SHAWN F. DORIUS AND JEFFREY SWINDLE
Iowa State University; University of Michigan
Email: sdorius@iastate.edu

Developmental Idealism in Internet Search Data

ABSTRACT Scholarship on developmental idealism demonstrates that ordinary people around the world tend to perceive the level of development and the specific characteristics of different countries similarly. We build on this literature by examining public perceptions of nations and development in internet search data, which we argue offers insights into public perceptions that survey data do not address. Our analysis finds that developmental idealism is prevalent in international internet search queries about countries. A consistent mental image of national development emerges from the traits publics ascribe to countries in their queries. We find a positive relationship between the sentiment expressed in autocomplete Google search queries about a given country and its position in the global developmental hierarchy. People in diverse places consistently associate positive attributes with countries ranked high on global development indices and negative characteristics with countries ranked low. We also find a positive correlation between the number of search queries about a country and the country's position in indices of global development. These findings illustrate that ordinary people have deeply internalized developmental idealism and that this informs their views about countries worldwide. KEYWORDS developmental idealism, culture, national images, social hierarchy, internet search data

INTRODUCTION

This study leverages Google search data to investigate whether developmental idealism is evident in the attributes people associate with countries around the world. Our interest is to extend understanding of how people in different places define development by contrasting the attributes lay publics associate with countries in different developmental categories (as defined by influential international organizations such as the United Nations or World Bank). To do this, we collect and analyze a corpus of Google search queries about countries.

Our inquiry draws on recent theory and empirical research on the worldwide prevalence of cultural models regarding the nation-state and national development that scholars have referred to as developmental idealism (Thornton 2001, 2005; Thornton, Dorius, and Swindle 2015; see also research on *world society* such as Hwang 2006; Meyer and Hannan 1979; Meyer et al. 1997). As a generalized worldview, *developmental idealism* (hereafter DI) constitutes a coherent network of perceptions, beliefs, values, expectations, roles, and scripts about how the world works, how it is organized, how a person should live in the world, and what goals individuals and societies should pursue (Thornton 2001, 2005; Thornton, Dorius, and Swindle 2015). One of the organizing principles of DI is *developmental hierarchy*, or the idea that countries can be socially ordered according to attributes thought to be

Sociology of Development, Vol. 5, Number 3, pps. 286–313. electronic ISSN 2374-538X. DOI: https://doi.org/10.1525/sod.2019.5.3.286.

related to development, such as health, wealth, education, and gender equality (Merry, Davis, and Kingsbury 2015; Towns and Rumelili 2017). Developmental hierarchy derives from social scientific theories that define development as progress away from the characteristics of traditional life and toward the attributes of modern life (Nisbet 1969, 1986). Illustrative of the attributes ascribed to societies imagined to be modern and developed are free markets, universal education, health, security, personal freedom, democracy, low fertility, gender equality, happiness, and the rule of law. Countries portrayed as less developed are depicted as having high fertility and mortality, famine, disease, low education, corruption, arranged marriages, political instability, and a great many other features long associated with "traditional" society. A hierarchical view of the world according to perceived levels of development has been widely disseminated by social scientists, societal elites, colonizers, and religious missionaries, among other mechanisms (Thornton, Dorius, and Swindle 2015).

Emerging evidence indicates that belief in developmental hierarchy has now spread to ordinary people in many countries. International social survey data show that irrespective of their country of origin, people tend to rate countries in nearly the same hierarchical order, and that this ordering of countries closely matches global development indices such as the Human Development Index and GDP per capita (Binstock et al. 2013; Csánóová 2013; Dorius 2016; Swindle, Dorius and Melegh 2019; Kiss 2017; Lai and Mu 2016; Lai, Mu, and Thornton 2015; Melegh et al. 2013, 2016; Thornton and Yang 2016; Thornton et al. 2012). Some have interpreted the cross-national uniformity of people's developmental ratings as evidence of a relatively universalized understanding of development among both ordinary people and societal elites. Our interest is to leverage Google search data in service of measuring the beliefs and attitudes of ordinary people from different places toward various countries of the world. In particular, we are interested to know what kinds of attributes people freely associate with different countries and whether these attributes are related to hierarchical views of countries.

In the research that follows, we report results from an exploratory study that leverages a large qualitative bank of Google search queries about countries' attributes. Each search is a "speech act" (Searle 1969) that captures a searcher's belief about what a given country is like. We use summary data on the most common queries of this type made by English-language Google users in each of nearly 200 countries. These data constitute a real-time, efficient source of information about the nature of beliefs about countries. They are not equivalent to survey data, but rather constitute a valuable and complementary source of mass qualitative observations that have been quantified and aggregated in terms of prevalence and geographic location (Bail 2014; Gross and Mann 2017; Lazer et al. 2009; Salganik 2017). They allow us to substantially expand the geographic scope of research on how people think about countries, including the prevalence of DI in Google search queries. We expect that the attributes Google users associate with countries align with DI narratives that link positive societal characteristics to countries thought to be developed and negative societal characteristics to countries imagined as less developed. This leads us to develop specific hypotheses about: (1) the prevalence of developmental content in public perceptions of countries; (2) the relationship between public sentiment about a given country and its position in developmental indices; and (3) the relationship between public interest in a country and the country's position in development indices.

THEORY AND APPROACH

The Importance of National and Developmental Perceptions

Public perceptions of countries, which we refer to as *national perceptions*, and public perceptions of societal development, which we refer to as *developmental perceptions*, are of interest because of their influence on social action and the organization of the world. Like other widespread stereotypes, biases, and prejudices, national and developmental perceptions influence human behavior. Illustrative of the importance of such perceptions is the large marketing industry devoted to shaping public perceptions of countries, where societal elites expend considerable national resources to brand their countries as modern, developed, and in possession of unique cultural heritage (Anholt 2010; Rivera 2008). These two types of perceptions—national and developmental—are conceptually distinct but overlap and inform one another. For example, if a person views Sweden as developed, then they might project characteristics of Sweden onto other countries they view as developed. Likewise, if a person believes that declining religiosity increases national development, then they might assume that developed countries are less religious. Positive perceptions of countries can lead to more tourism, more foreign direct investment, more favorable sovereign credit ratings and trade deals, and a host of other social and economic benefits to society. Negative perceptions of a country can have the opposite effects, leading to lower quality of life and shorter life expectancy.

Our claim that national and developmental perceptions affect social life derives from a long line of social science research. This view motivated, for example, one of UNESCO's first research investigations—a cross-national survey titled "National Stereotypes and International Understanding"—shortly after the founding of the United Nations following World War II (Buchanan 1951; Klineberg 1951; see also Rangil 2013:69–77). The authors argued that erroneous perceptions of one nationality by another both directly and indirectly contributed to international hostilities. Survey respondents from Australia, Mexico, the United States, and several European countries were asked to evaluate members of their own country as well as Americans, Russians, French, Chinese, and British nationals. Respondents selected positive and negative attributes (from a predetermined list) they associated with each country. A general pattern that emerged from the data was that majorities in every surveyed country associated positive attributes with the countries of Northwest European ancestry and negative attributes with countries that have been historically associated with Orientalism or otherwise perceived as outside "the West." Russia and China, for example, were most frequently associated with negative characteristics such as *backward* and *cruel*.

Much of the existing research on national and developmental perceptions uses survey data, which has allowed researchers to generate population-level estimates of various developmentalist beliefs. The downside is that certain components of perceptions go unobserved on surveys. From a cognitive-theoretical perspective, some aspects of people's perceptions may be too subtle, hidden, implicit, or unconscious to be captured in surveys alone, especially perceptions that are most foundational in driving individual behavior (Lizardo 2017; Patterson 2014). This is not to say that surveys are not a powerful tool for measuring perceptions—they are—but to say that there is more to perceptions than is collected from

surveys (Johnson-Hanks et al. 2011; Vaisey 2009, 2014). Next we outline how the design of surveys measuring national and developmental perceptions have been deployed to produce insights into cross-national differences of public opinion.

Existing Research on National and Developmental Perceptions with Surveys

The design and findings of two recent surveys fielded by Melegh and colleagues (2013, 2016) among publics in Europe highlight the contributions of, and gaps in, existing research on national and developmental perceptions. They also demonstrate the constraints of the sole reliance on survey data for gleaning insights into these types of perceptions.

Mental images of development. Melegh et al.'s (2013, 2016) surveys first asked respondents to rate countries' level of development, without being given a definition of "development." This approach ensured that respondents would rate countries according to their own understanding of development rather than the researchers' understanding (Thornton et al. 2012). These and other surveys using the same research design, conducted in dozens of countries, show that people in different places tend to hierarchically organize the nations of the world by perceived level of development in a very similar order (Binstock et al. 2013; Csánóová 2013; Kiss 2017; Lai and Mu 2016; Lai et al. 2015; Thornton and Yang 2016; Thornton et al. 2012). This research thereby provides a glimpse of how people think about countries and their development. In effect, they constitute mental maps of development as they exist in the minds of ordinary people around the world.

The innovation in the surveys by Melegh et al. (see also Dorius 2016; Swindle, Dorius and Melegh 2019) was to follow up the rating exercise by asking respondents what attributes they were thinking about when they rated countries by perceived level of development. They offered respondents 11 possible answers, from which respondents identified the economy, governance, education, science and technology, freedom, gender equality, and fertility rates as most central to their thinking. Here, Melegh at al.'s findings speak not only to respondents' mental *maps* of development, but to their mental *images* of development, as an ideal type and not related to a specific nation. However, it is unclear whether respondents had different or additional attributes in mind when rating countries, since they could only choose from a predetermined list. Are the mental images of development captured in surveys fully reflective of peoples' mental images of development? Or are they missing portions of their visions of development?

Goodness and development Researchers' provision of possible societal characteristics that frame respondents' thinking about development is also limited with respect to sentiment. The preset lists of attributes found in most surveys are largely descriptive and lacking emotion. This is limiting because perceptions of nations and development are emotion-laden, not merely descriptive (Binstock and Thornton 2007; Thornton 2005). Open-ended questions about why they gave a country a particular development score could capture more of the emotional valence behind respondents' national and developmental perceptions. Yet even then, the survey setting prompts respondents for their feelings on the spot, as opposed

to observing emotions in a "natural" setting, and thus may not adequately capture respondents' feelings about development or different countries.

