Culture and Queries: Google search data as a reflection of national culture

Author: B.J.B. Bredenoord
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT, Search engine data from Google Trends is frequently used for cross-country comparison, but it is unknown if Google search behavior reflects cultural traits. This study takes an explorative approach and aims to explore the idea that Google search engine data and national culture can be linked. For several English- and Spanish speaking countries, this paper compares the relative frequency of culture expressing keywords in Google search queries, and link this to national culture using the Hofstede and GLOBE models. Using another sample of 42 countries, we analyze search volume data about searches for year numbers in the future and present in Arabic numerals (e.g. '2015', '2016'). For the English- and Spanish speaking countries that are analyzed, significant relationships are found between national culture and the relative occurrence of culture expressing words for the dimensions power distance, individualism vs. collectivism, and long-term vs. short-term orientation. The analysis of the other sample of 42 countries shows that cultures with a high future orientation tend to search more often for year numbers in the future. The idea that Google search queries reflect culture may enable new forms of cross-cultural research. This knowledge may encourage the development of new methods of data collection for cross cultural comparison using Google search data.

Supervisors: Dr. A.B.J.M. Wijnhoven, Dr. C. Amrit

Keywords

National culture, Hofstede, GLOBE, Internet, Google Trends, Search queries, Search engine, Language.

1. INTRODUCTION

Massive amounts of data are collected these days as a result of society's interaction with IT systems. An example is search engine query data from Google. This enabled scientist to study flu trends (Ginsberg et al., 2009), changes in financial markets (Preis, Reith, Stanley, 2010), and even recreational fishing harvests (Carter, Crosson, Liese, 2015). Google search engine query data is freely available via Google Trends, a Google service that publishes the relative (not absolute) Google search volume for a keyword for a country or region. This service can be used to compare the popularity of keywords between different periods and regions. The fact that data can be gathered for virtually almost every country allows researchers to use these data for cross-country comparison (e.g. Troelstra, Bosdriesz, De Boer, Kunst, 2014; Vaughan, Romero-Frías 2014; Walasek, Brown, 2015; Benthaus, Neufeld, 2015; Qu, Manrique, Johnson, Restrepo, Johnson, 2016).

There are many different reasons why countries may differ on the type of information they search for via Google. One important reason may be cultural differences, because culture is very deeply rooted in people's minds (Hofstede, 2001). Different 'dimensions' of culture have been linked to many features of online behavior, values, and preferences. For example our preference for either horizontal (mouth-oriented) or vertical (eye-oriented) emoticons (Park, Baek, Cha, 2014), and website structure (Cyr, Head, 2013). To the best of our knowledge, culture has not been linked to search behavior however. This paper aims to show that a country's Google search behavior is a reflection of its national culture.

To find differences between countries, we can focus on the relative popularity of words. Previous research has shown that word use may reflect cultural values, because word use shows what is prioritized (Greenfeld, 2016). The relationship between word use and culture has been studied before in texts, such as books (Greenfeld, 2016; Twenge et al., 2012 A; Twenge et al., 2012 B), or the way bilingual people translate small stories (Na & Choi, 2009). The difference with this research is however that we will study word use in Google search queries, whereas most other research focuses on written texts.

Hofstede (2001) and House, et al. (2002) conceptualized culture into several measurable 'dimensions' for which a score (e.g. on a 1 to 100 scale) can be assigned to a country. This allows comparing multiple countries. Franke & Richney (2010) argue that it is necessary to compare at least 7-10 countries to make credible international generalizations. There are two languages which are spoken in more than 10 countries that were included in Hofstede's analysis. These languages are English and Spanish. This research will therefore compare countries that speak one of these languages. We will not only study the relative popularity of certain words, but also use year numbers in Arabic numerals (e.g. '2015', '2016'). The latter approach will be used for the cultural dimension future orientation.

This paper presents the idea that Google search queries reflect culture, which enables new forms of cross-cultural research. This knowledge may encourage the development of new methods of data collection for cross cultural comparison using Google search data. This would offer many benefits, since Google Trends is available for free, and data collection about entire countries can be done quickly, conveniently, and without the use of surveys.

