFranza et al., 2020; Garg et al., 2018; Pennington et al., 2014). Second, the Twitter data set (Pennington et al., 2014) captures people's opinions and interactions that have had an immense impact on social, political, business, and even health outcomes. Common Crawl and Twitter both provide pretrained embeddings using the GloVe algorithm. Third, the Google News text corpus (Mikolov et al., 2013) captures contemporary news content. Fourth, the Google Books text corpus (Hamilton et al., 2016) covers millions of books in diverse 20th-century genres that have been converted into a text corpus. Google data sets use embeddings created with the word2vec algorithm.

Validation of Pretrained Embeddings' Ability to Capture Association Between Names and Professions/Cities

In subsequent studies, we use pretrained word embeddings from three text corpora (Common Crawl, Twitter, and Google News), developed using the algorithms word2vec and GloVe to capture the nominative determinism effect. While these have been validated and used extensively in different research articles and industry applications, the objective of this validation was to ensure their ability to capture associations between constructs—specifically, between names and professions/cities. To test whether these pretrained embeddings are useful in detecting the association between names and professions/cities, we created lists of famous people and the professions/cities associated with them. To develop the list, we consulted various websites where famous people and their associated cities/professions were available. We tried to retain names that were universally famous as opposed to being popular in a particular country. We had to avoid some names and associated places such as Shakespeare and Stratford-on-Avon because the associated place did not have a single name, and consequently we would have difficulty in finding word embedding for it. Furthermore, we made sure not to include people who had places named after them such as Bolivia and Simon Bolivar, Colombia and Columbus. We refer to the list as the validation list (as depicted in Table 1). The cosine-similarity measure between famous people's names and professions/cities' names should capture the association between name and associated name/place. For example, Elvis and Memphis should show a greater cosine similarity than Elvis and Moscow.

Procedure

We first obtained the pretrained embedding for each person's name, profession name, and city name in the validation list. For each of the three text data sets, we then computed the cosine similarity between person name and profession name and also the cosine similarity between person name and city name. For example, we computed the cosine similarity between word vectors for Elvis and Memphis, which formed a compatible pair, as we would expect Elvis to be most closely associated with the city of Memphis. We also computed the cosine similarity of Elvis with each of the other 14 city names in our validation list; this formed the incompatible pairs. We averaged the cosine similarities for each of the incompatible pairs to obtain one value (e.g., Elvis and Macedonia, Elvis and Moscow). We then tested whether the cosine similarity for Elvis and Memphis was higher than the cosine similarity for Elvis and incompatible city names.

Table 1
Validation List of Famous People

Names	Cities	Names	Profession
Alexander	Macedonia	Armstrong	Astronaut
Bachchan	Bombay	Diana	Princess
Beckham	Manchester	Forbes	Publisher
Buffett	Omaha	Hughes	Aviator
Chanel	Paris	Lewis	Sprinter
Churchill	Kent	Monet	Painter
Elvis	Memphis	Morgan	Pirate
Gandhi	Delhi	Newton	Scientist
Jung	Basel	Nightingale	Nurse
Lennon	Liverpool	Pele	Footballer
Newton	Cambridge	Phelps	Swimmer
Obama	Chicago	Shakespeare	Author
Putin	Moscow	Shakira	Singer
Schwarzenegger	Hollywood	Teller	Magician

Because of the relatively small sample size (15 pairs of names and professions/cities), we ran a permutation test, as it does not assume a sampling distribution but generates the distribution by resampling the observed data.

Results of Validation Test

We first conducted the same analysis for person name and profession name for each of the three text data sets (see Table 2, for the aggregate results). The cosine similarity and permutation tests are depicted for Common Crawl (Table B1 and Figure B3 in Appendix B), Twitter (Table B2 and Figure B4 in Appendix B), and Google News (Table B3 and Figure B5 in Appendix B). The results indicate that the cosine similarity of person names and compatible profession names is significantly greater than that of person names and incompatible profession names. For example, in the Twitter embeddings, the cosine similarity of Monet to painter (compatible profession name) is 0.395, while the cosine similarity of Monet with all other (incompatible) profession names is 0.151 (see Table B2 in Appendix B). Therefore, embeddings do have the ability to capture association between names and professions.

We ran a similar validation test for person and city names. Again, the permutation test indicated that the cosine similarity of person name and compatible city name was significantly greater than that of person name and incompatible city names (see Table 3, for the aggregate results). The cosine similarity and permutation tests are depicted for Common Crawl (Table B4 and Figure B6 in Appendix B), Twitter (Table B5 and Figure B7 in Appendix B), and Google News (Table B6 and Figure B8 in Appendix B). For example, the Twitter data set Elvis–Memphis has a greater cosine similarity (0.326) than Elvis with incompatible city names (0.172; see Table B5 in Appendix B). Therefore, embeddings have the ability to capture association between names[1] and cities.

Study 1: Do First Name Letters Predict Profession and City Choices in Real-Life Data Sets?

The main objective of Study 1 was to examine whether the first letter of a person's name indeed affected profession and city choices in real-life data sets after controlling for potential confounding variables. The reason we focused on the first letter of a person's name is because literature suggests that the name–letter effect indicates a greater preference for the initials of one's first or last name. While some research has indeed found a name–letter effect for noninitials (Hoorens, 2014; Hoorens & Todorova, 1988; Koole et al., 2001; Nuttin, 1987). Others have failed to find the name–letter effect for noninitials (Koole et al., 1999, Study 3). The general finding is that name–letter preference is stronger for initials than noninitials (Hoorens, 2014). Because most of the literature has focused on initial letters, some scholars have labeled the name–letter rating task as "Name Initials Letter Task" (Sava et al., 2011) or the "Initial(s) Preference Task" (Sariyska et al., 2014; Schröder-Abé et al., 2007; Steinberg et al., 2007; Stieger & Burger, 2010; Vater et al., 2010; Zeigler-Hill, 2006). Furthermore, the literature does not distinguish whether such a name–letter preference depends on whether there is a diminishing preference as one moves from the

[1] We performed additional validation based on t-SNE visualization of embeddings; please see Appendix B.

Table 2
Validation Test Results Across the Three Large Language Models for Person Names and Profession Names

	Cosine similarity between person names and profession names		
Data sets	Mean compatible cosine similarity	Mean incompatible cosine similarity	Permutation *t* test
Common Crawl	0.26	0.10	$p < .01$
Twitter	0.30	0.15	$p < .01$
Google News	0.15	0.08	$p < .05$

initial to the later letters. Theory also does not investigate whether the number of letters in one's name affects the name–letter effect. Given this lack of theoretical finding in the literature, we focus on the first letter of a person's name. Moreover, in this research, we are focusing on nominative determinism, which is a downstream manifestation of the name–letter effect in the form of a preference for city or profession starting with the same first letter as one's first name. Therefore, we confine our predictions and findings to nominative determinism and not the name–letter effect, which is a more general preference. The literature on nominative determinism explains that the preference for one name and preference for a city or profession are mainly driven by the similarity of the first letter of one's first or last name and the first letter of the city or profession name (Jones et al., 2002; Pelham et al., 2002). That is, it is centered around the first letter.

We used three large language models based on the text corpora—Common Crawl, Twitter, and Google News—to test for nominative determinism.[2]

Procedure for Creating Data Set for Analysis

Step 1

We obtained pretrained embeddings for Common Crawl and Twitter (both developed using the GloVe algorithm) and Google News (developed using the word2vec algorithm) to examine the association between person names and profession names or city names. We followed the same procedure for each pretrained embedding data set.

Step 2

From the U.S. Social Security Administration website's publicly available data set, we chose all first names for the years of birth in the last decade (2011–2020). We did this to prevent undue influence of name trend in any particular year. Each record in the individual annual files has the format "name, sex, number," where name is 2–15 characters, sex is M (male) or F (female), and "number" is the number of occurrences of the name.

Step 3

To further prevent name popularity from influencing our results, we retained only the names that were included in all years of the decade (17,506 names). For example, if in certain years a greater share of baby boys were named Dennis and there was an increase in the number of dentist occupations, the share of people named Dennis becoming dentists would be more than expected when the underlying assumption is that name popularity is stable over time. Relatedly, base rates can influence the nominative determinism results. For example, when asking if too many people named Dennis, compared with, say, people named Chad, are dentists (vs. engineers), we need to consider both how many people are named Dennis and Chad and how many people are dentists and engineers. It could be that the name Dennis is more common than the name Chad, and there are more dentists than engineers in the population. Therefore, given base rates, there could be more people named Dennis who are dentists than engineers. Given these issues, we used the frequencies of the name data to control for such issues statistically. We also control for the frequency with which professions and cities occur in the embeddings (as described in Step 6).

Step 4

Ethnicity is a plausible explanation for nominative determinism in some domains, such as marriage decisions and location in which to settle (Simonsohn, 2011a). For example, the nominative determinism effect is more likely to emerge in ethnic groups because members of these groups often marry within their own ethnicity. Moreover, ethnic groups have different distributions of names than those in the general population. In addition, discrimination in professions due to ethnicity occurs frequently (Derous & Ryan, 2019); thus, we control for it. While we could obtain gender and frequency information (to control for base rate) from the social security data sets to control for various ethnic influences, we need the ethnicity information for the first names. Tzioumis (2018) provides information on the respective count and proportions of 4,250 first names across six mutually exclusive ethnic groups. We combined these names with the names retained in Step 3 to assign ethnicity information to first names (we were left with a final count of 3,410 names). To simplify our analysis, we then divided the ethnicity information into two categories, White and non-White (the pattern of results does not change when we include the different ethnicities).

Step 5

We obtained the word embedding for each retained name from the pretrained embeddings.

Step 6

This step describes how we created profession and city embeddings. We obtained the profession list from several career-related

Table 3
Validation Test Results Across the Three Large Language Models for Person and City Names

	Cosine similarity between person and city names		
Data sets	Mean compatible cosine similarity	Mean incompatible cosine similarity	Permutation *t* test
Common Crawl	0.40	0.26	$p < .01$
Twitter	0.33	0.15	$p < .01$
Google News	0.45	0.35	$p < .05$

[2] To access all data and codes, visit https://osf.io/gb5q9/?view_only=d7ad52586c164fc590ce97225fe0a11b.

websites, selecting only single-word professions, such as animator (vs. multiple-word professions such as aircraft instrument technician). Our list contained 508 professions. For testing the nominative determinism effect in the city-to-live-in choice, we obtained a list of cities from the SimpleMaps website (https://simplemaps.com/data/us-cities). Our list contained 14,856 cities. We again obtained the unique word embedding of each profession and city name from the pretrained embedding for each text corpus. That is, when running the analysis using the Common Crawl corpus, we obtained the embedding of person name, profession name, and city name from the Common Crawl corpus. This helped us test for nominative determinism using the millions of co-occurrences of these names in the language people use. While the frequency of names was directly available from the social security data files (Step 2), for frequency of occurrence of professions/cities, we queried the embeddings files for a count. In other words, the frequency of professions/cities is simply a count of the number of times the terms appeared in the pretrained embeddings. Finally, because the unit of analysis is names, we averaged the profession/city name embedding by letter so that we could compare letters with names. For example, all the profession/city names starting with the letter "A" were grouped together for comparison with person names (from Step 5).

Step 7

After obtaining the word embeddings of person names and profession/city names, we used cosine similarity to determine the level of association of person name with profession/city names for each corpus. This allowed us to compute cosine similarity between compatible person name and profession/city name embeddings (e.g., between names starting with "A" such as Adam and profession names starting with "A" such as artist or city names starting with "A" such as Aspen). We also computed cosine similarity between incompatible person names and profession/city names (e.g., Adam and "B," "C," … , "Z" profession/city names). For example, by comparing the cosine similarity of the word embeddings between Adam and artist and Adam and lawyer, we can determine whether Adam is more likely to be an artist or a lawyer. Higher values of cosine similarity indicate higher similarity.

Step 8

A potential alternative explanation for the nominative determinism effect could be the presence of alliteration (Lea et al., 2008) or natural clumping of words starting with the same letters.[3] For instance, there might be a natural tendency of words starting with the letter "a" to cluster in the same semantic space. If that is the case, the phenomenon of association of names with professions/cities would be explained by alliteration rather than by nominative determinism. In the next section, in which we provide details of variables used in the data set, we explain how we test for this alternative explanation.