The emotions, or sentiments, underlying individuals' national and developmental perceptions are of interest because DI defines modern society as good and desirable, and DI scholarship confirms that people in many different countries view modern society as positive and valuable (Allendorf and Thornton 2015; Thornton, Dorius, and Swindle 2015). It thus follows that people may conceptualize development and goodness as flowing together. Do people tend to express more positive sentiments when thinking about and describing countries perceived as developed than they do when considering countries they view as less developed?

Visibility Melegh et al. (2013, 2016) also analyzed their respondents' rankings of countries' level of development in a novel way. Unlike previous research, which had treated "don't know" responses to country-rating questions as missing data, Melegh et al. explicitly examined the prevalence of "don't know" responses. They observed that respondents were more likely to report "don't know" when rating countries with lower scores on global development indices. They theorized that the visibility of a nation among publics and the perceived level of development of a nation co-occur because people tend to fixate on countries they view as powerful and developed. People may be more interested in countries they perceive as developed or that rank high on global indices of development, such as the Human Development Index, because (a) they are more likely to be exposed to information about high-status countries, and (b) people tend to model their behaviors and styles of life after actors they perceive as having high status (Fiske 2011).

But it is hard to see how to generalize Melegh et al.'s findings on a country's visibility and its perceived level of development. Those surveys were conducted in a few Eastern European countries, a region where scholars argue that people tend to have some peculiarities in their national and developmental perceptions (Swindle, Dorius and Melegh 2019; Melegh 2006; Todorova 1997). The issue of generalizability is a persistent challenge in research on national and developmental perceptions because of the cost and labor required to field large, multinational surveys. This leads us to consider alternative methods for collecting and measuring such perceptions.

Given the gaps and opportunities in the literature on national and developmental perceptions based on extensive survey data collection and analyses, we propose that data containing people's unprompted language about countries provide novel insights into mental images of development, the relationship between perceptions of a nation's goodness and development, and a country's visibility among various publics. Before making the case for such data and how they can be gathered, we first consider how perceptions inform, and are embedded in, the language and cultural keywords people use in everyday life.

Measuring National and Developmental Perceptions in Language

National and developmental perceptions, like other stereotypes and beliefs, are often manifest in language. Certain words, phrases, and metaphors offer a window into speakers' personal cultural schemas. When these "cultural keywords" are written, spoken, or read, the

beliefs on which they are based can spread to those who are exposed to them, where they can then influence subsequent behavior (D'Andrade 2005; Franzosi 2010; Franzosi, de Fazio, and Vicari 2012; Ignatow 2016; Quinn 2005; Strauss 2005; see also de Saussure 1964; Searle 1969).

In a recent study that explored the historical prevalence of cultural keywords associated with national development and social hierarchy, Swindle (2019) argued that the language of development used in books reinforces and propagates belief in a developmental hierarchy. Swindle analyzed data gleaned from millions of English-language books and found that cultural keywords that hierarchically classify societies, such as *savage* and *civilized*, *primitive* and *advanced*, *Third World* and *First World*, and *developing* and *developed*, have been common for at least 300 years, albeit with some variation. We model the research presented in this article after Swindle's theoretical approach to the relationship between people's language and their national and developmental perceptions.

Consider one contemporary example that illustrates how a person's perceptions are embedded in their language and how emotional sentiment is signaled by their word choice. On January 11, 2018, during a meeting to discuss immigration with members of Congress, U.S. President Donald Trump reportedly refused to offer additional visas for immigrants from El Salvador, Haiti, and several African nations, asking, "Why are we having all these people from sh**hole countries come here?" (Davis, Stolberg, and Kaplan 2018). This comment reflects a starkly hierarchical conception of the world, contrasting the United States with "sh**hole countries." This degree of vulgarity and pejorative negativity would frequently go undetected by surveys owing to social desirability norms, but such views can be captured in observations of people's language use in more "natural" settings.

Perceptions in Internet Search Data

Internet search data offer a unique opportunity to measure perceptions of countries and development. When people use the internet to seek information, they sometimes implicitly disclose their beliefs about the objects or issues in question (e.g. restaurants, countries, or musicians). Fortunately, the companies that run popular search engines store individuals' search queries, collecting a trove of novel data on people's language use and perceptions that can provide a range of sociological insights. Investigations have used online search data to measure economic activity, predict influenza spread, detect infectious disease outbreak, forecast stock market volatility, monitor population-level suicide risk and depression incidence, and aid in the diagnosis of HIV (Ginsberg et al. 2009; Jena et al. 2013; McCarthy 2010; McLaren and Shanbhogue 2011; Wilson and Brownstein 2009; Yang et al. 2010). Researchers have also used search data to measure human perceptions about the prevalence of racist attitudes, how such attitudes relate to health outcomes, and endorsement of various conspiracy theories (Chae et al. 2015; DiGrazia 2017; Stephens-Davidowitz 2014).

The appeal of search data over conventional survey data for our global inquiry regarding public perceptions of countries rests on their volume, timeliness, and geographic scope (Lazer et al. 2009; Yang et al. 2010). The production of search data never stops, and in many cases data are available in real time—a far cry from the slower and more cross-sectional orientation of most social survey data collections (Bail 2014). The scale of search data is well

beyond conventional social scientific data systems, both in the rate at which they are generated and in their geographic scope, extending to anyone around the world with an internet connection. Finally, they can be obtained with far fewer resources than conventional methods because search engine firms have made some of them publicly available to developers and researchers, though usually only in aggregated form (Salganik 2017).

It has also been argued that search data are less susceptible to social desirability bias (Stephens-Davidowitz 2017). The idea here is that internet users are afforded a perceived, and often real, level of anonymity that can yield novel data. For example, research on racial ideology shows that the anonymity of the internet creates a safe space for the expression of racist attitudes and beliefs that have long been deemed unacceptable in public life (Bargh and McKenna 2004; Steinfeldt et al. 2010; Stephens-Davidowitz 2014). This is attractive for our interest in national and developmental perceptions because people who have especially hierarchical views of the world might be more likely to express such views privately online than in a survey interview where their identity is known.

Another appealing feature of search data is that the sentiment of expressions can be taken into account. Though they do not capture the full range of meaning and emotion that can be grasped through qualitative research methods, search data offer more opportunity to measure emotion than traditional survey data, through sentiment analysis. Sentiment, which can be positive, neutral, or negative, may be inferred from the positivity or negativity of the words in a given search query, especially adjectives. For example, a restaurant with a large number of reviews that include the terms *terrible*, *poor service*, and *lost reservations* reflects a decidedly negative customer sentiment toward the restaurant. Sentiment analysis uses human coders and computational methods such as machine learning to study people's opinions, sentiments, emotions, and attitudes, as expressed in written language (Liu et al. 2012; Taboada et al. 2011). This means that the sentiments expressed in internet search data can be quantified and compared.

What Internet Search Data Cannot Do

Search data are not without limitations. At present, they suffer from lack of transparency, poor replication, questions concerning measurement stability and reliability, and issues related to generalizability of findings to known populations (Mellon 2014, 2017; Mellon and Prosser 2017; Salganik 2017). Because nearly all search data are privately held by search engine firms, it is rare for a researcher to gain access to the raw data or to the algorithms and related data-generating technologies that influence search data results. This lack of transparency inhibits open science and makes it difficult to evaluate internet search data as fully as we would like (Lazer et al. 2014). Moreover, the search technologies and algorithms themselves are frequently updated, sometimes making replication of studies impossible. Another criticism of search data is that the anonymity afforded by the internet enhances the importance of group-level social identities, which may lead to greater reliance on old national-ethnic stereotypes (Baker and Potts 2013; Bargh and McKenna 2004; Spears et al. 2002). This has led some to argue that search engines perpetuate harmful and inaccurate stereotypes under the premise of algorithmic integrity (Graham and Sengupta 2017; Noble 2018).

Among the most important differences between survey data and internet search data is representation. Search data do not constitute representative samples of known populations, but rather tend to be aggregate measures of online behaviors, often disassociated from particular users. This provides excellent estimates of online behavior as it actually is, but not estimates of the *average* online behaviors of a randomly selected sample from a known population. The many "digital divides" that exist in terms of internet access, use, and ability compound the challenge of generalization from search data (Dimaggio et al. 2001; Guillén and Suárez 2005). In many countries, internet users (including Google users) tend to be more educated, wealthy, young, male, and urban, and divides also fall along national-level characteristics such as government regime and size of economy, but these individual- and national-level divides are greatly narrowing over time (Fatehkia, Kashyap, and Weber 2018; Garcia et al. 2018; Rath 2016; Stier 2017; Straumann and Graham 2016). Search data nonetheless hold the potential to offer valuable insights that can extend research on national and developmental perceptions.

Research Hypotheses

In the research presented below, we collected Google search data to produce a data set of the most prevalent characteristics associated with countries in Google search queries. Google's "autocomplete" function tabulates the most commonly occurring words in search queries about specific countries. As an example, Figure 1 shows the attributes that English-language Google search queries originating in the United States most commonly associated with China and Norway when asking "Why is [country] so . . . ?" We find this search query especially appealing because it represents someone who is seeking a cause ("why") to explain a belief or perception ("is [country] so . . . ?"). People who ask "Why is Switzerland so rich" are not asking *whether* Switzerland is rich but *why* Switzerland is rich. The searcher perceives Switzerland as rich and presumably is interested to learn the cause.[1] We interpret these attributes from search queries as approximations of *public* perceptions about the target country. While we are not the first to use this search query to collect data about countries (Straumann and Graham 2014), ours is the first attempt to relate these data to DI scholarship.