The purpose of this article is to answer the question: How does a country's relative search volume of pronouns, year numbers, and other culture expressing keywords in Google search queries reflect its national culture?

This paper will be mainly explorative and will be structured as follows; we will first explain national culture, its dimensions, and identify words of which the use may indicate either a high or low score on a cultural dimension. We will then analyze the relative search volume of these words using Google Trends to compare search behavior of 11 English speaking and 12 Spanish speaking countries with its national culture. We will analyze these two groups of countries by themselves, and also explore the option of combining the results in one larger group. Using another sample of 42 countries, we will compare a country's future orientation with the relative search volume for year numbers in the future. After discussing the results, we will mention several limitations of the research and end with a conclusion.

2. LITERATURE BACKGROUND

In this overview of the literature we will first explain national culture, and present two models of national culture (Hofstede and GLOBE). Secondly, we will identify words which relative frequency may reflect a country's score on a cultural dimension. We will use at least one word of which the use implies a low country score on the dimension, and one word which use implies a high score on that dimension. The relative use of these words in Google searches will be compared with the actual country scores. We will name these words '*culture expressing words'*.

2.1 National culture

Hofstede (1994) defined national culture as "the collective programming of the mind which distinguishes the members of one category of people from another" (p. 1). A country's culture is common to most but not all citizens (Hofstede, 2001), and should not be seen as an accurate description of personality traits of the statistically average culture member (McCrea, Terracciano, Realo, Allik, 2008). Several models that describe culture have been introduced. In this paper we will use two important ones, Hofstede, and GLOBE. Both models measure a country's culture as a result of the combination of different cultural 'dimensions' (i.e. measurable aspect of culture).

The first model, the Hofstede model, is the most widely used model for comparing cultures (De Mooij & Hofstede, 2010). Hofstede originally surveyed employees at IBM. The results of these surveys were used to evaluate countries on a 0 to 100 scale for every cultural dimension. The model originally consisted of four dimensions (*'power distance', 'individualism vs. collectivism', 'masculinity vs. femininity', 'uncertainty avoidance'*). The number of dimensions was expanded with *'long-term orientation'* in 2001, and *'indulgence vs. restraint'* in 2010. In this analysis we will not use this last dimension because it can be seen as complementary to Hofstede's long-term orientation dimension (Hofstede, 2011). The dimensions *'masculinity vs. femininity',* and *'uncertainty avoidance'* were also excluded because we were not able to identify words that can be said to reflect culture on a strong theoretical basis based on the available literature.

The second model that will be used is the more recent GLOBE model (House et al., 2004). The GLOBE model is largely based on the Hofstede model. Both models can therefore be seen as complementary. Three GLOBE dimensions are intended to measure the same construct as three Hofstede dimensions, and have the same name in both models. The Hofstede dimension long-term orientation was changed slightly into GLOBE's future orientation measure. Even though Hofstede's long-term dimension, and GLOBE's future orientation dimension describe a region's orientation towards time, both measures are not intended to measure the same construct. The difference will be explained later. Because many of the GLOBE dimensions are intended to measure the same as the Hofstede dimensions (House

et al., 2002), we will exclude most of them and only use the dimension 'future orientation'.

We will now describe three Hofstede dimensions, and one GLOBE dimension into further detail. For every dimension, we will also include some culture expressing words for which we argue that its relative frequency of use by individuals may characterize that individual's score on a cultural dimension.

2.2 Power distance

Power distance refers to the extent to which the unequal distribution of power is accepted by those not in power (Hofstede, 2001 p. 82). This inequality consists in all cultures, but the degree to which it is tolerated differs per culture. People in high power distance countries accept a certain degree of hierarchy, whereas people in low power distance countries prefer power equality (Hofstede, 2001 p. 80). Ahmed, Mouratidis, & Preston (2009) note that many company websites in high power distance countries reflect this cultural aspect in the form of for example hierarchy information, authority figures, honorific titles, and official certification. According to these authors, hierarchy is not only accepted in these countries, but people also value information about hierarchy.