Step 9

We created the final data set by combining the relevant variables (details of variables given subsequently) for analysis.

Details of Variables Used in Final Data Set for Analysis

Names

This includes names of individuals retained from Step 5 along with their respective embeddings.

Condition

We created this categorical variable as the main predictor with two levels. First, the compatible pair occurs when the first letter of the person name is the same as the first letter of the profession/city names. Second, the incompatible pair occurs when the first letter of the person name is not the same as the first letter of the profession/city names.

Compatible Names and Professions/Cities

To capture the association between compatible pairs of person names and profession/city names, we measured the cosine similarity between the word embeddings of the names of compatible pairs of names and professions/cities. That is, we computed cosine similarities between the word vectors of A names and A profession/city names, B names and B profession/city names, and so forth.

Incompatible Names and Professions/Cities

This variable captures the cosine similarity between incompatible pairs of word embedding of person names and word embedding of profession/city names. That is, we computed cosine similarities between the word embedding of A person names and B profession/city names, A person names and C profession/city names, and so forth (e.g., Adam and bartender, caretaker, doctor; Adam and Boulder, Cleveland, Denver). We then averaged these to create a single measure of incompatible cosine similarities.

Gender

This variable captures the gender of the retained names, which helps us analyze and control for any gender influences in the nominative determinism effect.

Ethnicity

We added ethnicity information to person names using the procedure described in Step 4, based on Tzioumis's (2018) data set.

Frequency of Names

As described in Step 3, this variable is the total frequency of names as they appear in the U.S. social security data. This allows us to control for base rate representation of names.

Average Profession/City Frequency

We describe the basic procedure for generating this variable in Step 6. We counted the frequency of profession/city terms in word-embedding data sets. As the unit of analysis is at the level of names, we aggregated the profession/city frequency by letter. We averaged

[3] We thank Reviewer 1 for this suggestion.

all the professions/cities beginning with the letter "A" and matched these to the "A" names. Thus, all "A" names in a particular decade have the same values. The variables *frequency of names* and *average profession/city frequency* help us control for the base rate. Furthermore, a concern of critics of nominative determinism is the possibility of a spurious correlation between ratios of professions (lawyers to dentists; Pelham et al., 2002, Study 7) to ratios of names (La_ to Den_ names). According to Simonsohn (2011a), an increase over the years of both the relative proportion of lawyers and the relative proportion of La_ names would indicate a fallacious association between the variables. To address concerns with fallacious association, we use the frequency information of both names and professions/cities as covariates in the data sets.

Alliteration Score

To control for the alternative explanation of alliteration, we first randomly sampled a large set of words—around half a million—and from this, we preprocessed the words to remove URLs, emojis, numbers, punctuation, concatenated words, and foreign language words. We filtered out words that are not in the English language using a dictionary available in R package Hunspell (Ooms, 2022). We further annotated the words to obtain the parts of speech for each word and retained only the nouns. Nouns were retained because both people's names and profession/city names are nouns. Subsequently, we were left with 7,500–10,000 nouns across different data sets. We then computed the cosine similarity of nouns with one another. Intuitively, we used a data structure that in the first and second columns has the nouns and in the third column has the cosine similarity of the nouns in Columns 1 and 2. Then, for each row, we compared the first letter of the noun in Column 1 with the first letter of the noun in Column 2. If they were the same, they were indicated in Column 4 as "compatible" and "incompatible," if the first letters were not the same. The cosine similarity was calculated for each letter of the compatible and incompatible condition and served as the alliteration score. This alliteration score was used as a covariate in the mixed model, along with other covariates.

Cosine Similarities of Compatible and Incompatible Name Profession/City Pairs

We use the cosine similarities as our dependent variable to measure nominative determinism. Recall that the variable *compatible names and professions/cities* refers to the cosine similarity between compatible name and profession/city word vectors (e.g., Adam and accountant, Adam and Athens), and *incompatible names and profession/cities* refers to the cosine similarity between incompatible name and profession/city word vectors (i.e., Adam and bartender, caretaker, doctor, and so forth; Adam and Boulder, Cleveland, Denver, and so forth). Nominative determinism implies that the difference between the means of cosine similarities of compatible and incompatible name profession/city pairs is statistically significant. In other words, if nominative determinism exists, the cosine similarity between the compatible pair of word vectors of names and word vectors of profession/cities (e.g., Adam and Athens, Adam and accountant) will be statistically higher than the cosine similarity between the incompatible pair of word vectors of names and word vectors of professions (e.g., Adam and Boulder, Adam and bartender).

Analysis and Results

This analysis aimed to test if nominative determinism exists in profession and city choice and to control for alternative explanations underlying the nominative determinism effect. We utilized three real-life language models based on large text corpora (Common Crawl, Twitter, and Google News) to examine the nominative determinism effect. In the results, we first report the nominative determinism effect in profession choice and then in city choice.

Model-Building Procedure Across Data Sets

Considering the nested structure of the data, in which the cosine similarities of compatible and incompatible name and city/profession pairs (Level 1) were nested within each name (Level 2), the analysis used a mixed-effects model. For all the models, we kept the cosine similarities as the outcome variable. We first compared the linear regression model (Model 1) with the random intercept model (Model 2). The goal was to determine whether the nesting structure inherent in the data warrants the use of a mixed-effects model over a simpler linear regression model. In general, model fit methods for comparisons indicated that the model with random intercept is considerably better across data sets. We next compared the random intercept model (Model 2) with the model that nested names under the letters (Model 3). In Model 4, we added the covariates we wanted to control for, such as average profession/city frequency, gender, race, alliteration scores, and total frequency of names (see Tables B7–B12 in Appendix B, for details).

The Nominative Determinism Effect in Profession Choice

In the full model (Model 4), in all three data sets (Common Crawl, Twitter, and Google News), we find a main effect of condition such that cosine similarity of person names with compatible profession names is more than the cosine similarity of person names and incompatible profession names (see Tables B7–B9 in Appendix B). There is a significant difference between the estimated marginal means of difference scores of compatible names and professions versus incompatible names and professions (for Common Crawl, $M_{\text{compatible}} = 0.139$ vs. $M_{\text{incompatible}} = 0.130$, $p = .0002$; for Twitter, $M_{\text{compatible}} = 0.154$ vs. $M_{\text{incompatible}} = 0.142$, $p < .0001$; for Google News, $M_{\text{compatible}} = 0.141$ vs. $M_{\text{incompatible}} = 0.128$, $p < .0001$).

Thus, across all three data sets, we find evidence for the nominative determinism effect. Our results are consistent with the identity expression argument that professions congruent with profession seekers' names (e.g., Dennis and dentist) implicitly provide a source for one's identity expression and choice.

Evidence of the nominative determinism effect for professions is an important finding because of the various questions about the existence of the effect in the real world stemming from concerns such as large variation in point estimates, small sample sizes, methodological issues, and alternative explanations (Dyjas et al., 2012; Gallucci, 2003; Simonsohn, 2011a). After controlling for diverse variables that might potentially influence the results, we show that a significant nominative determinism effect exists in profession choice, using multiple large language models based on large corpora. We next examine whether nominative determinism exists for city choice.

Nominative Determinism Effect in City Choice

Arguably, this is a more conservative test of the phenomenon, as people change professions far more easily than the cities in which they live. According to a Bureau of Labor Statistics report (Toossi, 2002), people born in the years 1957–1964 engaged in an average of 12.4 professions from the age of 18 to 54. In addition, a Health and Retirement Study (2015), a longitudinal project at the University of Michigan's Institute for Social Research, shows that the median distance children live from their mothers in the United States is 18 miles, and only 20% live more than a couple hours' drive from their parents. Karahan and Li (2016) report that less than 3% of working-age adults relocate to a different state each year. Thus, a conservative test for the nominative determinism effect would be people's choice of the city in which to live.

As in the case of profession choice, we used the three data sets of Common Crawl, Twitter, and Google News to examine the nominative determinism effect for city choice. In all three data sets, Model 4 (full model) shows a main effect of condition, such that the cosine similarity of person names and compatible city names is significantly greater than that of person names and incompatible city names (see Tables B10–B12 in Appendix B). There is a significant difference between the estimated marginal means of cosine scores of compatible names and cities versus incompatible names and cities (for Common Crawl, $M_{\text{compatible}} = 0.426$ vs. $M_{\text{incompatible}} = 0.403$, $p < .0001$; for Twitter, $M_{\text{compatible}} = 0.259$ vs. $M_{\text{incompatible}} = 0.246$, $p < .0001$; for Google News, $M_{\text{compatible}} = 0.517$ vs. $M_{\text{incompatible}} = 0.479$, $p < .0001$).

Overall, across multiple real-life text data sets, evidence shows the presence of the nominative determinism effect in professions and cities while controlling for diverse variables that could potentially affect the findings. Although the effect is consistent, these data sets only provide a snapshot because they are cross-sectional in nature. In Study 2, we explore the evolution of the nominative determinism effect in the 20th century using a longitudinal data set.

Study 2: Testing for Nominative Determinism in Profession and City Choice in the 20th Century Using Decade-by-Decade Word Embeddings

The main objective of this study was to examine whether nominative determinism indeed affects profession and city choices using large language models based on text corpus that captures millions of occurrences of people's names and their career names over several decades. The second objective was to examine whether gender disparity in the nominative determinism effect exists in profession and city choices and how it evolved over the 20th century. A third objective was to examine the psychological process underlying the nominative determinism effect by examining the role of an identity-expressing variable—namely, education. We suggest that names capture part of people's identity and that if people have competing identities, such as those emanating from their education, they are likely to rely less on just their names, thereby mitigating the nominative determinism effect.

The procedure was the same as in Study 1. However, instead of one pretrained embedding, we used the Google Books decade-by-decade embeddings (Hamilton et al., 2016), which contain linguistic and cultural trends as reflected in millions of books written over the period between 1900 and 2000, to assess the difference in cosine similarities of compatible versus incompatible pairs of names and professions. We also tested the mediating role of education in the nominative determinism effect. We created a data set for each decade by combining the relevant variables, with the final data set created by combining all of the decade data sets. All the variables were the same as in Study 1, except for two additional variables: decade and enrollment in higher education. We also did not create the condition variable (the categorical variable used in Study 1 with two levels: compatible names and professions/cities and incompatible names and professions/cities); instead, we used a difference score to capture nominative determinism. For nominative determinism to exist, the difference score between the means of cosine similarities of compatible and incompatible name and profession/city pairs should be statistically significant.

The decade variable has values from 1 to 10, where 1 corresponds to the decade 1900–1909, 2 to 1910–1919, and so forth. By including this temporal component in the data set, we address a related age-distribution argument of Simonsohn's (2011a). That is, if Walter, relative to Dennis, is the name of an older person, Walter will be less likely to be employed and less likely to be a dentist than Dennis. This is a valid criticism when using cross-sectional data that capture only snapshots at a particular moment in time, given that our data span 10 decades, any age-related concerns should be reduced.

The enrollment variable refers to the total fall enrollment at higher education institutions at the start of each decade. These data come from the National Center for Education Statistics (Snyder, 1993) and are reported by gender. We collated this exogeneous variable to test our theorizing.

Model-Building Approach

Considering the nested structure of the data, in which the difference score between cosine similarities of compatible and incompatible name and profession/city pairs in each decade (Level 1) was nested within each name (Level 2), the analysis used a longitudinal mixed-effects model. For all the models, we kept the difference score as the outcome variable and entered decade, gender, and their interaction into the model as fixed effects (except the null model); we entered intercepts for names as random effects.

We first compared the linear regression model (Model 1) with the random intercept model (Model 2). The goal was to determine whether the nesting structure inherent in our data set warranted the use of a mixed-effects model over a simpler linear regression model. All the model fit methods for comparisons indicated that the model with random intercept was considerably better.