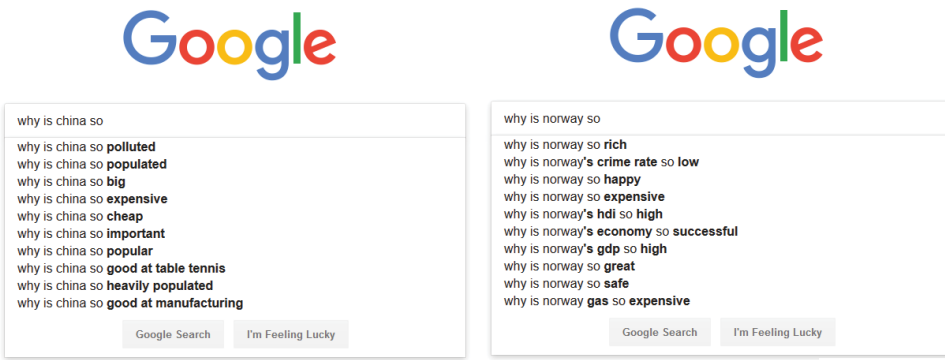

FIGURE 1. Measuring national and developmental perceptions in Google search data

The preceding considerations imply three hypotheses regarding public perceptions of countries for which Google search data are uniquely suited. These hypotheses are motivated by existing survey research on the prevalence of DI among lay publics and based on our theory regarding linkages between national and developmental perceptions and key concepts in theory about DI.

Our first hypothesis deals with the developmental content of peoples' perceptions of the attributes of countries. The accumulating research suggests that DI has given people mental maps of world developmental hierarchy. In that map, Western European countries, including Western European diaspora countries, are the most developed, while the countries of sub-Saharan Africa are the least developed. We seek to extend understanding of *mental maps* of development to *mental images* of development. To do this, we identify and contrast country attributes associated with countries classified as less developed on global development measures against those classified as highly developed. We expect that the characteristics people associate with a given country in their Google search queries depend on the country's position in the world developmental hierarchy. Countries that place high on global development indices are likely to be associated with the national characteristics DI defines as reflective of high development, such as wealth, health, education, and freedom, while countries ranked low on global development measures will be associated with characteristics that DI defines as representing low development, such as poverty, morbidity, illiteracy, gender inequality, and violence.

Our second hypothesis involves the affective orientation, or sentiment, of world publics toward more and less developed countries, as defined by their positions in developmental indices. Given the theoretical claim that perceptions of goodness and development go hand in hand, we expect that the overall sentiment toward a country, inferred from the attributes people ascribe to them, will be closely associated with its level of development. The more developed a country is, the more positive we expect the sentiment of global search queries about it to be.

Our third hypothesis relates a country's *visibility* in online search data to its placement along developmental indices. When the same set of countries are consistently rated near the top of various world rankings and are discussed often and favorably in the press (Csánóová 2013), people are more likely to be aware of them and seek out additional information about them. In contrast, countries that frequently place low in world rankings tend to have fewer connections to other countries and, as a result, are discussed less frequently. Poor countries, for example, are less likely to receive tourists, foreign investment, and trade in goods and services than rich countries. Accordingly, we expect that people will conduct fewer Google search queries about countries that rank low on development indices.

In summary, we propose that individuals' national and developmental perceptions are tightly linked to their exposure to and internalization of DI. DI has been disseminated throughout the world and greatly informs the way large majorities in many countries perceive the world. Perceptions about countries and development are therefore likely to reflect the messages of DI, in particular the arrangement of the world along a spectrum of developed and developing societies, in similar order to what is found in many world development metrics. Specifically, we assess: (1) the attributes people commonly associate with countries

(*images of development*) and whether these characteristics match developmental narratives; (2) the relationship between the overall sentiment (*goodness*) of the national characteristics that people relate to more versus less developed countries as defined by global development indices; and (3) the relationship between the relative number of search queries about countries' attributes (*visibility*) and countries' position on global development indices.

DATA AND METHODS

Data Collection

Data for this study were obtained from the search prediction database behind Google's autocomplete function (hereafter "autocomplete"). Internet users who use Google's search engine will have interacted with this product. Autocomplete is the program that attempts to guess what a Google user is looking for and recommends up to 10 similar queries made by other searchers. It does this by accessing historical search data to identify either exact matches or similar queries, which are then presented to the searcher, as illustrated in Figure 1. To make predictions, the Google autocomplete algorithm relies on an individual user's personal search history (the words and phrases this person has used in previous search queries), the search queries of other people in the user's area, the search language, and trending stories (Google 2017).

Google has made some of its autocomplete search data available to developers and researchers through an application programming interface, or API. An API offers a stable method for obtaining structured data from the Google search history database. Autocomplete suggestions receive modest filtering, largely based on the prevailing moral standards against, for example, hate speech, violence, pornographic or related adult content, personally identifiable material, and some illegal activity, such as piracy (Diakopolous 2015:405). Because the Google search database relies on place-based information to localize its search predictions—think how unhelpful a search for "best restaurants near me" would be if location was not considered—it is possible to leverage country-specific Google domains (e.g. google.co.*ca* for Canada and google.co.*jp* for Japan) to gather Google search histories as they emanate from individual countries.[2]

Data collection, which occurred from August 28 to September 3, 2016, was based on the search query "Why is [country] so . . . ?" We replaced [country] with the names of each of 194 countries,[3]conforming terms from our analysis including Hong Kong and Puerto Rico, and extracted the top 10 Google search terms associated with each place. This approach gave us the 10 most common queries about each country as compiled from Google search queries by people in various countries. We refer to the country from which the search emanated as the *public* and we refer to the target country (the one named in the search) as the *stimulus* country. From each public, we collected the top 10 suggestions following our search query, and we did this for each of the 194 stimulus countries. This approach produced a data set of 376,360 cells, including 10 cells for each public–stimulus country dyad (194 publics × 194 stimulus countries × 10 possible suggestions = 376,360 cells). A highly salient stimulus country, for which data for the top 10 attributes was generated from every searching public, would yield 1,940 terms (10 terms × 194 publics), while a small, relatively unknown country would result in far fewer suggestions. We excluded prior personal search history from

consideration in the search queries we collected. We also virtually changed the location of our computer to each different country (*public*). We only provided our virtual location at the country level so that no further geographic information, such as city, informed the algorithm in the data we collected.

Data preparation

Our data collection yielded 228,528 *total* search results, or terms. Theoretically, there could have been 228,528 unique terms, but in actuality the vast majority were suggested many times, and only 403 *unique* terms appeared. Furthermore, 61 of these 403 terms were directed at something other than a country. Suggestions in which the target was not the country included: "Why is Kuwait Airways cheap," "Why is the Honduras airport dangerous," "Why is the Bahrain dinar so strong," and "Why is the Rock of Gibraltar famous." These four examples focused on an airline, an airport, a currency, and a geological structure, respectively, rather than the country. While these suggestions are similar to our query (which is why autocomplete suggested them), they are not explicitly focused on countries. Because our interest was to develop an attribute list that followed from search queries beginning with "Why is [country] so . . .," we excluded the 61 non-conforming terms from our analysis.[4]

This left us with 342 unique country attributes from all the autocomplete suggestions. That such a large number of queries (228,528) reduces to such a small set of unique country attributes (342) is consistent with prior scholarship which finds that people in many different places hold similar perceptions about countries (Thornton et al. 2012). Nearly 80% of all attributes were expressed with a single word (e.g. *hot, dirty*, *peaceful*). Other country attributes were expressed in a short phrase (e.g. *good at soccer, hard to conquer, densely populated*). The longest attributes, comprising less than 0.5% of unique queries, were six words long, followed by five words (1.5%), four words (2.5%), three words (10%), and two words (8.2%).

We also performed a number of data-cleaning procedures common in the computational analysis of textual data, including removing extra spaces (two or more adjacent spaces in the text) and modest word stemming. Word stemming, which involves reducing a word to its root form (populated → populate; named → name), was necessary to ensure that words in our data set could be matched to the same root word in the sentiment data set (described below).

Table 1 illustrates the structure of the cleaned search term data set. Column 1 identifies the public from which the Google search term originated. Column 2 identifies the country being described (stimulus country), and column 3 lists the suggestions given after the phrases "Why is [Austria] so . . ." and "Why is [Venezuela] so . . . ." As illustrated in the table, terms include those about one's own country (Austria), and those emanating from one country (UK) toward another (Venezuela). Table 1 also illustrates instances of missing values in our search matrix, where English-language queries about Venezuela that emanated from the UK produced only seven of a potential 10 top search term results. Notice also that the attribute *expensive* was associated with both Austria and Venezuela, demonstrating an instance in which two quite different countries with regard to culture, geography, and economic levels can be associated with the same national characteristic in Google search queries.

TABLE 1. Structure of Google search queries about countries

Public	Stimulus	Search term
Austria	Austria	beautiful
Austria	Austria	clean
Austria	Austria	cold
Austria	Austria	expensive
Austria	Austria	fearful of nationalism and liberalism
Austria	Austria	good at recycling
Austria	Austria	happy
Austria	Austria	racist
Austria	Austria	rich
Austria	Austria	small
United Kingdom	Venezuela	bad
United Kingdom	Venezuela	broke
United Kingdom	Venezuela	corrupt
United Kingdom	Venezuela	dangerous
United Kingdom	Venezuela	expensive
United Kingdom	Venezuela	mess up
United Kingdom	Venezuela	poor
United Kingdom	Venezuela	..
United Kingdom	Venezuela	..
United Kingdom	Venezuela	..

Note: *Public* is the country from which the query originated; *stimulus* is the country named in the query.