We argue that if people accept a certain degree of unequal power distribution, they may be preoccupied with hierarchy, which may be represented by the relative use of the keyword 'hierarchy' (or 'hierarchies'), whereas people that prefer a more equal distribution of power might be more interested in the search term 'equality'[1].

2.3 Individualism vs. collectivism

People in individualist countries look primarily after themselves, and their own interest. Collectivist cultures instead assume that any person belongs to one or more 'in-groups'. These groups protect the interest of the members, but expect loyalty in return (Hofstede & De Mooi, 2010). In individualist societies, self-actualization and freedom are important. People's identity is based upon the individual. In collectivist societies more emphasis is placed on harmony, and group consensus, and a person's identity is based upon the social system of which the person belongs (Hofstede, 2001).

2.3.1 Strategy one: pronouns

Hofstede (2001 p. 227) already notes that collectivist countries are more 'we' conscious, whereas individualistic societies are more 'I' conscious. Gardner, Gabriel & Lee (1999); Na & Choi (2009); Twenge, Campbell, Gentile (2012 A); Uz (2014); Hamamura & Xu (2015), and Yu et al. (2016) used the relative use of pronouns in texts to study individualism and collectivism. Research showed that people may use natural language in the form of full sentences and question type search queries instead of keywords (Aula, Khan, Guan, 2010). In such cases, people may use pronouns in search queries. The relative use of these pronouns may reflect individualism.

2.3.2 Strategy two: 'get', and 'give'

Not-pronoun words can also be used for studying the relationship between word use and individualism. Greenfeld (2009), and Hofstede (2001, p. 209 + 227) show clear links between the concepts of Gemeinschaft (community) vs. Gesellschaft (society) on the one hand, and collectivism vs. individualism on the other. These concepts were first described by German sociologist Tönies in the late 19th century. He described Gemeinschaft societies as poorer, more rural, less concerned with material schooling, and more concerned with welfare of other people than Gesellschaft societies. Maybe more

importantly, Gemeinschaft societies are related to more collectivist values, whereas Gesellschaft societies showed individualistic values (Greenfeld, 2009). Even though the concepts of Gemeinschaft and Gesellschaft are not intended to reflect differences between countries, but rather a urban society with a rural society, they are often linked to national culture. This link has been mentioned theoretically by Hofstede (2001 p. 227) and Greenfeld (2009), and empirically by Christenson (1984). Still, there is limited recent empirical support for a clear relationship between the two sets of concepts. Greenfeld (2016) related the difference between contributing to the welfare of other people and obtaining something for oneself was related to the use of 'give', relative to the use of 'get'.

2.3.3 Strategy three: other words by Twenge et al. (2012 B)

Twenge, Campbell & Gentile (2012 B) looked at other words that describe either individualism or collectivism. Their study resulted in a list with 20 individualistic words and 20 collectivist words. They constructed this list by asking 53 persons to come up with 5 individualistic and 5 collectivist words, and 55 persons to rate these words on individualism and collectivism. These are also the words that we will use for comparing search queries in different countries. Examples of words that are found to be individualistic are: 'individual', 'unique', 'independent'. Examples of collectivist words are: 'community', 'team', 'group'. Twenge et al. also include some variations of these words (e.g. 'individual', 'individually').

2.4 Long-term orientation

Hofstede's long-term vs. short-term orientation dimension was previously called Confucian dynamism (Hofstede, 2001). This dimension was introduced by Hofstede in 2001 and was suggested by Eastern scholars. The Western minds of the creators of the Hofstede surveys had different perception of the importance of some values than Eastern minds. This dimension may therefore be somewhat strange to Westerners. Even though the questions were developed to study dynamics in values taught by Chinese philosopher Confucius, it turned out that the dimension was generalizable to countries that never heard from him (Hofstede, 2001 p. 355).