We next compared the random intercept model (Model 2) with the intercept-only model (Model 3), and the model with a first-order autoregressive error structure (Model 4). Model 4 does not assume independence of random effects; instead, it specifies that the correlations between the repeated measurements in each decade for each name decay as the temporal distance increases. Finally, we also compared Model 4 with Model 5, which included time-varying (e.g., name frequency in each decade, alliteration scores) and time-invariant (e.g., gender, race) covariates. The results for the presence of the nominative determinism effect were consistent across all the models (see Tables B13 and B16 in Appendix B). We focus on Model 5 (full model), which controls for all the covariates, for our discussion.

Nominative Determinism in Profession Choice Over Decades

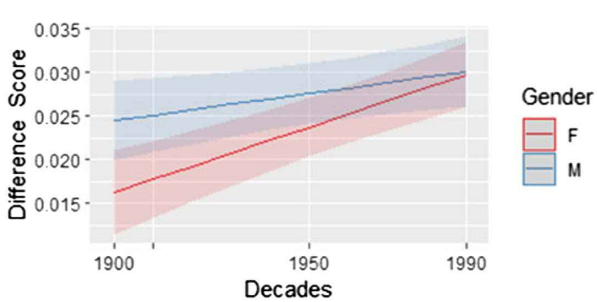
The aims of this analysis were threefold: (a) to examine whether the nominative determinism effect exists in profession choice over decades; (b) to examine gender differences in the nominative determinism effect, if any; and (c) to test for psychological processes underlying the nominative determinism effect by using the identity-expressing variable of education. Model 5 (Tables B13 in Appendix B) shows a significant intercept (.015, $p < .01$). This implies that the similarity of name to profession was greater when these shared the same letter than when they did not (i.e., a pattern consistent with the nominative determinism effect). While the magnitude of the intercept is relatively small, it is important to note that the intercept value does not lend itself to regular interpretation because the difference score is actually a difference-of-difference score (the difference between the similarity of word embeddings of compatible name and profession pairs less the difference between the similarity of word embeddings of incompatible name and profession pairs). Thus, after controlling for diverse variables that might influence the results, we show that a significant nominative determinism effect exists in profession choice using a large text data set that is not susceptible to criticisms of cherry-picking.

Interaction of Decade and Gender in the Nominative Determinism Effect

We find a Decade × Gender interaction ($\beta = -.001$, $p < .05$) in the nominative determinism effect. Decomposing the interaction, we find that men had a greater difference score than women. However, the difference between the genders diminished over time, as indicated by the negative interaction term. The marginal effects plot in Figure 1 depicts this interaction. In essence, the interaction indicates that while there is a significant difference in the nominative determinism effect between men and women during the early 20th century, as we move from earlier to later decades, there is only a marginal difference in how men display the nominative determinism effect, indicating that at a conceptual level, their ability to choose professions has not changed much. However, for women, an increasing nominative determinism effect indicates that their ability to choose a profession that matches their name (and, thus, their self-identity) has improved. Prior research suggests that because women are more likely to change their last name after marriage, their first names are a more significant part of their identity, and thus they show a stronger nominative determinism effect (Pelham et al., 2002). By contrast, our findings suggest that in the early 20th century, women's limited career choices constrained their ability to express their identity in this way; however, as more options emerged for women to express their self-identity over time, the difference in the nominative determinism effect between men and women diminished. Thus, our results are consistent with the argument that professions congruent with profession seekers' names implicitly provide a source for identity expression.

Figure 1

Marginal Effects Plot



Note. M = male; F = female. See the online article for the color version of this figure.

Nominative Determinism in City Choice Over Decades

To test for nominative determinism in city choice over decades, we again use the Google Books corpus, which captures millions of occurrences of people's names and the cities they live in over several decades and, thus, is less likely to suffer from sample size issues. The aims of this analysis were (a) to examine whether the nominative determinism effect exists in the city-to-live-in choice and (b) to examine the pattern of the nominative determinism effect over the decades of the last century. The model-building approach was similar to the choice of profession over decades. We focus on Model 5 (full model), which controls for all the covariates for our discussion.

Model 5 (Tables B16 in Appendix B) shows a significant intercept (.039, $p < .01$), consistent with the nominative determinism effect. A significant intercept implies that the similarity of names to cities was greater when they shared the same letter than when they did not. Given the various criticisms, finding evidence of the existence of the nominative determinism effect in a conservative test lends additional support to the findings.

Absence of Interaction of Decade and Gender in the Nominative Determinism Effect

Unlike in the choice of profession, we did not find any evidence of a Decade × Gender interaction in choice of city. Recall that the Decade × Gender interaction in the choice of profession demonstrated a significant difference in the nominative determinism effect between men and women in the early 20th century; however, there was not much difference in the nominative determinism effect later in the century. We speculated that in the early 20th century, women's career choices were limited due to availability of profession choice; however, the difference in the nominative determinism effect between men and women diminished as profession options increased. In the case of city choice, it is possible that the steady reduction in interstate migration over the decades has influenced both genders equally (Karahan & Li, 2016), and thus we do not find any evidence of interaction of decade and gender.

Psychological Processes Underlying the Nominative Determinism Effect

The presence of Decade × Gender interaction in the choice of profession allows us to test for our proposed psychological process. Theoretically, we suggest that the nominative determinism effect occurs when people have the ability to choose a profession in line with their identity, specifically when their name captures the sum total of who they are. If this is the case, we should find that when competing identities become salient, people become less reliant on their names to convey who they are. Prior research indicates that education provides people with another set of identities (Donald et al., 2019; Tomlinson, 2013). If so, we should find that when education enrollment increases, it acts as a suppressor variable. That is, people's educational identity will suppress the influence of their name on career choice. When using enrollment in higher education institutions as a mediator, we should find that over decades, as enrollment increases, it acts as a suppressor variable and reduces the nominative determinism effect. Furthermore, we should find this effect for both men and women; that is, gender should not moderate the suppressing role of the mediator (enrollment in a higher education institution). We propose no difference across gender because both men and women are likely to rely on an educational identity.

As we have multilevel data, we tested this hypothesis following the procedure outlined in Tingley et al. (2014) and implemented it in an R package mediation. As a first step, we fitted two random intercept models. The first model had enrollment in a higher education institution as the outcome variable, and we entered decade, gender, and their interaction into the model as fixed effects. The second model had the difference score as the outcome variable, and we entered decade, gender, their interaction, average job frequency, alliteration score, ethnicity(White), name frequency, and enrollment in a higher education institution as fixed effects. We used these fitted models in the mediation analysis. To estimate the confidence interval around the treatment effect, direct effect, and average total effect, we performed 1,000 simulations using the quasi-Bayesian Monte Carlo method based on normal approximation (Imai et al., 2010).

We estimated mediation effects for men and women (see Tables B14 and B15 in Appendix B). For men, we found that the average total effect was significant (estimate = 0.000718, $p = .002$). This means that over decades, the difference score did change significantly for men. When we decompose this effect into direct and indirect effects, we find that the portion of the average total effect that is transmitted through enrollment in a higher education institution (i.e., indirect effect) is statistically significant and negative (estimate = −.001496, $p < .0001$), indicating that enrollment in higher education institutions acts as a suppressor. By contrast, the direct effect (excluding the mediator) is significant and positive (estimate = .002214, $p < .0001$).

For women, the results were similar: The average total effect was statistically significant (estimate = .00133, $p < .0001$). This indicates that over decades, the difference score increased for women. When we decompose this effect into direct and indirect effects, the pattern of results was similar to that of men. We find that the portion of the average total effect that is transmitted through enrollment in a higher education institution (i.e., indirect effect) is statistically significant and negative (estimate = −.00168, $p < .0001$). The direct effect (excluding the mediator) is also significant (estimate = .003, $p < .0001$).

These findings reflect an inconsistent mediation (Cliff & Earleywine, 1994; MacKinnon et al., 2000; Tzelgov & Henik, 1991). In mediation analysis, opposite signs in the direct and mediated effects of an independent variable are known as inconsistent mediation models (Davis, 1985). This is in contrast with regular mediation models, in which the direct and mediated effects have the same sign. Consider a hypothetical example of inconsistent mediation in which the suppression effect is present (McFatter, 1979). Suppose we are interested in the relationship among workers' intelligence (X), level of boredom (M), and the number of errors made on an assembly line task (Y). We could argue that the direct effect of intelligence on errors would be negative (workers with higher intelligence will make fewer errors) and the indirect effect of intelligence on errors mediated by boredom would be positive (the higher the intelligence of workers, the more they will be bored, and the greater the boredom, the higher the number of errors). Overall, in our context, the mediation analysis suggests that as enrollment in higher educational institutions increased, it suppressed the nominative determinism effect for both men and women. Our analysis supports the argument that the implicit preference for one's own name is a means for implicit identity expression, and as identities can be expressed in many ways, alternative means of implicit identity expression, such as education, dampen the implicit preference for one's own name.

General Discussion

An important aspect of people's life and general well-being is their ability to make choices. In this research, we provide evidence of how these choices are shaped by implicit biases. In the profession domain, most theoretical models of organizational attraction and profession choice depict the process as being driven by rational conscious thinking (Kanfer et al., 2001). For example, empirical research suggests that profession search is a goal-directed behavior subject to self-regulation processes such as goal setting, self-monitoring, and self-reactions (Song et al., 2006; van Hooft et al., 2005). Scant research has examined the role of implicit processes or context in important life choices (Highouse & Hoffman, 2001). In our study, drawing from implicit social cognition literature, we examine how the rational processes underlying profession and city-to-live-in choices may be influenced by implicit self-enhancement in the form of people's preference for their first names. Using the latest NLP techniques, we rely on diverse text language models based on large corpora to find a consistent nominative determinism effect in cross-sectional embeddings as well as over the decades of the 20th century. This is important because prior research has questioned the nominative determinism effect, given its lack of appropriate controls.

We also find notable differences in the pattern of the nominative determinism effect across the century for profession versus city-to-live-in choices. While men show a consistent pattern of the nominative determinism effect across the decades for profession choices, women show a much lower effect in the early part of the 20th century, though as time progresses, the effect increases. This Decade × Gender interaction is consistent with the self-identity explanation, which suggests that as alternative means of identity expression become available, the nominative determinism effect diminishes. Mediation by education (as an alternative way to express identity) bears out this theorizing. While we found a significant impact of the nominative determinism effect on the choice of city in which to live, the Gender × Decade interaction was not significant. Future research might explore these differences in the nominative determinism effect in greater detail and assess its application in different domains.

Contributions

While prior research assumes the existence of nominative determinism in profession and city-to-live-in choices, the evidence was said to be of questionable reliability (Gallucci, 2003) and inadequate given the limited names and professions/cities or the use of a proprietary data set (Anseel & Duyck, 2008). Simonsohn (2011a) raised the issue of unobservable heterogeneity and suggested other plausible explanations. Research had also raised questions about the nature of the methods and analyses used (Dyjas et al., 2012; Gallucci, 2003). In our studies, we find evidence for the presence of the nominative determinism effect using a variety of approaches to control for these issues.

Our theoretical contributions are threefold. First, we present evidence of the presence of the nominative determinism effect in profession and city-to-live-in choices across multiple real-life data sets and also across a century, showing the stability of the effect even after controlling for alternative explanations such as alliteration, base frequency, and reverse causality. Second, we report differences in the pattern of the nominative determinism effect for profession versus city-to-live-in choices across the 20th century. Third, we elaborate on the suppressing mediating role of alternative identity expression such as education in the relationship between implicit preference for one's name and job choices.

Our study also offers methodological contributions. First, we use publicly available data sets to encourage reproducibility. Second, we help alleviate the criticism of cherry-picking that can be levied on small data sets by using large text data sets. Third, we examine the effect both cross-sectionally and across an entire century, thereby highlighting the robustness of the effect. Fourth, our use of a pretrained and prevalidated word-embedding method not only effectively addresses the common criticism of experiments (i.e., lack of ecological validity, as the entire internet can be scoured for text data) but also introduces a powerful method from computational social sciences to infer associations present in the real world (Bhatia, 2017, 2019; Garg et al., 2018; Jaidka et al., 2020; Kozlowski et al., 2019; Mooijman et al., 2018).