Additional Measures

Sentiment-scoring Google search queries Our second hypothesis asks whether the positivity or negativity of Google search queries about countries systematically varies by a country's level of development based on global development indices. With the country as our primary unit of analysis, we reduced all of the search terms for each country down to a single number expressing the average sentiment embodied in terms about the country. To accomplish this, we linked our search term data set to the SenticNet lexicon, a publicly available sentiment dictionary containing sentiment scores for 50,000 positive and negative words and short phrases (Cambria and Hussain 2015).[5] As is common in the sentiment analysis of textual data, single words (unigrams) can have a score from −1 (very negative) to +1 (very positive). A phrase such as *good at sports* would receive two sentiment scores, one for *good* (+0.66) and one for *sports* (−0.04). Because *at* is not scored in the SenticNet lexicon, it receives no sentiment score. Words like *successful, great, good*, and *famous* are illustrative of search terms that have high positive sentiment, and words such as *poor, violent*, and *dirty* are terms that our scoring method identified as having a highly negative sentiment. Scores for individual terms are derived from machine learning techniques in which the scoring allocation algorithm is "trained" on a large number of text corpora with the help of human coders.

After scoring terms from Google search queries by sentiment, we computed a single sentiment score for each country from the weighted average of the sentiment scores of all terms about the country.[6] Our weighting variable was the number of times each attribute was associated with a country, which ensured that terms associated with a country by many different publics had more influence on that country's overall sentiment score than did an attribute that was only infrequently associated with it. For example, *successful* was a top-10 search attribute for China in 187 publics, whereas *stupid* was associated with China only once. Without frequency weighting, a commonly searched attribute like *successful* and an infrequently searched attribute like *stupid* would have equal influence in the calculation of China's summary sentiment score. The country-level weighted sentiment scores ranged from a low of −0.64 (Nuaru) to a high of 0.51 (United Arab Emirates), with mean of −0.12 and standard deviation of 0.24.

Scoring countries by level of development We also linked our search data to two measures of national development regularly produced by the United Nations, one continuous and the other categorical. We used the 2015 Human Development Index (HDI), which assigns a development score to each country with possible values from 0 to 1, as a continuous measure of national development. We used the 2016 UN development classification of countries as low, medium, high, or very high as a categorical measure.

Scoring countries according to visibility Our third hypothesis asks whether a country's position on development indices is related to its visibility among world publics. To measure visibility, we leverage the fact that some countries were the target of many Google search queries from many countries, and other countries were the target of a small number of queries from a few countries. In the analysis that follows, we use the relative completeness of the data for each country as a proxy for national visibility. Countries with complete data (10 search results for each of 194 searching publics) are interpreted as highly salient to world publics, while a country with few search results about it is a largely "invisible" country. While this is an admittedly imperfect measure of a country's visibility to the general public, data completeness does give us some insight into the relationship between a country's position on global development indices and the frequency with which it is the target of Google search queries.

RESULTS

Images of Development

Descriptive review of Google search queries Our first hypothesis posits that the characteristics publics associate with countries are reflective of DI. The characteristics they ascribe to nations exemplify their national and developmental perceptions. We expect that the search attributes of nations depict a world divided up into "developed" and "developing" societies, and provide general images of what a developed society looks like. Qualitatively, the country characteristics used in Google search queries tend to cluster around a small number of themes, including the economy, polity, natural environment, safety/security, demographic regime, culture/people, and national reputation. Table 2 lists terms that are illustrative of each of these thematic areas.

TABLE 2. Thematic content of Google search queries about countries

Theme	Search terms
Economy	rich, expensive, poor, cheap, high rent, wealthy, broke, GDP low, impoverished, in debt, prosperous
Polity	corrupt, stable/unstable, liberal, right-wing, left-wing, free, hard to govern, socialist, conservative, democratic
Natural environment	hot, cold, beautiful, big, small, dry, rainy, humid, windy, flat, mountainous, dusty, green, warm, cloudy, biodiverse, prone to natural disasters, icy, weather bad
Safety/security	dangerous, violent, safe, peaceful, clean, dirty, (un)healthy, water clean, scary, trashy, dangerous, dirty, ugly, violent
Demographic regime	life expectancy low/high, populated densely/sparsely populated, overpopulated, underpopulated, unpopulated, low populated, population small, population young, empty, birth rate high, death rate high, infant mortality high
Culture/people	boring, (un)happy, sexist, racist, homophobic, (ir)religious, weird, mean, nice, angry, profane, catholic, suicidal, crazy, productive, angry, annoying, dumb, smart, extreme, spiritual, anti-semitic, unique
National reputation	important, popular, awesome, special, successful, great, famous
Generalized development	good, bad, civilized, uncivilized, developed, underdeveloped, backward, barbaric/barbarous, savage, technologically advanced, behind at technology, westernized, HDI low, rank high, doing well, messed up, underrated, innovative, urbanized, advanced

Note: Search terms listed above are an illustrative, rather than an exhaustive, list of traits in each category

Economic attributes include words such as *rich, poor, broke*, and *impoverished*. Attributes associated with a country's polity include *corrupt, stable/unstable, liberal*, and *right-wing*. Other terms, such as *weak* and *powerful*, may be simultaneously reflective of perceptions of a country's government, military, or economy. Attributes associated with the natural environment include *hot, cold, rainy, humid, biodiverse*, and *prone to natural disasters*. Attributes that reflect safety, security and general well-being include *dangerous, violent, safe, peaceful*, (*un*)*healthy*, and *water clean*.

Attributes reflecting a country's demographic regime include *life expectancy low/high, populated, birth rate high*, and *infant mortality high*. Descriptions of a country's culture and people include attributes such as *boring*, (*un*)*happy, sexist, racist, homophobic*, (*ir*)*religious*, and *weird.* Characteristics that appear to reflect national identity/reputation include *important, popular, awesome, special, successful, great*, and *famous*. And finally there were also many terms that reflect a generalized understanding of development that do not clearly fit the already mentioned themes: terms such as (*un*)*civilized, developed, backward, barbaric*, and *savage*.

The content of Google search queries clearly overlaps with the developmental discourse emanating from social scientific writings, official publications of world development institutions, and the work of many international non-governmental organizations. The content

domains we identify have significant overlap with those found in prior DI research (Melegh et al. 2013, 2016). Two content domains in our search data that were not included in DI surveys were safety/security and the natural environment. To date, DI scholarship has not investigated the linkages between these two content domains and developmental discourse, though environmental narratives of development and institutional capability recently have gained greater prominence in both development theory and public policy discourse (e.g. Andrews 2013; United Nations Development Programme 2011).

Word densities of search data Figure 2 displays the 40 most searched attributes, each of which accounts for at least 0.5% of all terms in our data set. The most queried attribute, *poor*, comprised approximately 11% of all terms and was associated with 122 of 195 countries (63% of all stimulus countries). The next most frequent terms were *expensive* (associated with 42% of countries), *hot* (28%), *rich* (27%), *cheap* (21%), *dangerous* (18%), and *corrupt* (19%). In agreement with prior research (Straumann and Graham 2014), we find that the economic terms *poor*, *expensive*, *rich*, and *cheap* are among the most prevalent terms associated with countries, collectively accounting for nearly 25% of all terms. This also agrees with the work of Melegh et al. (2013, 2016), who found that respondents most frequently cited the economy as something they were thinking about when rating countries on development. The salience of *corrupt* in our results also agrees with recent DI scholarship which showed that publics associate "good governance" with development (Thornton et al. 2017). Other terms with high search prevalence included *hot, dangerous, corrupt, bad, important, popular, small*, and *cold*. As shown in Figure 2, a relatively small number of terms, many of which are steeped in DI, appear to dominate publics' imagination about countries.

Poor was strongly associated with countries that are often grouped together in the academic and policy literature as "the developing world." Notable exceptions included Eastern Europe, portions of Southern Europe, and the "emerging economies" or BRIC countries (Brazil, Russia, India, and China), each of which was also associated with *poor*. No country in which the majority of the population is of Northwest European ancestry was associated with the term *poor*; nor was South Korea or Japan. In contrast to the geographic prevalence of the trait *poor*, publics' use of *rich* was far more nuanced. *Rich* was extensively associated with North, West, and Central European countries and with European diaspora countries (i.e. Australia, Canada, New Zealand, and the United States), but it was also associated with regional economic leaders such as Chile and Argentina in South America (which are also largely Western European diaspora countries), Saudi Arabia in the Middle East, South Africa and Botswana in Africa, and China, Japan, and South Korea in Asia (Table 3).

We further decomposed the *rich–poor* distinction by coding stimulus countries into two categories (Table 3). Countries in the left column were associated with *rich* and never with *poor*. The right three columns list countries that were associated with *poor* and never with *rich*. Countries identified as "rich" by Google searchers are frequently found at the top of world development indices. Many of these countries are ranked, for example, among the world's happiest, most free, most healthy, and most prosperous, according to indices produced and disseminated by Gallup, Freedom House, the World Health Organization, and the United Nations, respectively. Countries in the right three columns are more

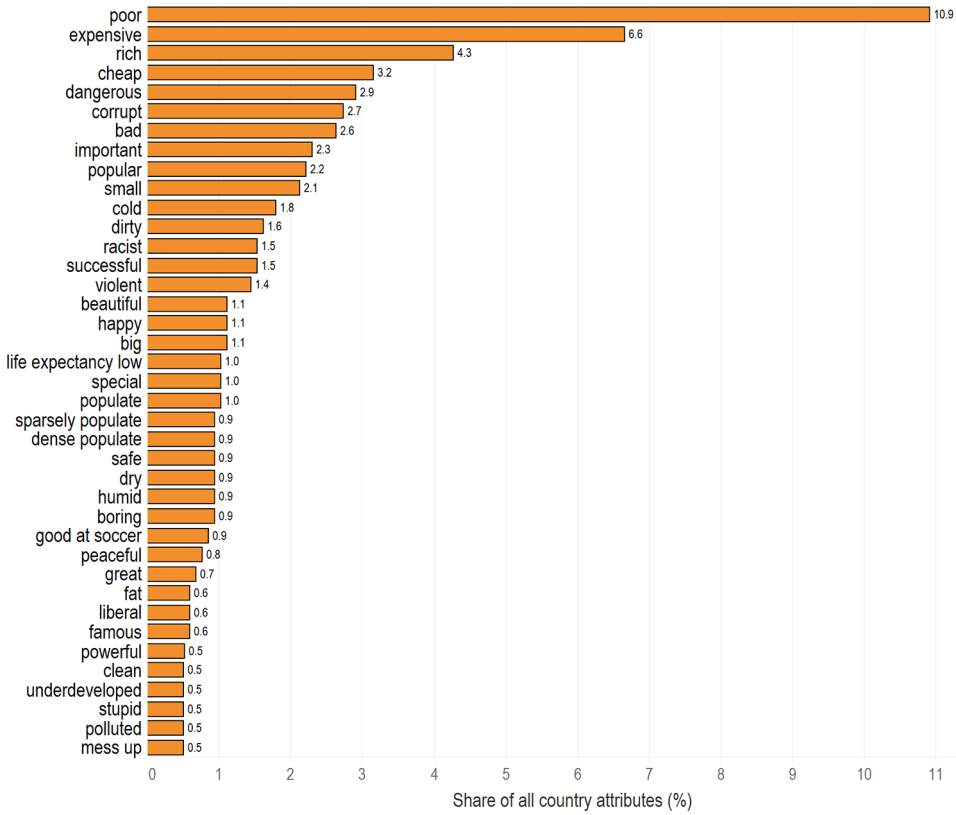

FIGURE 2. Top 40 attributes as a share of all search terms from Google search queries about countries
Note: Bar height is scaled to the share of all terms, where, for example, searches including the term *rich* comprised 4.3% of all searches in our database.