The dimension involves the extent to which a country exhibits a future-oriented perspective rather than a conventional historic or short-term point of view. Values included in long-term orientation are perseverance, thrift, adaptation, ordering relationships by status, and having a sense of shame. The opposite pole, short-term orientation, includes personal steadiness, stability, respect for tradition, social obligations, and preservation of 'face' (Hofstede, 2001 p. 354). Differences in thinking styles and cognitive processes also play a major role (Hofstede, 2001 p. 362)

Two versions of country scores on the long-term orientation dimension scale are available. One is based on the so called 'Chinese Value Survey', and the other is based on the 'World Value Survey' (Gerecke, House, 2013). The country scores differ slightly between both versions. We will use the Hofstede scores that are based on results derived from the World Value Survey, because these scores are more recent and are available for more countries (Gerecke, House, 2013). Unfortunately, not all countries have been studied by Hofstede on the long-term orientation dimension. Therefore, for not all countries is a score available. For the English-speaking countries, there is no long-term orientation score available for Jamaica. For the Spanish speaking countries, there is no score available for Costa Rica,

---

[1] Originally, we also intended to include 'equalities' after gathering data it turned out however that search volume was so

low that no data was obtainable for some countries. We therefore excluded this term from the analysis.

Ecuador, Guatemala, and Panama. These countries are therefore excluded for the long-term orientation dimension. Just like for the dimension individualism vs. collectivism, we use multiple strategies to estimate a country's level of long-term orientation, one based on moral values and one based on the relative popularity of '*why*' and '*what*'.

2.4.1 Strategy one: Moral values

For short- long term orientation, Hofstede operationalizes his dimensions according to persistence, perseverance, status, thrift on the long-term side, and a focus on steadiness, stability, 'face', and tradition on the other. As culture expressing words we will therefore use '*persistence*', '*perseverance*', '*shame*', (high long-term orientation) and '*tradition*', '*dignity*', '*reputation*' (two last as alternative for 'face') (short-term orientation) as culture expressing words.

2.4.2 Strategy two: 'why' vs. 'what'

As said above, one of the differences that is associated with short-term or long-term orientation is a preference for one of two cognitive thinking styles. These two thinking styles can be either analytical (also known as specific), or holistic (also known as synthetic). A preference for a holistic thinking style is related to long-term orientation, whereas a preference for an analytical thinking style is associated with short-term orientation (Hofstede, 2001, p. 364).

Nisbett, Choi, Peng, & Norenzayan (2001) see a holistic thinking style as: "an orientation to the context or field as a whole, including attention to relationships between a focal object, and (…) a preference for explaining and predicting events on the basis of such relationships". On the other side of the continuum we find the analytical thinking style, which is involves 'detachment of the object from its context, a tendency to focus on attributes of the object to assign it to categories' (Nisbett et al., 2001). Analytical thinking in this situation should not be confused with analytical skills necessary to be successful in mathematics or science (Hofstede, 2001 p. 364).

Western, short-term thinkers can be seen as analytic, focusing on an object. Eastern, long-term thinkers can be seen as holistic, focusing on the relationship between objects. An example of the difference between the two thinking styles can be given by Masuda & Nisbett (2001). They showed American (short-term orientation, analytical thinking) and Japanese (long-term orientation, holistic thinking) individuals a short video about an underwater scene. They asked the participants to describe what they saw. Many Americans described what they saw in the forefront; fish. Many Japanese said that they saw an aquarium. They spoke more about what was going on in the background and were more likely to mention interdependencies between the different elements in the water. This and other research showed that Eastern thinkers are much more likely to see events with reference to its context than Western thinkers. Western thinkers instead, emphasize the main elements itself.

Searching via Google is per definition a sign for a perceived need of information. As we have seen however, people may differ in the whether they emphasize either the object itself, or the relationships and interdependencies between objects. For question type searches, this emphasis may be reflected by the type of words they use in their search query. According to Morgan, Moni & Jobling (2009), the word '*why*' relates to 'exploration of understanding including cause and effect, feelings, attitudes and behaviors'. We argue that people who stress the importance on the relationship between different entities may search more often for '*why*', because this word implicitly recognizes that a state of event is the result of some other event, which suggests that a relationship between objects is assumed. The use of the word '*what*' was placed on a different

level of thinking than '*why*' by Morgan et al. (2009). We argue that people who put emphasis on the object itself will be more likely to search for '*what*'. On first sight, the use of the search term 'how' might also seem to express an implicit assumption of a relationship between elements. However, 'how' can also be used in combination with many different words, for example 'much' or 'many' (e.g. 'how much', 'how many'). Such question-based searches do not imply that a relationship between elements is assumed.