Limitations and Future Research

One limitation of our research is that we used only single-word professions (vs. multiword professions such as aircraft maintenance technicians) and single-word cities, and thus our findings are generalizable only to similar one-word professions and cities. With advances in word-embedding techniques, future research could broaden the search criteria. Second, due to computational constraints, we used a randomly selected subset of nouns to control for alliteration. Future research can use all the nouns in corpora to control for alliteration. Third, while word embeddings based on NLP systems make excellent predictive models (LeCun et al., 2015; Schmidhuber, 2015) and, as such, were appropriate to use in our context to test for the nominative determinism effect, a criticism from a psychological standpoint is that representations of words are mostly based on text-based patterns and weakly linked to world knowledge (Lake & Murphy, 2023). In addition, text-based theories are in sharp contrast to perceptual symbol systems as a representation of words and meanings (Andrews et al., 2014; Louwerse, 2007), thus raising the concern that distributional models may not capture such a representation. A similar concern is that distributional models do not adequately capture affective systems. While these criticisms of word-embedding techniques do not directly affect our findings, research that investigates these issues would be fruitful.

References

Aaker, J., & Schmitt, B. (2001). Culture-dependent assimilation and differentiation of the self: Preferences for consumption symbols in the United States and China. *Journal of Cross-Cultural Psychology*, *32*(5), 561–576. https://doi.org/10.1177/0022022101032005003

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, *115*(3), 715–753. https://doi.org/10.1162/003355300554881

Allport, G. W. (1937). *Personality: A psychological interpretation* (pp. xiv, 588). Henry Holt.

Alter, A. (2013). *Drunk tank pink: And other unexpected forces that shape how we think, feel, and behave*. Penguin Press.

Andrews, M., Frank, S., & Vigliocco, G. (2014). Reconciling embodied and distributional accounts of meaning in language. *Topics in Cognitive Science*, *6*(3), 359–370. https://doi.org/10.1111/tops.12096

Anseel, F., & Duyck, W. (2008). Unconscious applicants: A systematic test of the name–letter effect. *Psychological Science*, *19*(10), 1059–1061. https://doi.org/10.1111/j.1467-9280.2008.02199.x

Anseel, F., & Duyck, W. (2009). Implicit letter preferences in job choice: An experimental test of the role of cognitive load. *The Journal of Psychology*, *143*(2), 207–223. https://doi.org/10.3200/JRLP.143.2.207-224

Arnold, F., & Kuo, E. C. Y. (1984). The value of daughters and sons: A comparative study of the gender preferences of parents. *Journal of Comparative Family Studies*, *15*(2), 299–318. https://doi.org/10.3138/jcfs.15.2.299

Ashforth, B. E., & Schinoff, B. S. (2016). Identity under construction: How individuals come to define themselves in organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, *3*(1), 111–137. https://doi.org/10.1146/annurev-orgpsych-041015-062322

Bailey, M., Johnston, D. M., Kuchler, T., Stroebel, J., & Wong, A. (2019). *Peer effects in product adoption*. National Bureau of Economic Research. https://doi.org/10.3386/w25843

Baroni, M., & Lenci, A. (2010). Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, *36*(4), 673–721. https://aclanthology.org/J10-4006.pdf

Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. *Journal of Marketing*, *84*(1), 1–25. https://doi.org/10.1177/0022242919873106

Bhatia, S. (2017). Associative judgment and vector space semantics. *Psychological Review*, *124*(1), 1–20. https://doi.org/10.1037/rev0000047

Bhatia, S. (2019). Predicting risk perception: New insights from data science. *Management Science*, *65*(8), 3800–3823. https://doi.org/10.1287/mnsc.2018.3121

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, *3*, 993–1022.

Boghrati, R., & Berger, J. A. (2022). *Quantifying gender bias in consumer culture* (SSRN Scholarly Paper No. 4004777). Social Science Research Network. https://doi.org/10.2139/ssrn.4004777

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186. https://doi.org/10.1126/science.aal4230

Callan, V. J., & Kee, P.-K. (1981). Sons or daughters? Cross-cultural comparisons of the sex preferences of Australian, Greek, Italian, Malay, Chinese and Indian parents in Australia and Malaysia. *Population and Environment*, *4*(2), 97–108. https://doi.org/10.1007/BF01255800

Cliff, N., & Earleywine, M. (1994). *All predictors are "mediators" unless the other predictor is a "suppressor"* [Unpublished manuscript].

Davis, J. A. (1985). *The logic of causal order*. Sage Publications. https://doi.org/10.4135/9781412986212

de Saussure, F. (2011). *Course in general linguistics*. Columbia University Press.

DeFranza, D., Mishra, H., & Mishra, A. (2020). How language shapes prejudice against women: An examination across 45 world languages. *Journal of Personality and Social Psychology*, *119*(1), 7–22. https://doi.org/10.1037/pspa0000188

Derous, E., & Ryan, A. M. (2019). When your resume is (not) turning you down: Modelling ethnic bias in resume screening. *Human Resource Management Journal*, *29*(2), 113–130. https://doi.org/10.1111/1748-8583.12217

Dion, K. L. (1983). Names, identity, and self. *Names*, *31*(4), 245–257. https://doi.org/10.1179/nam.1983.31.4.245

Domingos, P. (2015). *The master algorithm: How the quest for the ultimate learning machine will remake our world* (pp. xxi, 329). Basic Books.

Donald, W. E., Baruch, Y., & Ashleigh, M. (2019). The undergraduate self-perception of employability: Human capital, careers advice, and career ownership. *Studies in Higher Education*, *44*(4), 599–614. https://doi.org/10.1080/03075079.2017.1387107

Dyjas, O., Grasman, R. P. P. P., Wetzels, R., van der Maas, H. L. J., & Wagenmakers, E.-J. (2012). What's in a name: A Bayesian hierarchical analysis of the name–letter effect. *Frontiers in Psychology*, *3*, Article 334. https://doi.org/10.3389/fpsyg.2012.00334

Festinger, L. (1957). *A theory of cognitive dissonance* (pp. xi, 291). Stanford University Press.

Firth, J. R. (1957). A synopsis of linguistic theory, 1930–1955. *Studies in linguistic analysis* (pp. 10–32). Oxford Press.

Frommer, M. (1982). Names. *ETC: A Review of General Semantics*, *39*(2), 106–108.

Gallucci, M. (2003). I sell seashells by the seashore and my name is Jack: Comment on Pelham, Mirenberg, and Jones (2002). *Journal of Personality and Social Psychology*, *85*(5), 789–799. https://doi.org/10.1037/0022-3514.85.5.789

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(16), E3635–E3644. https://doi.org/10.1073/pnas.1720347115

Garten, J., Hoover, J., Johnson, K. M., Boghrati, R., Iskiwitch, C., & Dehghani, M. (2018). Dictionaries and distributions: Combining expert knowledge and large scale textual data content analysis: Distributed dictionary representation. *Behavior Research Methods*, *50*, 344–361. https://doi.org/10.3758/s13428-017-0875-9

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4–27. https://doi.org/10.1037/0033-295X.102.1.4

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). *Diachronic word embeddings reveal statistical laws of semantic change* [Conference session]. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Stroudsburg, PA, United States. https://doi.org/10.18653/v1/P16-1141

Health and Retirement Study. (2015). *2015 life history mail survey*. https://hrsdata.isr.umich.edu/data-products/2015-life-history-mail-survey

Hetts, J. J., Sakuma, M., & Pelham, B. W. (1999). Two roads to positive regard: Implicit and explicit self-evaluation and culture. *Journal of Experimental Social Psychology*, *35*(6), 512–559. https://doi.org/10.1006/jesp.1999.1391

Highouse, S., & Hoffman, J. R. (2001). Organizational attraction and job choice. *International Review of Industrial and Organizational Psychology*, *16*, 37–64.

Hoorens, V. (2014). What's really in a name–letter effect? Name–letter preferences as indirect measures of self-esteem. *European Review of Social Psychology*, *25*(1), 228–262. https://doi.org/10.1080/10463283.2014.980085

Hoorens, V., & Nuttin, J. M. (1993). Overvaluation of own attributes: Mere ownership or subjective frequency? *Social Cognition*, *11*(2), 177–200. https://doi.org/10.1521/soco.1993.11.2.177

Hoorens, V., & Todorova, E. (1988). The name letter effect: Attachment to self or primacy of own name writing? *European Journal of Social Psychology*, *18*(4), 365–368. https://doi.org/10.1002/ejsp.2420180406

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*(4), 309–334. https://doi.org/10.1037/a0020761

Jaidka, K., Giorgi, S., Schwartz, H. A., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2020). Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(19), 10165–10171. https://doi.org/10.1073/pnas.1906364117

James, W. (1890). *The principles of psychology* (Vol. 2). Henry Holt.

Jones, J. T., Pelham, B. W., Carvallo, M., & Mirenberg, M. C. (2004). How do I love thee? Let me count the Js: Implicit egotism and interpersonal attraction. *Journal of Personality and Social Psychology*, *87*(5), 665–683. https://doi.org/10.1037/0022-3514.87.5.665

Jones, J. T., Pelham, B. W., Mirenberg, M. C., & Hetts, J. J. (2002). Name letter preferences are not merely mere exposure: Implicit egotism as self-regulation. *Journal of Experimental Social Psychology*, *38*(2), 170–177. https://doi.org/10.1006/jesp.2001.1497

Jurafsky, D., & Martin, J. (2020). *Speech and language processing* (3rd ed. draft). https://web.stanford.edu/~jurafsky/slp3/3.pdf

Kanfer, R., Wanberg, C. R., & Kantrowitz, T. M. (2001). Job search and employment: A personality-motivational analysis and meta-analytic review. *Journal of Applied Psychology*, *86*(5), 837–855. https://doi.org/10.1037/0021-9010.86.5.837

Karahan, F., & Li, D. (2016). What caused the decline in interstate migration in the United States? *Liberty street economics*. Federal Reserve Bank of New York. https://ideas.repec.org/p/fip/fednls/87158.html

Kim, H. S., & Drolet, A. (2003). Choice and self-expression: A cultural analysis of variety-seeking. *Journal of Personality and Social Psychology*, *85*(2), 373–382. https://doi.org/10.1037/0022-3514.85.2.373

Kleine, R. E., III, Kleine, S. S., & Kernan, J. B. (1993). Mundane consumption and the self: A social-identity perspective. *Journal of Consumer Psychology*, *2*(3), 209–235. https://doi.org/10.1016/S1057-7408(08)80015-0

Koole, S. L., Dijksterhuis, A., & Van Knippenberg, A. (2001). What's in a name: Implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology*, *80*(4), 669–685. https://doi.org/10.1037/0022-3514.80.4.669

Koole, S. L., & Pelham, B. W. (2003). On the nature of implicit self-esteem: The case of the name letter effect. In S. J. Spencer, S. Fein, M. P. Zanna, & J. M. Olson (Eds.), *Motivated social perception: The Ontario symposium* (Vol. 9, pp. 93–116). Lawrence Erlbaum.

Koole, S. L., Smeets, K., Van Knippenberg, A., & Dijksterhuis, A. (1999). The cessation of rumination through self-affirmation. *Journal of Personality and Social Psychology*, *77*(1), 111–125. https://doi.org/10.1037/0022-3514.77.1.111

Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review*, *84*(5), 905–949. https://doi.org/10.1177/0003122419877135

Lake, B. M., & Murphy, G. L. (2023). Word meaning in minds and machines. *Psychological Review*, *130*(2), 401–431. https://doi.org/10.1037/rev0000297

Lea, R. B., Rapp, D. N., Elfenbein, A., Mitchel, A. D., & Romine, R. S. (2008). Sweet silent thought: Alliteration and resonance in poetry comprehension. *Psychological Science*, *19*(7), 709–716. https://doi.org/10.1111/j.1467-9280.2008.02146.x

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Levy, S. J. (1959). Symbols for sale. *Harvard Business Review*.