heterogeneous, though the list is dominated by countries rated low on various development indices. This suggests that the perceptions of world publics concerning the wealth and poverty of nations, as expressed in Google search queries, are deeply interconnected with the global developmental hierarchy, a finding that agrees with previous DI survey research (Binstock et al. 2013; Csánóová 2013; Dorius 2016; Swindle, Dorius and Melegh 2019; Kiss 2017; Lai and Mu 2016; Lai, Mu, and Thornton 2015; Melegh et al. 2013, 2016; Thornton and Yang 2016; Thornton et al. 2012).

National attributes by level of development To gain insights into the images of development in the minds of people who use Google to search for information about countries, we assigned a score to each search term that reflected the average HDI score of countries that were ever associated with the term. For example, the average HDI score of countries that were ever associated with *poor* was 0.66, while the average HDI score of countries ever associated with *safe* was 0.87. Table 4 lists 50 search attributes, including those with the 25 lowest and 25 highest average HDI scores, excluding very low-frequency terms, and sorted by frequency of occurrence in our data set.

TABLE 3. Rich and poor countries according to Google search queries about countries

Associated with "rich," but never "poor"	Associated with "poor," but never "rich"		
Australia	Afghanistan	Gambia	Palestine
Austria	Albania	Georgia	Papua New Guinea
Azerbaijan	Algeria	Ghana	Paraguay
Bahrain	Angola	Greece	Peru
Belgium	Armenia	Guatemala	Philippines
Brunei	Bangladesh	Guyana	Poland
Canada	Belarus	Haiti	Portugal
Chile	Belize	Honduras	Puerto Rico
Cyprus	Benin	Hungary	Republic of the Congo
Denmark	Bhutan	India	Romania
Finland	Bolivia	Indonesia	Russia
France	Bosnia and Herzegovina	Iraq	Rwanda
Germany	Brazil	Italy	Samoa
Hong Kong	Bulgaria	Ivory Coast	Senegal
Iceland	Burkina Faso	Jamaica	Serbia
Israel	Burundi	Kenya	Sierra Leone
Japan	Cambodia	Laos	Solomon Islands
Kazakhstan	Cameroon	Latvia	Somalia
Kuwait	Cape Verde	Lesotho	Spain
Liechtenstein	Central African Republic	Lithuania	Sri Lanka
Luxembourg	Chad	Macedonia	Tajikistan
Netherlands	Colombia	Madagascar	Tanzania
New Zealand	Costa Rica	Malawi	Thailand
Norway	Croatia	Mali	Togo
Qatar	Cuba	Mexico	Tonga
San Marino	Czech Republic	Moldova	Tunisia
Saudi Arabia	Dem. Rep. of the Congo	Mongolia	Turkey
Seychelles	Djibouti	Morocco	Turkmenistan
Singapore	Dominica	Mozambique	Uganda
Slovenia	Dominican Republic	Myanmar	Ukraine
Sweden	Ecuador	Namibia	Uruguay
Switzerland	Egypt	Nepal	Vanuatu
United Arab Emirates	El Salvador	Nicaragua	Venezuela
United Kingdom	Ethiopia	Niger	Vietnam
United States	Fiji	Pakistan	Zimbabwe

TABLE 4. Terms from Google search queries about countries that are most frequently associated with countries that receive low and high scores on the Human Development Index (HDI)

Search terms associated with low-HDI countries			Search terms associated with high-HDI countries		
Term	**Mean HDI**	**Occurrences**	**Term**	**Mean HDI**	**Occurrences**
poor	0.66	24,958	safe	0.87	2,141
dangerous	0.66	6,635	boring	0.88	2,138
life expectancy low	0.50	2,340	great	0.90	1,553
special	0.67	2,337	liberal	0.88	1,365
populate	0.67	2,335	clean	0.85	1,170
sparsely populate	0.66	2,150	awesome	0.87	782
underdeveloped	0.51	1,170	strict	0.87	776
mess up	0.62	1,167	suicidal	0.87	586
homophobic	0.63	779	weird	0.91	585
birth rate high	0.43	585	flat	0.93	581
hard to infect	0.52	391	healthy	0.91	393
poor rent high	0.62	391	evil	0.87	392
mean	0.48	390	prosperous	0.92	391
good at running	0.50	390	british	0.86	390
undernourished	0.51	390	atheist	0.87	390
undeveloped	0.55	390	good at winter olympics	0.94	390
economy bad	0.58	390	catholic	0.89	389
oddly shaped	0.64	390	productive	0.91	389
unhappy	0.64	390	green	0.92	389
inflation high	0.66	390	cloudy	0.92	386
isolated	0.67	390	rainy	0.94	386
crazy	0.60	389	smart	0.90	196
overpopulation	0.60	389	good at sport	0.92	196
disgusting	0.63	198	important to china	0.92	196
strong	0.40	196	good at speed skating	0.92	196

Note: *Mean HDI* is the average HDI of countries ever associated with the listed national characteristic. *Occurrences* is the frequency with which the term appeared in our data set. The lowest-frequency terms are not listed.

Characteristics associated with low-HDI countries in English-language Google search queries suggest the belief that life in those countries is one of high fertility and mortality, short life expectancy, high prevalence of disease, and endemic poverty. With an average HDI score of 0.43, *birth rate high* had among the lowest HDI score associations. *Poor*, which appeared more than 24,000 times in our data set, was the term most frequently associated with low-HDI countries. The image of underdevelopment reflects a Hobbesian world where life is "solitary, poor, nasty, brutish, and short" (Hobbes [1651] 1996). Low-HDI countries were perceived as *isolated* (Hobbes's "solitary"), *poor, disgusting*, and *dangerous*,

and where *life expectancy is low*. The data also show that words such as *underdeveloped* and *undeveloped* were frequently associated with low-HDI countries.

Among the search attributes most consistently associated with high-HDI countries, *safe, boring, great, liberal*, and *green* appeared with greatest frequency in our data set. The mental image of development, as measured by attributes associated with high-HDI countries, suggests a very different life than what is inferred from attributes about low-HDI countries. In high-HDI countries, people are perceived as excelling in sports; life is seen as *prosperous, rainy, green*, and *safe*; and people are viewed as *productive, healthy, smart, liberal*, and *awesome*.

Of course, high-HDI countries do have negative stereotypes, including being perceived as *weird, evil*, and *suicidal*, but these negative attributes are far outweighed by positive ones. Conversely, low-HDI countries are sometimes associated with positive attributes (e.g. *special* or *strong*), but again, such characteristics are far outweighed by negative ones.

Goodness and Development

Our second hypothesis concerns a positive association between the affective orientation (sentiment, emotion) of a country's attributes as expressed in Google search queries and the country's placement in global development indices. To measure public sentiment toward countries, we averaged the sentiment scores of all search terms associated with each country, weighted by term frequency, which we report in Figure 3. With few exceptions, the data

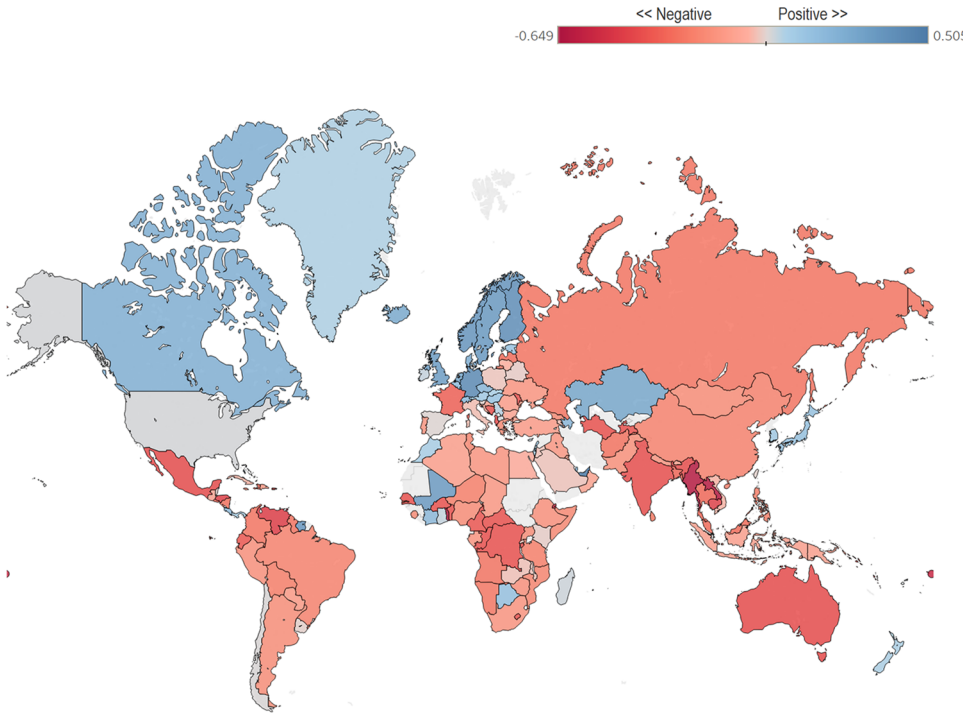

FIGURE 3. Average sentiment toward countries, according to Google search queries about countries

Note: Average sentiment refers to the weighted average of the sentiment scores of all search teams directed at a country. Countries associated with many positive search terms will have a high average sentiment, while countries associated with many negative search terms will have a negative average sentiment.

show a strong positive bias toward countries of North and Central European ancestry. Ranked by the average sentiment expressed in Google search queries about each country, we found that sentiment toward Nordic countries, including Finland (ranked 3rd), Norway (4th), Sweden (6th), Iceland (11th), and Demark (14th), is among the most favorable of any in the world. Other countries in the North and Central European country-group for which global sentiment was exceptionally favorable included Luxembourg, the Netherlands, Germany, Switzerland, Hungary, Belgium, and Canada (a Northwest European diaspora country). Highly favorable perceptions of North and Central European countries closely aligns with other DI research (Thornton et al. 2012; Swindle, Dorius and Melegh 2019) and with nation branding research (Anholt 2010).