Liu, Q. B., & Karahanna, E. (2017). The dark side of reviews: The swaying effects of online product reviews on attribute preference construction. *Management Information Systems Quarterly*, *41*(2), 427–448. https://doi.org/10.25300/MISQ/2017/41.2.05

Louwerse, M. M. (2007). Symbolic or embodied representations: A case for symbol interdependency. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 107–120). Lawrence Erlbaum.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, *1*(4), 173–181. https://doi.org/10.1023/A:1026595011371

McCullough, B. D., & McWilliams, T. P. (2011). Students with the initial "A" don't get better grades. *Journal of Research in Personality*, *45*(3), 340–343. https://doi.org/10.1016/j.jrp.2011.03.006

McFatter, R. M. (1979). The use of structural equation models in interpreting regression equations including suppressor and enhancer variables. *Applied Psychological Measurement*, *3*(1), 123–135. https://doi.org/10.1177/014662167900300113

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *26*, 3111–3119.

Miller, L. K. (2013). All text considered: A perspective on mass digitizing and archival processing. *The American Archivist*, *76*(2), 521–541. https://doi.org/10.17723/aarc.76.2.6q005254035w2076

Molloy, R., Smith, C. L., & Wozniak, A. (2011). Internal migration in the United States. *The Journal of Economic Perspectives*, *25*(3), 173–196. https://doi.org/10.1257/jep.25.3.173

Mooijman, M., Hoover, J., Lin, Y., Ji, H., & Dehghani, M. (2018). Moralization in social networks and the emergence of violence during protests. *Nature Human Behaviour*, *2*(6), 389–396. https://doi.org/10.1038/s41562-018-0353-0

Murdock, G. P. (1945). The common denominator of cultures. In R. Linton (Ed.), *The science of man in the world crisis* (pp. 123–143). Columbia University Press.

Nuttin, J. M. (1985). Narcissism beyond gestalt and awareness: The name letter effect. *European Journal of Social Psychology*, *15*(3), 353–361. https://doi.org/10.1002/ejsp.2420150309

Nuttin, J. M., Jr. (1987). Affective consequences of mere ownership: The name letter effect in twelve European languages. *European Journal of Social Psychology*, *17*(4), 381–402. https://doi.org/10.1002/ejsp.2420170402

Obodaru, O. (2016). Forgone, but not forgotten: Toward a theory of forgone professional identities. *Academy of Management Journal*, *60*(2), 523–553. https://doi.org/10.5465/amj.2013.0432

Ooms, J. (2022). *hunspell: High-performance stemmer, tokenizer, and spell checker* (R package Version 3.0.2) [Computer software]. https://CRAN.R-project.org/package=hunspell

Oyserman, D. (2009). Identity-based motivation and consumer behavior. *Journal of Consumer Psychology*, *19*(3), 276–279. https://doi.org/10.1016/j.jcps.2009.06.001

Oyserman, D., & Yoon, K.-I. (2009). Neighborhood effects on racial–ethnic identity: The undermining role of segregation. *Race and Social Problems*, *1*(2), 67–76. https://doi.org/10.1007/s12552-009-9007-1

Padarian, J., & Fuentes, I. (2019). Word embeddings for application in geosciences: Development, evaluation, and examples of soil-related concepts. *Soil*, *5*(2), 177–187. https://doi.org/10.5194/soil-5-177-2019

Pelham, B. W., Mirenberg, M. C., & Jones, J. T. (2002). Why Susie sells seashells by the seashore: Implicit egotism and major life decisions. *Journal of Personality and Social Psychology*, *82*(4), 469–487. https://doi.org/10.1037/0022-3514.82.4.469

Pennington, J., Socher, R., & Manning, C. (2014). *GloVe: Global vectors for word representation* [Conference session]. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar https://doi.org/10.3115/v1/D14-1162

Pyszczynski, T., Greenberg, J., & Solomon, S. (1999). A dual-process model of defense against conscious and unconscious death-related thoughts: An extension of terror management theory. *Psychological Review*, *106*(4), 835–845. https://doi.org/10.1037/0033-295X.106.4.835

Ramu, G., & Tavuchis, N. (1986). The valuation of children and parenthood among the voluntarily childless and parental couples in Canada. *Journal of Comparative Family Studies*, *17*(1), 99–116. https://doi.org/10.3138/jcfs.17.1.99

Rezaei, H. (2022). *Word embeddings are capable of capturing rhythmic similarity of words*. arXiv preprint. https://doi.org/10.48550/arXiv.2204.04833

Ruder, S. (2016). *An overview of gradient descent optimization algorithms*. arXiv preprint. https://doi.org/10.48550/arXiv.1609.04747

Sariyska, R., Reuter, M., Bey, K., Sha, P., Li, M., Chen, Y.-F., Liu, W.-Y., Zhu, Y.-K., Li, C.-B., Suárez-Rivillas, A., Feldmann, M., Hellmann, M., Keiper, J., Markett, S., Young, K. S., & Montag, C. (2014). Self-Esteem, personality and internet addiction: A cross-cultural comparison study. *Personality and Individual Differences*, *61*, 28–33. https://doi.org/10.1016/j.paid.2014.01.001

Sava, F. A., Maricutoiu, L. P., Rusu, S., Macsinga, I., & Virga, D. (2011). Implicit and explicit self-esteem and irrational beliefs. *Journal of Cognitive and Behavioral Psychotherapies*, *11*(1), 97–111. https://www.researchgate.net/publication/235417899_Implicit_and_explicit_self-esteem_and_irrational_beliefs

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Schrage, M. (2014). *The innovator's hypothesis: How cheap experiments are worth more than good ideas*. MIT Press. https://econpapers.repec.org/bookchap/mtptitles/0262028360.htm

Schröder-Abé, M., Rudolph, A., Wiesner, A., & Schütz, A. (2007). Self-esteem discrepancies and defensive reactions to social feedback. *International Journal of Psychology*, *42*(3), 174–183. https://doi.org/10.1080/00207590601068134

Sedikides, C., & Gregg, A. P. (2008). Self-enhancement: Food for thought. *Perspectives on Psychological Science*, *3*(2), 102–116. https://doi.org/10.1111/j.1745-6916.2008.00068.x

Sedikides, C., & Strube, M. J. (1997). Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. *Advances in Experimental Social Psychology*, *29*, 209–269. https://doi.org/10.1016/S0065-2601(08)60018-0

Simonsohn, U. (2011a). Spurious? Name similarity effects (implicit egotism) in marriage, job, and moving decisions. *Journal of Personality and Social Psychology*, *101*(1), 1–24. https://doi.org/10.1037/a0021990

Simonsohn, U. (2011b). Spurious also? Name-similarity effects (implicit egotism) in employment decisions. *Psychological Science*, *22*(8), 1087–1089. https://doi.org/10.1177/0956797611413937

Snibbe, A. C., & Markus, H. R. (2005). You can't always get what you want: Educational attainment, agency, and choice. *Journal of Personality and Social Psychology*, *88*(4), 703–720. https://doi.org/10.1037/0022-3514.88.4.703

Snyder, T. D. (1993). *120 years of American education: A statistical portrait*. National Center for Education Statistics. https://eric.ed.gov/?id=ED355277

Solomon, M. R. (1983). The role of products as social stimuli: A symbolic interactionism perspective. *The Journal of Consumer Research*, *10*(3), 319–329. https://doi.org/10.1086/208971

Song, Z., Wanberg, C., Niu, X., & Xie, Y. (2006). Action–state orientation and the theory of planned behavior: A study of job search in China. *Journal of Vocational Behavior*, *68*(3), 490–503. https://doi.org/10.1016/j.jvb.2005.11.001

Steinberg, J. A., Karpinski, A., & Alloy, L. B. (2007). The exploration of implicit aspects of self-esteem in vulnerability–stress models of depression. *Self and Identity*, *6*(2–3), 101–117. https://doi.org/10.1080/15298860601118884

Stekel, W. (1911). Die verpflichtung des namens [The obligation of the name]. *Zeitschrift für Psychotherapie und Medizinische Psychologie*, *3*, 110–114.

Stewart, B. M., & Zhukov, Y. M. (2009). Use of force and civil–military relations in Russia: An automated content analysis. *Small Wars & Insurgencies*, *20*(2), 319–343. https://doi.org/10.1080/09592310902975455

Stieger, S., & Burger, C. (2010). Implicit and explicit self-esteem in the context of internet addiction. *Cyberpsychology, Behavior, and Social Networking*, *13*(6), 681–688. https://doi.org/10.1089/cyber.2009.0426

Stryker, S. (1980). *Symbolic interactionism: A social structural version*. Blackburn Press.

Stubbs, M. (1996). *Text and corpus analysis: Computer-assisted studies of language and culture*. Blackwell.

Tafarodi, R. W., Mehranvar, S., Panton, R. L., & Milne, A. B. (2002). Putting oneself in the task: Choice, personalization, and confidence. *Personality and Social Psychology Bulletin*, *28*(5), 648–658. https://doi.org/10.1177/0146167202288009

Tataru, C. A., & David, M. M. (2020). Decoding the language of microbiomes using word-embedding techniques, and applications in inflammatory bowel disease. *PLOS Computational Biology*, *16*(5), Article e1007859. https://doi.org/10.1371/journal.pcbi.1007859

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software*, *59*(5), 1–38. https://doi.org/10.18637/jss.v059.i05

Tomlinson, M. (2013). *Education, work and identity*. Bloomsbury. https://eprints.soton.ac.uk/201535/

Toossi, M. (2002). A century of change: The U.S. labor force from 1950 to 2050. *Monthly Labor Review*, *125*(5), 15–28.

Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., Persson, K. A., Ceder, G., & Jain, A. (2019). Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature*, *571*(7763), 95–98. https://doi.org/10.1038/s41586-019-1335-8

Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, *37*, 141–188. https://doi.org/10.48550/arXiv.1003.1141

Tzelgov, J., & Henik, A. (1991). Suppression situations in psychological research: Definitions, implications, and applications. *Psychological Bulletin*, *109*(3), 524–536. https://doi.org/10.1037/0033-2909.109.3.524

Tzioumis, K. (2018). Demographic aspects of first names. *Scientific Data*, *5*(1), Article 180025. https://doi.org/10.1038/sdata.2018.25

van Hooft, E. A. J., Born, M. P., Taris, T. W., van der Flier, H., & Blonk, R. W. B. (2005). Bridging the gap between intentions and behavior: Implementation intentions, action control, and procrastination. *Journal of Vocational Behavior*, *66*(2), 238–256. https://doi.org/10.1016/j.jvb.2004.10.003

Vater, A., Schröder-Abé, M., Schütz, A., Lammers, C. H., & Roepke, S. (2010). Discrepancies between explicit and implicit self-esteem are linked to symptom severity in borderline personality disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, *41*(4), 357–364. https://doi.org/10.1016/j.jbtep.2010.03.007

Watson, R. S. (1986). The named and the nameless: Gender and person in Chinese society. *American Ethnologist*, *13*(4), 619–631. https://doi.org/10.1525/ae.1986.13.4.02a00020

Wille, B., Beyers, W., & De Fruyt, F. (2012). A transactional approach to person–environment fit: Reciprocal relations between personality development and career role growth across young to middle adulthood. *Journal of Vocational Behavior*, *81*(3), 307–321. https://doi.org/10.1016/j.jvb.2012.06.004

Zeigler-Hill, V. (2006). Discrepancies between implicit and explicit self-esteem: Implications for narcissism and self-esteem instability. *Journal of Personality*, *74*(1), 119–144. https://doi.org/10.1111/j.1467-6494.2005.00371.x

Appendix A

Details of the Word-Embedding Algorithms

We needed a unique numerical representation of a word for two reasons: (a) converting a word to a unique numerical representation helps us analyze words just like numbers (e.g., finding similarity between two words through distance metrics) and (b) if the unique representation of a word captures the meaning of the word, this allows us to move beyond a count-based approach (which views each word as presence or absence and does not inform us about the word's meaning) to compare two words by their meaning. This quality of word embedding distinguishes it from other text analysis methods such as topic modeling (Blei et al., 2003) that can uncover latent topics in a document or words in a topic but are unable to compare words by their meanings. Topic models cannot be used if we want to determine the semantic (meaning-based) similarity of two words (e.g., Is cake considered tastier than a salad? Is Apple considered more premium than Dell? Is Dennis likely to live in Denver and be a dentist?).