Sentiment toward the countries of Southern Europe was more mixed, though skewed slightly negative, while sentiment toward much of Eastern Europe, including Russia, was generally negative. Russia in particular stood out for its association with a large number of negative characteristics (*corrupt, poor, racist, crazy, homophobic, cold, violent, bad, weird, backward, dangerous*, and *evil*). The attitudinal divide between Eastern and Western Europe illustrated in Figure 3 closely aligns with the well-documented east–west developmental slope of Europe (Melegh 2006; Wolff 1994).

The data also reveal several outlier countries. Australia is a rich, Western country that was associated with an unusually large number of negative characteristics (e.g. *dangerous, cold, dry, empty, hot, racists, scary, strict, weird*). Another outlier country is Kazakhstan, which was associated with just four attributes, all either neutral or positive (*big, cold, rich, sparsely populated*). The United States was another interesting country. With an HDI of 0.92 in 2015, we would expect the US to score high in terms of Google search sentiment. Instead, it was associated with a mix of positive and negative characteristics (*cold, fat, popular, populous, powerful, racists, religious, rich, stupid*, and *violent*) that gave it a middling sentiment score.

Visibility and Development

Our third hypothesis posited a positive relationship between a country's level of development and its salience in the minds of Google searchers, as measured by the percentage of missing suggestions about each country in our data set. Twenty-two countries (11%) had no English-language queries beginning with the phrase "Why is [country] so . . . ." These were typically small island nations such as Kiribati and Niue; Uzbekistan is a notable exception.[7] Another 40% of countries were associated with six or fewer unique search traits. Countries with few queries included small countries (e.g. Latvia, Moldova, Andorra, United Arab Emirates), island nations, a number of countries in Africa, and more recently independent states (e.g. Kazakhstan, Tajikistan, Czech Republic). Sixty-seven countries had complete data, or 10 suggestions about the country from each of the 194 searching publics. Nearly all countries in this group were either large or categorized as highly developed by the United Nations. Viewed globally, the least searched countries were in Africa, Central Asia, and to a lesser extent in Southeast Asia and South America (see Figure 4, where country visibility is depicted based on the percentage of 1,940 possible search results for a country that are missing).

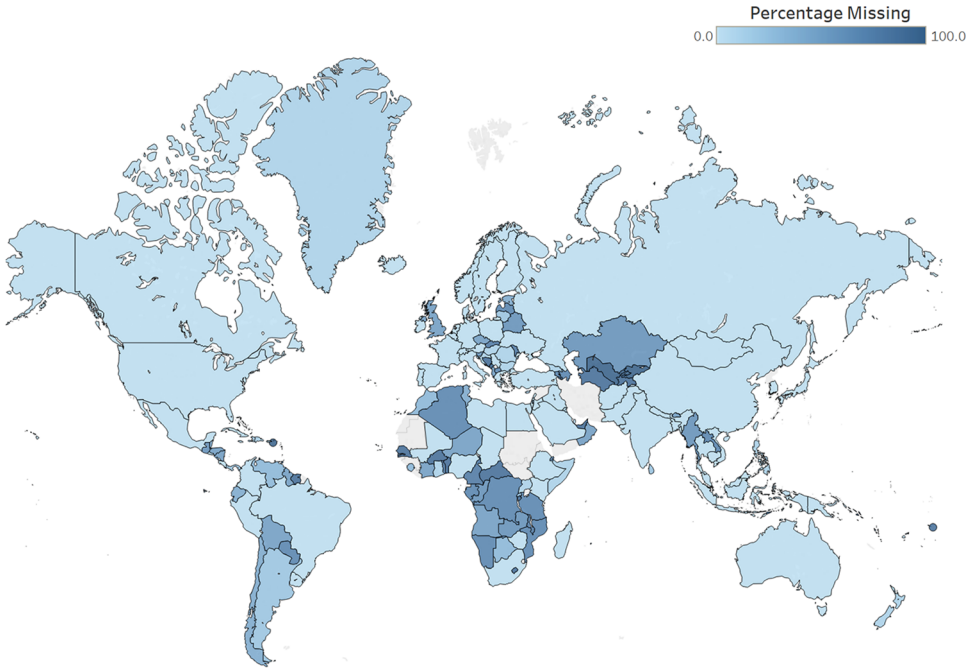

FIGURE 4. Country visibility according to percentage of missing data in Google search queries about countries

Note: Darker shading identifies low-visibility countries: those for which there were few or no search queries. Lighter shading identifies highly visible countries: those for which there were many search queries.

TABLE 5. Relationship between level of development, national visibility, and sentiment, by category of development

UN development classification*	Mean HDI	Visibility (% missing)	Mean sentiment
Low	46.8	48.0	−0.23
Medium	63.2	45.0	−0.22
High	75.3	34.1	−0.13
Very high	87.8	18.8	+0.06

Note: * Per UN Development Programme categories (2016).

In our data, the frequency of internet search queries about countries is a linear function of HDI (Table 5). Countries classified as low in development (HDI < 0.55) had an average data missingness of 48%. In contrast, countries classified as very highly developed (HDI > 0.80) had an average missingness of less than 20%. Across all countries, the correlation between frequency of searches and HDI was 0.34, indicating a positive relationship between position on development indices and visibility. Controlling for population size yielded an even larger correlation of 0.44 between frequency of searches about countries (visibility) and country HDI scores.

We detected a similar pattern when we compared sentiment scores of countries across UN developmental categories (low, medium, high, very high). The mean sentiment of terms associated with countries that the UN categorized as low in development was −0.23, meaning that on average, attributes associated with low-HDI countries have a negative connotation. Even among countries in our data set that were classified as very highly developed, the average sentiment score was only 0.06. The correlation between the sentiment of terms associated with a country and its HDI score was 0.55, while the correlation between a country's sentiment scores and its visibility was 0.36. In aggregate, the search terms directed at very highly developed countries had positive connotations, while the aggregate sentiment scores of countries in all other developmental categories were negative.

CONCLUSION

Our study used English-language Google search queries about countries to evaluate central propositions of DI scholarship. The data we analyzed suggest that when people search the internet for information about countries, their search queries are laden with developmental language. This sort of language is related to the prominence of DI schemas in the personal culture of many searchers. That we were able to detect a clear signal of DI in the aggregated queries of people in nearly 200 different countries provides further evidence of the contemporary global pervasiveness of DI thinking among world publics.

Our analysis of Google search queries provides further evidence that the countries of North and Central Europe, including European diaspora countries, enjoy a perceptual advantage over other countries, at least among English-speaking internet users. Rich Western countries are perceived as technologically advanced and where life is long and people are healthy, safe, free, happy, smart, and rich. Life in equatorial countries is perceived as the opposite: unsafe, dirty, overpopulated, poor, savage, and backward. The consistency with which "developed" country attributes were associated with high-HDI countries and "underdeveloped" country attributes with low-HDI countries reveals a global public that is attuned to the social hierarchy propagated by international development organizations. The consistency with which publics perceive a country favorably or unfavorably according to its position on development indices suggests that culture—in this case, DI—is an important source of stability in the global developmental hierarchy of nations.

We found that publics associate positive characteristics with high-HDI countries and generally negative characteristics with low-HDI countries. The data we collected project the belief that development is good and underdevelopment is bad, at least in terms of the sentiment embodied in the terms we analyzed. If we consider the sentiment scores as rough measures of the affective, or emotional, content of queries about countries, then we are inclined to state that global sentiment toward high-HDI countries is positive while sentiment toward low-HDI countries is negative, irrespective of whether queries emanate from rich or poor countries. We suspect that the association between sentiment and a country's position on global development indices has real-world consequences for how nationalities interact, both locally and globally. Nonetheless, the correlations are not as high as the correlations

found in the prior literature between survey measures of perceived levels of national development and actual HDI scores. This suggests that other subjective concepts beyond DI influence the affective content of national perceptions. We suspect that these latent global sentiments toward countries structure interactions between nationalities in ways that advantage some groups and disadvantage others. Additional research is needed to evaluate the relationship between publics' sentiment about certain nations and the status of foreign relations between countries.

Limitations

Our view is that internet search data represent a valuable source of information from which to harvest novel social and cultural insights. Nevertheless, search data are not without their limitations. First, we only collected terms that followed our search query in the English language. We do not know how different our results would be if we replicated our study in different languages. Second, we only collected search terms from Google. Data from other search engines, particularly large, non-English search engines such as Chinese-language search engine Baidu, might reveal different patterns. Third, our data collection reflects a single snapshot in time, so we cannot speak to the stability of the patterns we observed. Fourth, the overall negative sentiment expressed in the search terms in our data set calls into question whether our stem search query—"Why is [country] so . . . ?"—may have been biased toward negative responses and perhaps masked commonly ascribed positive country traits. Future research is needed to assess how search patterns vary with time, place, language, search engine, and stem search query.