Word2vec

The popular word-embedding algorithm word2vec (Mikolov et al., 2013) is actually a combination of two methods. We elaborate on one called skip-gram with negative sampling, which Hamilton et al. (2016) used to create their decade-by-decade embeddings. These pretrained and validated embeddings are freely available online along with their code (https://nlp.stanford.edu/projects/histwords/).

The underlying intuition in the word2vec algorithm is that it takes word "x" that is near to a target word—say, "cat"—and trains a binary classifier to answer the following question: Is word "x" likely to show up near word "cat"? The weights that the classifier learns from the task are the word embeddings (Jurafsky & Martin, 2020). Consider the following sentence: "Only the sparkling wine produced in France's Champagne region can be legally labeled with the name." If the context window is ±3 for the target word "wine," the context window will consist of {only, the, sparkling, produced, in, France}. If we denote target word "wine" as t and the other context words as set C = {C1, C2, C3, C4, C5, C6}, the probability that C is the correct context word given the target will be $P(+ |t,C)$. The probability that C is not the correct context word given the target will be $P(-|t,C)$. Thus, $P(-|t,C) = 1 - P(+ |t,C)$.

How is probability P computed? The way probability is computed can be understood based on similarity: If the embedding of a word is similar to the target embedding, the word should be nearer to the target. Similarity between two vectors can be computed by their dot product. The popular metric of similarity, cosine similarity, is simply a normalized dot product. In other words, similarity between a target word and context is simply their dot product:

$$\text{Similarity}(t, C) \approx t \cdot C. \tag{A1}$$

The problem here is that the dot product can take values from $-\infty$ to $+\infty$, so it must be converted to probability using a sigmoidal function, $\sigma(x)$ (as in logistic regression)

$$\sum(x) = \frac{1}{1 + e^{-x}}. \tag{A2}$$

Substituting the dot product (t·C) for x, we get the probability of word C being the appropriate context for target t.

$$P(+ |t, C) = \frac{1}{1 + e^{-t \cdot C}}. \tag{A3}$$

(*Appendices continue*)

For the sigmoid function to act as probability, we need to ensure that the sum of a word being the context word and the word not being the context word adds up to 1.

$$P(-|t, C) = 1 - \frac{1}{1 + e^{-t \cdot C}} = \frac{e^{-t \cdot C}}{1 + e^{-t \cdot C}}. \quad (A4)$$

While these equations do give us the probability for one word, in our case, context will have multiple words. The word2vec algorithm gets around the problem by making a simplifying assumption that all context words are independent, which allows for multiplication of the probabilities.

$$P(+|t, C_{1:k}) = \prod_{i=1}^{k} \frac{1}{1 + e^{-t \cdot Ci}}. \quad (A5)$$

Taking the logarithm of both sides, we get

$$\log P(+|t, C_{1:k}) = \sum_{i=1}^{k} \log \frac{1}{1 + e^{-t \cdot Ci}}. \quad (A6)$$

In summary, the algorithm trains a logistic regression-like classifier to assign probability on the basis of the similarity between the context window of words and the target word. The goal of the classifier is to return the probability that C_t is a real context word (true vs. false) given a tuple of the target word and the candidate context word (e.g., sparkling, wine—true) or the target and the irrelevant word (wine, football—false). This probability is derived from a sigmoid function of the dot product of target word embeddings with each context word embedding. From this, we can argue that the similarity of two words that occur in the same context (e.g., wine and Champagne) will be greater than words in a different context (e.g., wine and football). Thus, given the word embeddings for each target and context word, the probability can be computed.

Word2vec learns embeddings for the words by starting with an initial set of embedding vectors and then using optimization techniques to shift the embedding of each word to be more like the neighboring words and less like words that are distant from it. Again, consider the sentence, "Only the sparkling wine produced in France's Champagne region can be legally labeled with the name." If the target word is "wine" and the context window is ±2, we can consider some positive and negative training instances. For the classifier to be able to differentiate, we do need negative instances. Word2vec (skip-gram with negative sampling) uses more negative instances than positive ones based on a parameter. The negative instances are random words that can be anything but the target word. In the following example, we use parameter value 2 to generate the negative examples.

Positive examples +		Negative examples −			
t	C	t	C	t	C
wine	the	wine	football	wine	bat
wine	sparkling	wine	boat	wine	ship
wine	produced	wine	shoe	wine	lace
wine	in	wine	television	wine	car

With a randomly generated initial set of embeddings, given the positive and negative training examples, word2vec tries to adjust the embeddings such that (a) the similarity between the target word and context word pairs (t, C) from the positive examples is maximized and (b) the similarity between the target word and context word pairs (t, C) from the negative examples is minimized. More formally, taking one target/context pair (t, C) and k negative examples n1, …, nk, the learning objective L is as follows:

$$L(\theta) = \log P(+|t, C) + \sum_{i=1}^{k} \log P(-|t, n_i).$$

$$L(\theta) = \log \frac{1}{1 + e^{-t \cdot C}} + \sum_{i=1}^{k} \log \frac{1}{1 + e^{ni \cdot t}}. \quad (A7)$$

The algorithm then uses a stochastic gradient descent method (Ruder, 2016) to maximize the objective function by iteratively changing the embeddings for each target word and each context word or noise word in the vocabulary on the basis of the constraints. The algorithm creates two embeddings: target word embedding and context embedding. Typically, target word embeddings are used for subsequent analysis.

GloVe

Pennington et al. (2014) developed the GloVe word-embedding algorithm. Just like word2vec, GloVe first creates a co-occurrence matrix by counting the co-occurrence of each word with every other word in the corpus within a context window. Context windows help move beyond a simple count-based approach in text analysis and bring in Firth's hypothesis: that the meaning of a word can be captured by the words appearing along with it.

We explain GloVe using a relevant example. To test for nominative determinism, we examine whether person names starting with the letter "D" are more likely to co-occur in the same context window as profession (and city) names starting with the letter "D" than with any other letter. Then, we create a matrix that counts the number of times a target name (e.g., Dennis) occurs in the context of

(*Appendices continue*)

dentist (compatible attribute) or lawyer (incompatible attribute) in the document. If we define a context window of 10, then 10 words before and 10 words after would be considered for counting co-occurrence. GloVe moves beyond just a count and calculates the probability P_{ji} of a target word j occurring in the context of the attribute word i by using $P(j|i) = \frac{X_{ji}}{X_i}$. Here, X_{ji} is the number of times word j appears in the context of word i, and X_i is the number of times any word in the document appears in the context of i.

If nominative determinism exists in the data sets, we would find that the probability of Dennis (denoted by j) occurring in the context of dentist (denoted by i) is greater than the probability of Larry (denoted by k) occurring in the context of dentist ($P_{ji} > P_{ki}$ or $P_{ji}/P_{ki} > 1$). Or we would find the probability of Larry occurring in the context of lawyer (denoted by l) more than the probability of Dennis occurring in the context of lawyer ($P_{kl} > P_{jl}$ or $P_{kl}/P_{jl} > 1$). However, for profession names incompatible with both Dennis and Larry (e.g., artist, denoted by t), the probability of occurrence is likely to be similar in both contexts ($P_{tj} = P_{tk}$ or $P_{tj}/P_{tk} = 1$). The use of ratio of probability co-occurrences rather than a simple co-occurrences count helps discriminate relevant from irrelevant words and, thus, helps capture the meaning between two words.

The ratio of probability co-occurrences depends on three words i, j, and k, and thus the general model can be represented as $F(w_i, w_j, \vec{w}_k) = \frac{P_{ik}}{P_{jk}}$. Here, w_i is the word vector for word i, and w_j is the word vector for word j. This general model develops into a weighted least-squares cost function. Using a stochastic gradient-descent method, minimizing the cost function yields vectors (numerical dimensions) for each word. Words can be represented as vectors with any number of dimensions, but a smaller number of dimensions such as 30 or 50 lose information. The pretrained embeddings we use for Common Crawl and Twitter data sets have 300 and 200 dimensions for each word, respectively.

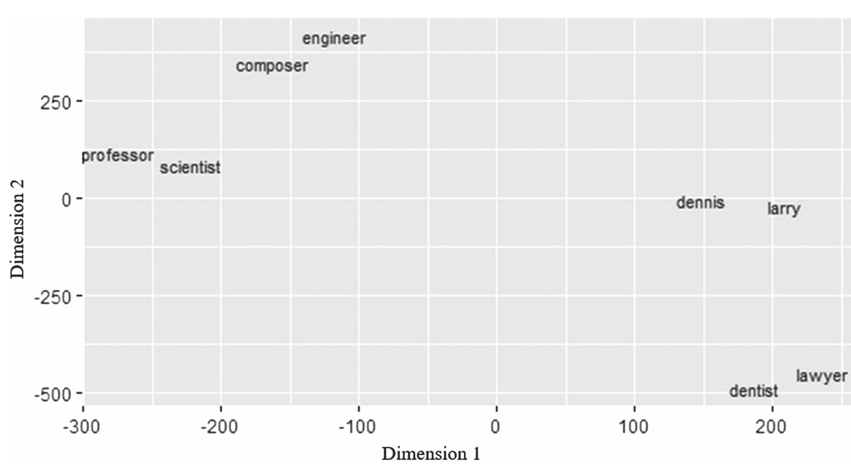
Appendix B

Additional Validation

As an additional illustration of the ability of these algorithms to detect nominative determinism, we provide a few selected two-dimensional visualizations of the embeddings using the t-distributed stochastic neighbor embedding (t-SNE) procedure. The t-SNE refers to a statistical method used in visualization of high-dimensional embeddings for easier human understanding. Figure B1 illustrates how, based on the projection of word embeddings, Dennis is close to the profession of dentist, and Larry is close to the profession of lawyer. Similarly, Figure B2 illustrates how George is close to Georgia and Kenneth is closer to Kentucky.

Figure B1

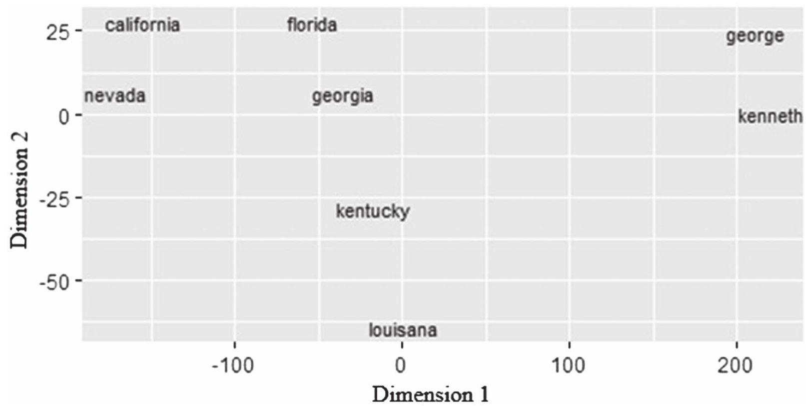
t-SNE Visualization of Names and Professions



Note. t-SNE = t-distributed stochastic neighbor embedding.

(*Appendices continue*)

Figure B2
t-SNE Visualization of Names and Cities

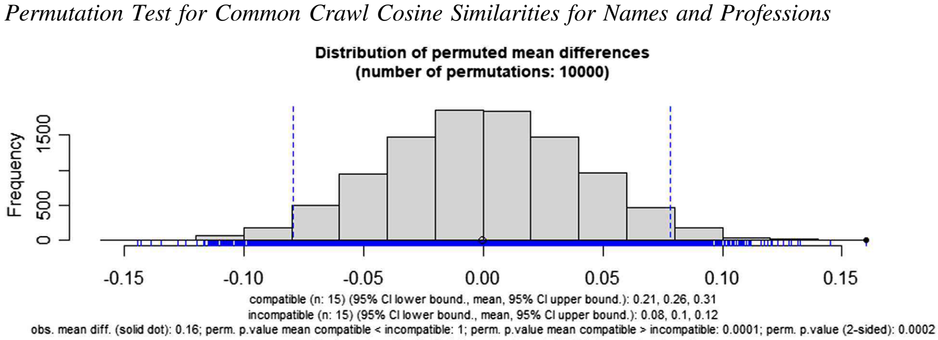


Note. t-SNE = t-distributed stochastic neighbor embedding.