Despite these limitations, we note that Google is the most used search engine in the world, and the majority of content on the Internet is in English and flows from wealthy countries, in particular the United States (Ballatore, Graham, and Sen 2017). Internet users may sometimes search in English even when it is not their native tongue, because they know from experience that there is simply far more information available in English. This is especially likely when they are searching for information about other countries, as English is increasingly known as the global lingua franca.

We conclude by noting that autocomplete suggestions, such as the ones we collected for our study, may introduce perceptions about countries that the internet searcher did not previously have or endorse and therefore could alter their prior national and developmental perceptions. Some have argued that Google and other autocompletion utilities are not just tools for *revealing* latent, negative perceptions and biases such as racism and sexism but also *disseminate* such perceptions (Baker and Potts 2013; Diakopoulos 2015; Miller and Record 2016; Noble 2018). Search algorithms such as Google's autocomplete may contribute to pathological stereotyping of social groups according to nationality by showing users the queries of prior users. This suggests that algorithmic bias is another mechanism by which old and sometimes long-discredited ways of thinking about people and places are transmitted across time, space, and cultures. Future research is needed to understand the extent to which technologies propagate developmentalist stereotypes and how such stereotypes influence intergroup interactions. ■

*Please direct correspondence to SHAWN F. DORIUS (sdorius@iastate.edu), Department of Sociology, Iowa State University, 317 East Hall, Ames, IA 50010. This research was made possible by support from the College of Agriculture and Life Sciences and the Department of Sociology, Iowa State University, a training grant (R01 HD078397) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), and NICHD research infrastructure grants (R24 HD041028 and P2CHD041028) to the Population Studies Center of the University of Michigan. We extend special thanks to the editors of this special issue, Keera Allendorf and Arland Thornton, as well as the members of the Developmental Idealism working group for their helpful comments.

REFERENCES

Allendorf, Keera, and Arland Thornton. 2015. "Caste and Choice: The Influence of Developmental Idealism on Marriage Behavior." *American Journal of Sociology* 121(1):243–87.

Andrews, Matt. 2013. *The Limits of Institutional Reform in Development: Changing Rules for Realistic Solutions*. Cambridge University Press.

Anholt, Simon. 2010. *Places: Identity, Image, and Reputation*. Basingstoke: Palgrave Macmillan.

Bail, Christopher A. 2014. "The Cultural Environment: Measuring Culture with Big Data." *Theory and Society* 43(3):465–82.

Baker, Paul, and Amanda Potts. 2013. "'Why Do White People Have Thin Lips?' Google and the Perpetuation of Stereotypes via Auto-Complete Search Forms." *Critical Discourse Studies* 10(2): 187–204.

Ballatore, Andrea, Mark Graham, and Shilad Sen. 2017. "Digital Hegemonies: The Localness of Search Engine Results." *Annals of the American Association of Geographers* 107(5):1194–1215.

Bargh, John A., and Katelyn Y. A. McKenna. 2004. "The Internet and Social Life." *Annual Review of Psychology* 55:573–90.

Binstock, Georgina, et al. 2013. "Influences on the Knowledge and Beliefs of Ordinary People about Developmental Hierarchies." *International Journal of Comparative Sociology* 54(4):325–44.

Binstock, Georgina, and Arland Thornton. 2007. "Knowledge and Use of Developmental Thinking about Societies and Families among Teenagers in Argentina." *Demografia* 50(5):75–104.

Buchanan, William. 1951. "Stereotypes and Tensions as Revealed by the UNESCO International Poll." *International Social Science Bulletin* 3(3):515–28.

Cambria, Erik, and Amir Hussain. 2015. *Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis*. Springer International.

Chae, David H., et al. 2015. "Association between an Internet-Based Measure of Area Racism and Black Mortality." *PLOS One* 10(4):e0122963.

Csánóová, Sabina. 2013. "Rank-Ordering Modernity: Perceptions of Global Hierarchies and Development in Hungary." Department of Sociology and Social Anthropology, Central European University, Budapest, Hungary.

D'Andrade, Roy G. 2005. "Some Methods for Studying Cultural Cognitive Structures." Pp. 83–104 in *Finding Culture in Talk: A Collection of Methods*, edited by N. Quinn. New York: Palgrave Macmillan.

Davis, Julie Hirschfeld, Sheryl Gay Stolberg, and Thomas Kaplan. 2018. "Trump Alarms Lawmakers with Disparaging Words for Haiti and Africa." *New York Times*, January 11.

de Saussure, Ferdinand. 1964. *Course in General Linguistics*. London: Owen.

Diakopoulos, Nicholas. 2015. "Algorithmic Accountability." *Digital Journalism* 3(3):398–415.

DiGrazia, Joseph. 2017. "The Social Determinants of Conspiratorial Ideation." *Socius* 3:2378023116689791.

DiMaggio, Paul, Eszter Hargittai, W. Russell Neuman, and John P. Robinson. 2001. "Social Implications of the Internet." *Annual Review of Sociology* 27(1):307–36.

Dorius, Shawn F. 2016. "Chinese and World Cultural Models of Developmental Hierarchy." *Chinese Journal of Sociology* 2(4):577–608.

Fatehkia, Masoomali, Ridhi Kashyap, and Ingmar Weber. 2018. "Using Facebook Ad Data to Track the Global Digital Gender Gap." *World Development* 107:189–209.

Fiske, Susan T. 2011. *Envy Up, Scorn Down: How Status Divides Us*. New York: Russell Sage Foundation.

Franzosi, Roberto. 2010. *Quantitative Narrative Analysis*. Los Angeles, CA: SAGE.

Franzosi, Roberto, Gianluca De Fazio, and Stefania Vicari. 2012. "Ways of Measuring Agency: An Application of Quantitative Narrative Analysis to Lynchings in Georgia (1875-1930)." *Sociological Methodology* 42(1):1–42.

Garcia, David, et al. 2018. "Analyzing Gender Inequality through Large-Scale Facebook Advertising Data." *Proceedings of the National Academy of Sciences* 115(27):6958–63.

Ginsberg, Jeremy, et al. 2009. "Detecting Influenza Epidemics Using Search Engine Query Data." *Nature* 457(7232):1012–14.

Google. 2017. "Search Using Autocomplete." Retrieved August 1, 2017 (https://support.google.com/websearch/answer/106230?hl=en).

Graham, Mark, and Anasuya Sengupta. 2017. "We're All Connected Now, So Why Is the Internet so White and Western?" *The Guardian*. Retrieved April 17, 2018 (http://www.theguardian.com/commentisfree/2017/oct/05/internet-white-western-google-wikipedia-skewed).

Gross, Neil, and Marcus Mann. 2017. "Is There a 'Ferguson Effect?' Google Searches, Concern about Police Violence, and Crime in U.S. Cities, 2014–2016." *Socius* 3:2378023117703122.

Guillén, Mauro F., and Sandra L. Suárez. 2005. "Explaining the Global Digital Divide: Economic, Political and Sociological Drivers of Cross-National Internet Use." *Social Forces* 84(2):681–708.

Hobbes, Thomas. [1651] 1996. *Leviathan*, edited by J. C. A. Gaskin. Oxford: Oxford University Press.

Hwang, Hokyu. 2006. "Planning Development: Globalization and the Shifting Locus of Planning." In *Globalization and Organization; World Society and Organizational Change*, edited by G. S. Drori, J. W. Meyer, and H. Hwang. New York: Oxford University Press.

Ignatow, Gabe. 2016. "Theoretical Foundations for Digital Text Analysis." *Journal for the Theory of Social Behaviour* 46(1):104–20.

Jena, Anupam B., Pinar Karaca-Mandic, Lesley Weaver, and Seth A. Seabury. 2013. "Predicting New Diagnoses of HIV Infection Using Internet Search Engine Data." *Clinical Infectious Diseases* 56(9):1352–53.

Johnson-Hanks, Jennifer A., Christine A. Bachrach, S. Phillip Morgan, Hans-Peter Kohler, and Lynette Hoelter. 2011. *Understanding Family Change and Variation: Toward a Theory of Conjunctural Action*. London: Springer.

Kiss, Tamás. 2017. "Escaping the 'Balkanizing' Gaze? Perceptions of Global and Internal Developmental Hierarchies in Romania." *East European Politics and Societies* 31(3):565–95.

Klineberg, Otto. 1951. "The Scientific Study of National Stereotypes." *International Social Science Bulletin* 3(3):505–14.

Lai, Qing, and Zheng Mu. 2016. "Universal, Yet Local: The Religious Factor in Chinese Muslims' Perception of World Developmental Hierarchy." *Chinese Journal of Sociology* 2(4):524–46.

Lai, Qing, Zheng Mu, and Arland Thornton. 2015. "World Culture at the Individual Level: Individual Conformity to the Schema of National Development." 4th Annual Conference of the Sociology of Development Section, Brown University.

Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. "The Parable of Google Flu: Traps in Big Data Analysis." *Science* 343(6176):1203–05.

Lazer, David, et al. 2009. "Computational Social Science." *Science* 323(5915):721–23.

Liu, James H., et al. 2012. "Cross-Cultural Dimensions of Meaning in the Evaluation of Events in World History? Perceptions of Historical Calamities and Progress in Cross-Cultural Data From Thirty Societies." *Journal of Cross-Cultural Psychology* 43(2):251–72.

Lizardo, Omar. 2017. "Improving Cultural Analysis: Considering Personal Culture in Its Declarative and Nondeclarative Modes." *American Sociological Review* 82(1):88–115.

McCarthy, Michael J. 2010. "Internet Monitoring of Suicide Risk in the Population." *Journal of Affective Disorders* 122(3):277–79.

McLaren, Nick, and Rachana Shanbhogue. 2011. "Using Internet Search Data as Economic Indicators." *Bank of England Quarterly Bulletin* 51(2):134–40.