Table B1
Common Crawl Cosine Similarities of Compatible Versus Incompatible Names and Professions

			Cosine similarities	
S. No.	Names	Professions	Compatible	Incompatible
1.	Armstrong	Astronaut	0.312	0.12
2.	Diana	Princess	0.45	0.128
3.	Forbes	Publisher	0.112	0.073
4.	Hughes	Aviator	0.197	0.124
5.	Lewis	Sprinter	0.146	0.16
6.	Monet	Painter	0.316	0.073
7.	Morgan	Pirate	0.246	0.129
8.	Newton	Scientist	0.168	0.093
9.	Nightingale	Nurse	0.251	0.107
10.	Pele	Footballer	0.257	0.026
11.	Phelps	Swimmer	0.412	0.093
12.	Shakespeare	Author	0.147	0.121
13.	Shakira	Singer	0.236	0.047
14.	Teller	Magician	0.414	0.155
15.	Wagner	Composer	0.172	0.094

Figure B3
Permutation Test for Common Crawl Cosine Similarities for Names and Professions



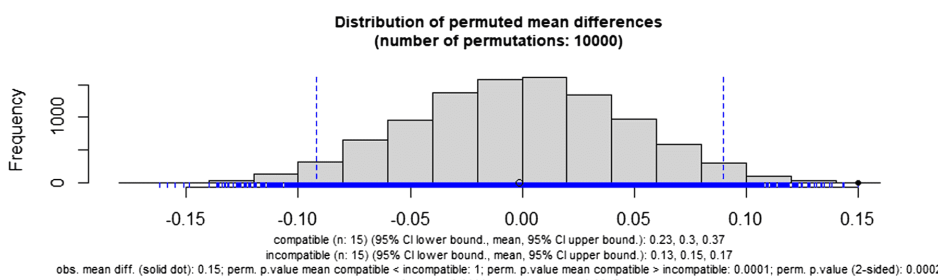
Note. CI = confidence interval; obs. mean diff. = observation mean difference; perm. = permutation. See the online article for the color version of this figure.

(*Appendices continue*)

Table B2
Twitter Cosine Similarities of Compatible Versus Incompatible Names and Professions

			Cosine similarities	
S. No.	Names	Professions	Compatible	Incompatible
1.	Armstrong	Astronaut	0.355	0.186
2.	Diana	Princess	0.575	0.143
3.	Forbes	Publisher	0.423	0.137
4.	Hughes	Aviator	0.14	0.228
5.	Lewis	Sprinter	0.223	0.214
6.	Monet	Painter	0.395	0.151
7.	Morgan	Pirate	0.191	0.199
8.	Newton	Scientist	0.15	0.084
9.	Nightingale	Nurse	0.067	0.104
10.	Pele	Footballer	0.274	0.069
11.	Phelps	Swimmer	0.54	0.159
12.	Shakespeare	Author	0.292	0.16
13.	Shakira	Singer	0.343	0.11
14.	Teller	Magician	0.285	0.176
15.	Wagner	Composer	0.199	0.143

Figure B4
Permutation Test for Twitter Cosine Similarities for Names and Professions



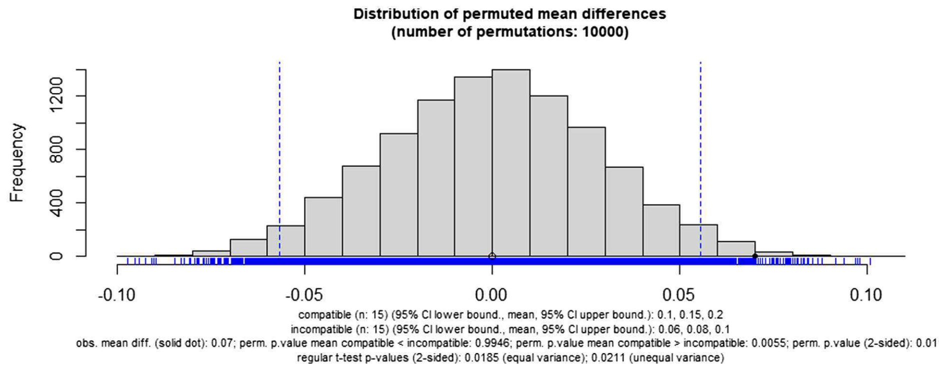
Note. CI = confidence interval; obs. mean diff. = observation mean difference; perm. = permutation. See the online article for the color version of this figure.

Table B3
Google News Cosine Similarities of Compatible Versus Incompatible Names and Professions

			Cosine similarities	
S. No.	Names	Professions	Compatible	Incompatible
1.	Armstrong	Astronaut	0.114	0.12
2.	Diana	Princess	0.275	0.094
3.	Forbes	Publisher	0.087	0.059
4.	Hughes	Aviator	0.013	0.055
5.	Lewis	Sprinter	0.055	0.042
6.	Monet	Painter	0.203	0.051
7.	Morgan	Pirate	0.099	0.036
8.	Newton	Scientist	0.046	0.079
9.	Nightingale	Nurse	0.224	0.207
10.	Pele	Footballer	0.276	0.058
11.	Phelps	Swimmer	0.204	0.053
12.	Shakespeare	Author	0.121	0.086
13.	Shakira	Singer	0.271	0.104
14.	Teller	Magician	0.18	0.12
15.	Wagner	Composer	0.043	0.065

(*Appendices continue*)

Figure B5
Permutation Test for Google News Cosine Similarities for Names and Professions

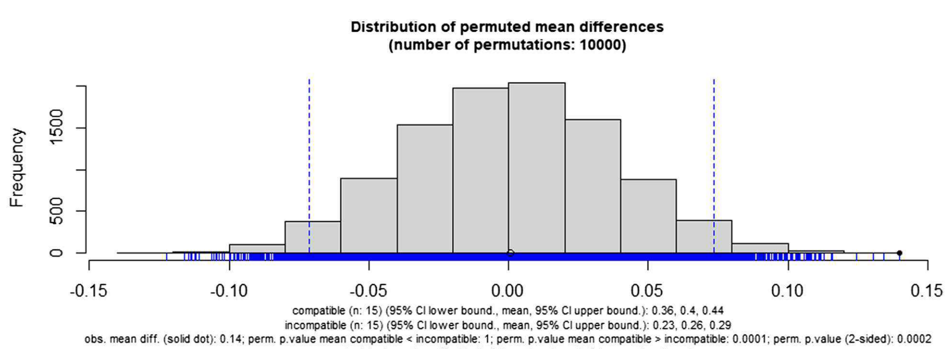


Note. CI = confidence interval; obs. mean diff. = observation mean difference; perm. = permutation. See the online article for the color version of this figure.

Table B4
Common Crawl Cosine Similarities of Compatible Versus Incompatible Names and Cities

S. No.	Names	Places	Cosine similarities Compatible	Incompatible
1.	Alexander	Macedonia	0.433	0.389
2.	Bachchan	Bombay	0.323	0.164
3.	Beckham	Manchester	0.441	0.304
4.	Buffett	Omaha	0.301	0.246
5.	Chanel	Paris	0.453	0.245
6.	Churchill	Kent	0.412	0.331
7.	Elvis	Memphis	0.486	0.312
8.	Gandhi	Delhi	0.46	0.256
9.	Jung	Basel	0.309	0.22
10.	Lennon	Liverpool	0.434	0.275
11.	Newton	Cambridge	0.482	0.332
12.	Obama	Chicago	0.462	0.279
13.	Putin	Moscow	0.459	0.185
14.	Schwarzenegger	Hollywood	0.343	0.165
15.	Thatcher	London	0.218	0.238

Figure B6
Permutation Test for Common Crawl Cosine Similarities for Names and Cities



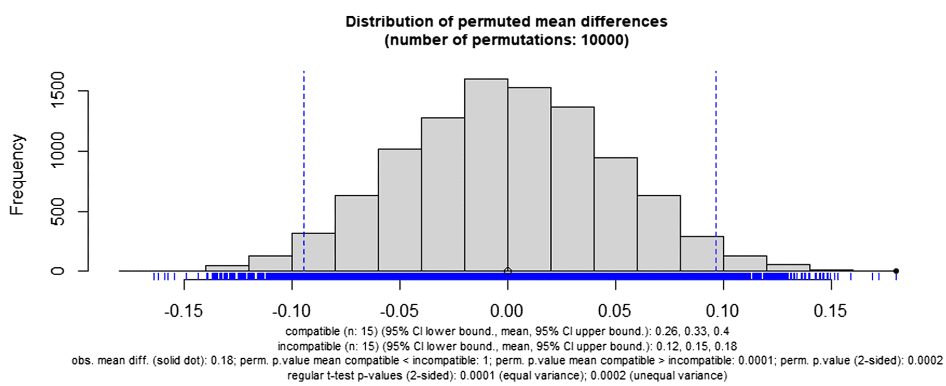
Note. CI = confidence interval; obs. mean diff. = observation mean difference; perm. = permutation. See the online article for the color version of this figure.

(*Appendices continue*)

Table B5
Twitter Cosine Similarities of Compatible Versus Incompatible Names and Cities

			Cosine similarities	
S. No.	Names	Cities	Compatible	Incompatible
1.	Alexander	Macedonia	0.136	0.238
2.	Bachchan	Bombay	0.214	0.085
3.	Beckham	Manchester	0.522	0.212
4.	Buffett	Omaha	0.277	0.093
5.	Chanel	Paris	0.514	0.168
6.	Churchill	Kent	0.201	0.156
7.	Elvis	Memphis	0.326	0.172
8.	Gandhi	Delhi	0.516	0.097
9.	Jung	Basel	0.135	0.135
10.	Lennon	Liverpool	0.392	0.152
11.	Newton	Cambridge	0.304	0.172
12.	Obama	Chicago	0.425	0.221
13.	Putin	Moscow	0.392	0.143
14.	Schwarzenegger	Hollywood	0.211	0.068
15.	Thatcher	London	0.348	0.181

Figure B7
Permutation Test for Twitter Cosine Similarities for Names and Cities



Note. CI = confidence interval; obs. mean diff. = observation mean difference; perm. = permutation. See the online article for the color version of this figure.

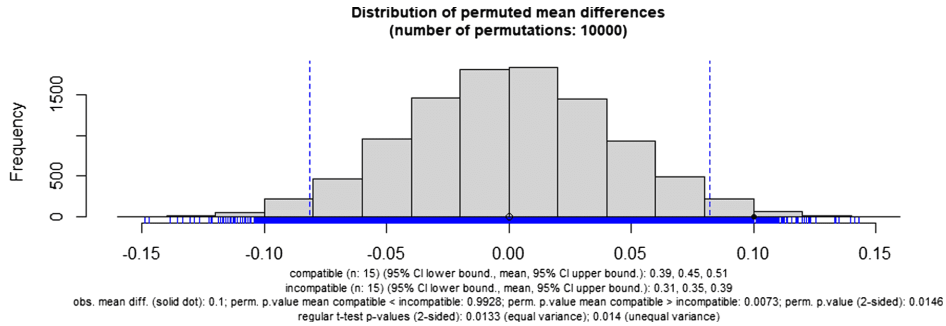
Table B6
Google News Cosine Similarities of Compatible Versus Incompatible Names and Cities

			Cosine similarities	
S. No.	Names	Cities	Compatible	Incompatible
1.	Alexander	Macedonia	0.434	0.466
2.	Bachchan	Bombay	0.504	0.363
3.	Beckham	Manchester	0.541	0.412
4.	Buffett	Omaha	0.344	0.221
5.	Chanel	Paris	0.417	0.351
6.	Churchill	Kent	0.382	0.391
7.	Elvis	Memphis	0.501	0.43
8.	Gandhi	Delhi	0.516	0.362
9.	Jung	Basel	0.14	0.143
10.	Lennon	Liverpool	0.658	0.389
11.	Newton	Cambridge	0.531	0.358
12.	Obama	Chicago	0.514	0.405
13.	Putin	Moscow	0.41	0.314
14.	Schwarzenegger	Hollywood	0.506	0.33
15.	Thatcher	London	0.31	0.254

(*Appendices continue*)

Figure B8

Permutation Test for Google News Cosine Similarities for Names and Cities



Note. CI = confidence interval; obs. mean diff. = observation mean difference; perm. = permutation. See the online article for the color version of this figure.

Table B7

Model Comparison Profession Choices (Common Crawl)

	Dependent variable			
	Cosine similarities			
	OLS	Linear mixed effects		
Variable	(1)	(2)	(3)	(4)
Condition	0.011***	0.011***	0.011***	0.010***
	(0.002)	(0.002)	(0.002)	(0.003)
Average job frequency				0.005
				(0.005)
Gender (male)				−0.007***
				(0.002)
Total frequency year				0.0003
				(0.001)
Ethnicity (White)				0.002*
				(0.001)
Alliteration score				0.001
				(0.002)
Constant	0.133***	0.129***	0.129***	0.133***
	(0.002)	(0.005)	(0.005)	(0.005)
Observations	2,800	2,800	2,800	2,800
R^2	0.008			
Adjusted R^2	0.007			
Log likelihood		3,905.284	3,905.284	3,886.979
Akaike inf. crit.		−7,802.569	−7,800.569	−7,753.959
Bayesian inf. crit.		−7,778.819	−7,770.882	−7,694.585
Residual *SE*	0.062 (*df* = 2,798)			
F statistic	21.409*** (*df* = 1; 2,798)			

Note. OLS = ordinary least square; inf. crit. = information criteria; *df* = degrees of freedom; *SE* = standard error.
* $p < .1$. ** $p < .05$. *** $p < .01$.

(*Appendices continue*)

Table B8
Model Comparison Profession Choices (Twitter)

	Dependent variable			
	Cosine similarities			
	OLS	Linear mixed effects		
Variable	(1)	(2)	(3)	(4)
Condition	0.014***	0.014***	0.014***	0.012***
	(0.002)	(0.002)	(0.001)	(0.001)
Average job frequency				0.003
				(0.004)
Gender (male)				0.014***
				(0.003)
Total frequency year				−0.002
				(0.001)
Ethnicity (White)				0.028***
				(0.001)
Alliteration score				0.002**
				(0.001)
Constant	0.145***	0.138***	0.139***	0.135***
	(0.001)	(0.005)	(0.005)	(0.004)
Observations	6,700	6,700	6,700	6,700
R^2	0.006			
Adjusted R^2	0.006			
Log likelihood		7,260.987	8,604.152	8,804.759
Akaike inf. crit.		−14,513.970	−17,198.300	−17,589.520
Bayesian inf. crit.		−14,486.740	−17,164.250	−17,521.420
Residual *SE*	0.084 (*df* = 6,698)			
F statistic	43.067*** (*df* = 1; 6,698)			

Note. OLS = ordinary least square; inf. crit. = information criteria; *df* = degrees of freedom; *SE* = standard error.
* $p < .1$. ** $p < .05$. *** $p < .01$.

Table B9
Model Comparison Profession Choices (Google News)

	Dependent variable			
	Cosine similarities			
	OLS	Linear mixed effects		
Variable	(1)	(2)	(3)	(4)
Condition	0.011***	0.011***	0.011***	0.013***
	(0.002)	(0.002)	(0.002)	(0.003)
Average job frequency				0.006
				(0.005)
Gender (male)				−0.007***
				(0.002)
Total frequency year				0.0003
				(0.001)
Ethnicity (White)				0.002*
				(0.001)
Alliteration score				−0.003
				(0.002)
Constant	0.133***	0.129***	0.129***	0.132***
	(0.002)	(0.005)	(0.005)	(0.005)
Observations	2,800	2,800	2,800	2,800
R^2	0.008			
Adjusted R^2	0.007			
Log likelihood		3,905.284	3,905.284	3,887.567
Akaike inf. crit.		−7,802.569	−7,800.569	−7,755.134
Bayesian inf. crit.		−7,778.819	−7,770.882	−7,695.760
Residual *SE*	0.062 (*df* = 2,798)			
F statistic	21.409*** (*df* = 1; 2,798)			

Note. OLS = ordinary least square; inf. crit. = information criteria; *df* = degrees of freedom; *SE* = standard error.
* $p < .1$. ** $p < .05$. *** $p < .01$.

(*Appendices continue*)

Table B10
Model Comparison City Choices (Common Crawl)

	Dependent variable			
	Cosine similarities			
	OLS	Linear mixed effects		
Variable	(1)	(2)	(3)	(4)
Condition	0.032***	0.032***	0.032***	0.024***
	(0.003)	(0.003)	(0.001)	(0.001)
Average city frequency				−0.004
				(0.003)
Gender (male)				0.002
				(0.004)
Total frequency year				0.004**
				(0.002)
Ethnicity (White)				0.019***
				(0.002)
Alliteration score				0.008***
				(0.001)
Constant	0.400***	0.394***	0.396***	0.402***
	(0.002)	(0.006)	(0.005)	(0.004)
Observations	6,648	6,648	6,648	6,648
R^2	0.021			
Adjusted R^2	0.021			
Log likelihood		5,340.420	9,483.339	9,635.826
Akaike inf. crit.		−10,672.840	−18,956.680	−19,251.650
Bayesian inf. crit.		−10,645.630	−18,922.670	−19,183.630
Residual *SE*	0.109 (*df* = 6,646)			
F statistic	143.832*** (*df* = 1; 6,646)			

Note. OLS = ordinary least square; inf. crit. = information criteria; *df* = degrees of freedom; *SE* = standard error.
* $p < .1$. ** $p < .05$. *** $p < .01$.

Table B11
Model Comparison City Choices (Twitter)

	Dependent variable			
	Cosine similarities			
	OLS	Linear mixed effects		
Variable	(1)	(2)	(3)	(4)
Condition	0.029***	0.029***	0.029***	0.013***
	(0.003)	(0.003)	(0.001)	(0.001)
Average city frequency				−0.013***
				(0.004)
Gender (male)				0.008**
				(0.004)
Total frequency year				0.001
				(0.002)
Ethnicity (White)				0.039***
				(0.002)
Alliteration score				0.017***
				(0.001)
Constant	0.242***	0.235***	0.235***	0.242***
	(0.002)	(0.009)	(0.008)	(0.006)
Observations	6,700	6,700	6,700	6,700
R^2	0.013			
Adjusted R^2	0.013			
Log likelihood		4,617.753	7,514.956	7,945.269
Akaike inf. crit.		−9,227.507	−15,019.910	−15,870.540
Bayesian inf. crit.		−9,200.267	−14,985.860	−15,802.440
Residual *SE*	0.124 (*df* = 6,698)			
F statistic	90.223*** (*df* = 1; 6,698)			

Note. OLS = ordinary least square; inf. crit. = information criteria; *df* = degrees of freedom; *SE* = standard error.
* $p < .1$. ** $p < .05$. *** $p < .01$.

(*Appendices continue*)

Table B12
Model Comparison City Choices (Google News)

Variable	Dependent variable: Cosine similarities — OLS (1)	Linear mixed effects (2)	Linear mixed effects (3)	Linear mixed effects (4)
Condition	0.046***	0.046***	0.046***	0.038***
	(0.005)	(0.005)	(0.005)	(0.006)
Average city frequency				−0.003
				(0.005)
Gender (male)				0.008
				(0.005)
Total frequency year				0.009***
				(0.003)
Ethnicity (White)				0.010***
				(0.003)
Alliteration score				0.011**
				(0.004)
Constant	0.482***	0.473***	0.473***	0.475***
	(0.004)	(0.007)	(0.007)	(0.007)
Observations	2,800	2,800	2,800	2,800
R^2	0.027			
Adjusted R^2	0.027			
Log likelihood		1,614.153	1,619.087	1,613.181
Akaike inf. crit.		−3,220.306	−3,228.173	−3,206.362
Bayesian inf. crit.		−3,196.557	−3,198.486	−3,146.988
Residual *SE*	0.137 (*df* = 2,798)			
F statistic	77.863*** (*df* = 1; 2,798)			

Note. OLS = ordinary least square; inf. crit. = information criteria; *df* = degrees of freedom; *SE* = standard error.
$^{*}p < .1.$ $^{**}p < .05.$ $^{***}p < .01.$

Table B13
Model Comparison Profession Choice (Decade)

Variable	Dependent variable: Difference score — Generalized least squares (1)	Linear mixed effects (2)	Linear mixed effects (3)	Linear mixed effects (4)	Linear mixed effects (5)
Decade	0.001***	0.001***		0.001***	0.001***
	(0.0004)	(0.0002)		(0.0003)	(0.0003)
Gender (male)	0.018***	0.007***		0.009***	0.009***
	(0.004)	(0.003)		(0.003)	(0.003)
Average job frequency					−0.003***
					(0.001)
Alliteration score					0.014***
					(0.001)
Ethnicity (White)					0.007***
					(0.001)
Name frequency					−0.00005
					(0.001)
Decade × Gender	−0.002***	−0.001**		−0.001**	−0.001**
	(0.001)	(0.0003)		(0.0004)	(0.0004)
Constant	0.014***	0.019***	0.026***	0.018***	0.015***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
Observations	12,662	12,662	12,662	12,662	12,662
Log likelihood	13,952.590	18,049.240	18,062.530	18,300.470	18,385.420
Akaike inf. crit.	−27,895.190	−36,086.480	−36,119.060	−36,586.940	−36,748.830
Bayesian inf. crit.	−27,857.960	−36,041.800	−36,096.720	−36,534.810	−36,666.930

Note. inf. crit. = information criteria.
$^{*}p < .1.$ $^{**}p < .05.$ $^{***}p < .01.$

(*Appendices continue*)

Table B14
Causal Mediation Analysis (Gender = Male)

Description	Estimate	95% CI lower	95% CI upper	p value
ACME	−0.001496	−0.002117	0.0	<2e−16***
ADE	0.002214	0.001408	0.0	<2e−16***
Total effect	0.000718	0.000251	0.0	.002**
Prop. mediated	−2.068104	−6.917837	−0.99	.002**

Note. CI = confidence interval; ACME = average causal mediation effect (total effect − direct effect); ADE = average direct effect (total effect − indirect effect); total effect = direct (ADE) + indirect (ACME); prop. mediated = ACME/total effect; prop. = proportion.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table B15
Causal Mediation Analysis (Gender = Female)

Description	Estimate	95% CI lower	95% CI upper	p value
ACME	−0.00168	−0.00236	0.0	<2e−16***
ADE	0.00300	0.00217	0.0	<2e−16***
Total effect	0.00133	0.00086	0.0	<2e−16***
Prop. mediated	−1.26441	−2.15328	−0.68	<2e−16***

Note. CI = confidence interval; ACME = average causal mediation effect (total effect − direct effect); ADE = average direct effect (total effect − indirect effect); total effect: Direct (ADE) + indirect (ACME); prop. mediated = ACME/total effect; prop. = proportion.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table B16
Model Comparison in City Choices (Decade)

	Dependent variable				
	Difference score				
	Generalized least squares	Linear mixed effects			
Variable	(1)	(2)	(3)	(4)	(5)
Decade	−0.003***	−0.002***		−0.002***	−0.001***
	(0.0001)	(0.0001)		(0.0001)	(0.0001)
Gender (male)	0.003**	0.0003		0.00005	0.0003
	(0.001)	(0.001)		(0.001)	(0.001)
Average city frequency					−0.003***
					(0.0004)
Alliteration score					−0.001***
					(0.0003)
Ethnicity (White)					0.003***
					(0.0005)
Name frequency					0.001***
					(0.0002)
Decade × Gender	−0.001***	−0.0003***		−0.0002	−0.0002
	(0.0002)	(0.0001)		(0.0001)	(0.0001)
Constant	0.048***	0.043***	0.028***	0.041***	0.039***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	12,667	12,667	12,667	12,667	12,667
Log likelihood	26,880.450	31,102.730	30,526.530	31,632.800	31,670.800
Akaike inf. crit.	−53,750.900	−62,193.460	−61,047.070	−63,251.600	−63,319.600
Bayesian inf. crit.	−53,713.670	−62,148.780	−61,024.730	−63,199.470	−63,237.690

Note. inf. crit. = information criteria.
* $p < .1$. ** $p < .05$. *** $p < .01$.

Received November 28, 2021
Revision received March 9, 2023
Accepted April 12, 2023 ■