Melegh, Attila. 2006. *On the East-West Slope: Globalization, Nationalism, Racism and Discourses on Central and Eastern Europe*. New York: Central European University Press.

Melegh, Attila, Tamás Kiss, Sabina Csánóová, Linda Young-DeMarco, and Arland Thornton. 2016. "The Perception of Global Hierarchies: South-Eastern European Patterns in Comparative Perspectives." *Chinese Journal of Sociology* 2(4):497–523.

Melegh, Attila, Arland Thornton, Dimiter Philipov, and Linda Young-DeMarco. 2013. "Perceptions of Societal Developmental Hierarchies in Europe and Beyond: A Bulgarian Perspective." *European Sociological Review* 29(3):603–15.

Mellon, Jonathan. 2014. "Internet Search Data and Issue Salience: The Properties of Google Trends as a Measure of Issue Salience." *Journal of Elections, Public Opinion, and Parties* 24(1):45–72.

Mellon, Jonathan. 2017. "Making Inferences about Elections and Public Opinion Using Incidentally Collected Data." Pp. 522–33 in *The Routledge Handbook of Elections, Voting Behavior and Public Opinion*, edited by J. Fisher, E. Fieldhouse, M. N. Franklin, and M. Cantijoch. New York: Routledge.

Mellon, Jonathan, and Christopher Prosser. 2017. "Twitter and Facebook Are Not Representative of the General Population: Political Attitudes and Demographics of British Social Media Users." *Research & Politics* 4(3):2053168017720008.

Merry, Sally Engle, Kevin Davis, and Benedict Kingsbury, eds. 2015. *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law*. Cambridge, MA: Cambridge University Press.

Meyer, J. W., and M. Hannan. 1979. *National Development and the World System*. Chicago, IL: University of Chicago Press.

Meyer, John W., John Boli, George M. Thomas, and Francisco O. Ramirez. 1997. "World Society and the Nation-State." *American Journal of Sociology* 103(1):144–81.

Miller, Boaz, and Isaac Record. 2016. "Responsible Epistemic Technologies: A Social-Epistemological Analysis of Autocompleted Web Search." *New Media & Society* 19(12):1945–63.

Nisbet, Robert A. 1969. *Social Change and History: Aspects of the Western Theory of Development*. Oxford: Oxford University Press.

Nisbet, Robert A. 1986. *The Making of Modern Society*. Brighton: Wheatsheaf.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.

Patterson, Orlando. 2014. "Making Sense of Culture." *Annual Review of Sociology* 40(1):1–30.

Quinn, Naomi. 2005. "How to Reconstruct Schemas People Share, from What They Say." Pp. 35–81 in *Finding Culture in Talk: A Collection of Methods*. New York: Palgrave Macmillan.

Rangil, Teresa Tomás. 2013. "Citizen, Academic, Expert, or International Worker? Juggling with Identities at UNESCO's Social Science Department, 1946–1955." *Science in Context* 26(1):61–91.

Rath, Badri Narayan. 2016. "Does the Digital Divide across Countries Lead to Convergence? New International Evidence." *Economic Modelling* 58:75–82.

Rivera, Lauren A. 2008. "Managing 'Spoiled' National Identity: War, Tourism, and Memory in Croatia." *American Sociological Review* 73(4):613–34.

Salganik, Matthew J. 2017. *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press.

Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

Spears, Russell, Tom Postmes, Martin Lea, and Anka Wolbert. 2002. "When Are Net Effects Gross Products? The Power of Influence and the Influence of Power in Computer-Mediated Communication." *Journal of Social Issues* 58(1):91–107.

Swindle, Jeffrey, Shawn F. Dorius, and Attila Melegh. 2019. "The Mental Map of National Hierarchy in Europe." SocArXiv. https://osf.io/pd4rb/

Steinfeldt, Jesse A., et al. 2010. "Racism in the Electronic Age: Role of Online Forums in Expressing Racial Attitudes about American Indians." *Cultural Diversity & Ethnic Minority Psychology* 16(3): 362–71.

Stephens-Davidowitz, Seth. 2014. "The Cost of Racial Animus on a Black Candidate: Evidence Using Google Search Data." *Journal of Public Economics* 118:26–40.

Stephens-Davidowitz, Seth. 2017. *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us about Who We Really Are*. New York: HarperCollins.

Stier, Sebastian. 2017. "Internet Diffusion and Regime Type: Temporal Patterns in Technology Adoption." *Telecommunications Policy* 41(1):25–34.

Straumann, Ralph, and Mark Graham. 2014. "The World through the Eyes of a Search Algorithm." *Oxford Internet Institute*. Retrieved September 6, 2017 (https://geography.oii.ox.ac.uk/the-world-through-the-eyes-of-a-search-algorithm/).

Straumann, Ralph K., and Mark Graham. 2016. "Who Isn't Online? Mapping the 'Archipelago of Disconnection'." *Regional Studies, Regional Science* 3(1):96–98.

Strauss, Claudia. 2005. "Analyzing Discourse for Cultural Complexity." Pp. 203–42 in *Finding Culture in Talk: A Collection of Methods*, edited by N. Quinn. New York: Palgrave Macmillan.

Swindle, Jeffrey. 2019. "Categorizing the World: Developmental Classification of Societies in English Books, 1700-2000." Unpublished Paper.

Taboada, Maite, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. "Lexicon-Based Methods for Sentiment Analysis." *Computational Linguistics* 37(2):267–307.

Thornton, Arland. 2001. "The Developmental Paradigm, Reading History Sideways, and Family Change." *Demography* 38(4):449–65.

Thornton, Arland. 2005. *Reading History Sideways: The Fallacy and Enduring Impact of the Developmental Paradigm on Family Life*. Chicago, IL: University of Chicago Press.

Thornton, Arland, Shawn F. Dorius, and Jeffrey Swindle. 2015. "Developmental Idealism: The Cultural Foundations of World Development Programs." *Sociology of Development* 1(2):277–320.

Thornton, Arland, Shawn Dorius, Jeffrey Swindle, Linda Young-DeMarco, and Mansoor Moaddel. 2017. "Middle Eastern Beliefs about the Causal Linkages of Development to Freedom, Democracy, and Human Rights." *Sociology of Development* 3(1):70–94.

Thornton, Arland, and Li-shou Yang. 2016. "Perceptions of Developmental Hierarchies in Taiwan: Conceptual, Substantive, and Methodological Insights." *Chinese Journal of Sociology* 2(4):547–76.

Thornton, Arland, et al. 2012. "Knowledge and Beliefs about National Development and Developmental Hierarchies: The Viewpoints of Ordinary People in Thirteen Countries." *Social Science Research* 41(5):1053–68.

Todorova, Maria, ed. 1997. *Imagining the Balkans*. New York: Oxford University Press.

Towns, Ann E., and Bahar Rumelili. 2017. "Taking the Pressure: Unpacking the Relation between Norms, Social Hierarchies, and Social Pressures on States." *European Journal of International Relations* 23(4):756–79.

United Nations Development Programme. 2011. *Human Development Report 2011: Sustainability and Equity—A Better Future for All*. New York: Palgrave Macmillan.

Vaisey, Stephen. 2009. "Motivation and Justification: A Dual-Process Model of Culture in Action." *American Journal of Sociology* 114(6):1675–1715.

Vaisey, Stephen. 2014. "Is Interviewing Compatible with the Dual-Process Model of Culture?" *American Journal of Cultural Sociology* 2(1):150–58.

Wilson, Kumanan, and John S. Brownstein. 2009. "Early Detection of Disease Outbreaks Using the Internet." *CMAJ* 180(8):829–31.

Wolff, Larry. 1994. *Inventing Eastern Europe: The Map of Civilization on the Mind of Enlightenment*. Palo Alto, CA: Stanford University Press.

Yang, Albert C., Norden E. Huang, Chung-Kang Peng, and Shih-Jen Tsai. 2010. "Do Seasons Have an Influence on the Incidence of Depression? The Use of an Internet Search Engine Query Data as a Proxy of Human Affect." *PLOS One* 5(10):e13728.

NOTES

1. It is likely that some (though very few) searchers do not believe this themselves, but rather believe that it is commonly believed by others and are curious to know common explanations for this belief. Even in such cases, this search is evidence of their awareness of this common public belief.

2. Aggregation of search data to the level of countries necessarily masks within-country variation in searches by more granular locales, such as states, cities, or regions.

3. Our method to identify stimulus countries was to restrict our search to official English-language country names. Some people might use the search terms England, Great Britain, United Kingdom, and UK as synonyms for England. Likewise, searches for United States, America, US, and USA, could all be used to identify the United States of America. Our purpose in restricting our data extraction to official country names was to minimize measurement error (e.g. "America" might refer to the Americas rather than the United States), though it also means that our results should be considered a lower bound for the total searches for countries.

4. Because Google does not reveal the details of its search algorithms, we cannot say why our data collection gathered a small number of search results that deviated from our target phrase, "Why is [country] so . . . ."

5. See sentic.net for details on sentiment analysis. Various methods have been deployed to develop sentiment dictionaries, but the most familiar to social scientists will be ones in which researchers present survey respondents with a list of words or phrases and ask them to evaluate words along one or more dimensions (usually polarity). The aggregate ratings of a word or phrase by survey respondents are used to develop a summary measure of its semantic orientation.

6. Google searches without matching sentiment scores in the SenticNet lexicon include *affected by brexit, badass, behind, biodiverse, british, mountainous, multicultural, often called a quagmire, perverted, targeted by isis, uncivilized, underpopulated, underrated, unexplored*, and *unpopulated*.

7. Countries excluded from our analysis due to no search data included Antigua and Barbuda, Ascension Island, British Virgin Islands, Cocos (Keeling) Islands, Cook Islands, Federated States of Micronesia, Guadeloupe, Kiribati, Kyrgyzstan, Montenegro, Montserrat, Niue, Norfolk Island, Pitcairn Islands, Helena, Saint Vincent and the Grenadines, Sao Tome and Principe, Timor-Leste, Tokelau, Tristan da Cunha, United States Virgin Islands, and Uzbekistan.