2.5 Future orientation

GLOBE's measure of future orientation can be seen as related to Hofstede's long-term orientation, but there are important differences. GLOBE's measure is defined as 'the extent to which members of a society or an organization believe that their current actions will influence their future, focus on investment in their future, believe that they will have a future that matters, believe in planning for developing their future, and look far into the future for assessing the effects of their current actions.' (House et al., 2004) Put very simply, GLOBE's measure of future orientation can be seen as the extent to which a country shows interest in the future. Hofstede's long-term orientation dimension instead, looks more at values associated with a focus on the long-term.

Just like for the other GLOBE dimensions, there are two versions of the future orientation dimension available, 'practices' (which reflect 'as is'), and 'values' (which reflects 'as should be'). GLOBE's 'values' measures do not reflect the direct behavior of people, values' focuses on the believe of how people 'should' behave, rather than how they really behave. In this paper we will use GLOBE's future orientation 'practices' measure, since we are interested in how societies really behave, not how people think people in their society should behave.

Preis, Moat, Stanley & Bishop (2012) studied search query data from Google to study the drivers that affect the orientation of internet users towards the future between different countries. They used data from Google Trends from 2010 in which they tried to find out how many searches were conducted for the terms '2009' and '2011'. They use year numbers in Arabic numerals to overcome language problems, so they can compare many countries. They found that there is a correlation between a country's economic success (GDP per capita) and orientation towards year numbers in the future. In their short paper, they suggested two possible explanations, (1) a focus on the future supports economic success, and (2) a country's GDP per capita affects the type of information sought online possibly due to economic influences on available Internet infrastructure.

Preis et al. did not link their measure of orientation towards the future to national culture, even though different perceptions of time orientation is one of the most studied aspects of national culture (e.g. Hofstede, 2001; House, Javidan, Hanges, Dorfman, 2002). In this article, we will take an approach which is similar to the one used by Preis et al., but we will link it to national culture. Since future orientation is correlated with GDP per capita, it is possible that the relationship between GDP per capita and future orientation found by Preis et al. was confounded by GDP per capita. We will therefore also look at the relationship between future orientation and the relative search volume of year numbers in the future while controlling for GDP per capita. (i.e. statistically excluding the effect that GDP per capita has on both variables).

3. METHOD

We will compare the relative search volume of culture expressing words on Google with a country's national culture. We will withdraw data from Google trends. Google trends data is

available from 2004 onwards. Google Trends is frequently used to track changes over time, but a different approach will be used here. We will use the average search volume of keywords using data from searches from January 2004, until January 2016. Because culture evolves slowly, we do not expect major cultural change within this time frame. Google Trends provides relative search volumes rather than absolute search volumes. It can therefore be used to compare to multiple contrasting phenomena, by comparing two sets of keywords, but not the absolute use of one keyword. Therefore, we need to balance high and low scoring keywords on one dimension.

This means that we will compare the average relative search frequency of two (sets of) opposing culture expressing keywords among countries. Based on the literature, several culture expressing words have been identified for various cultural dimensions. The frequent use of one of these keywords will be related to a high score on a cultural dimension, whereas a frequent use of the other will express a low score on the dimension. For every country, the relative search frequency of the first word will be divided by the relative search frequency of the second word. This results in a ratio. Our hypothesis is that this ratio reflects culture, and will therefore be correlated with the country's score on that cultural dimension[2]. To check this hypothesis, the correlation between this ratio and the country's score on the relevant dimension will be calculated. Kendall's tau will be used as measure of correlation because a quick examination showed that we cannot be entirely confident that all the used variables follow a bivariate normal distribution. The future orientation dimension is the only dimension for which we will not use keywords, but year numbers in Arabic numerals (e.g. '2015', '2016').

We will analyze countries of which either English or Spanish is an official language, and is also common as first language. Only those countries studied by Hofstede (2001) will be included. The English speaking countries that will be included are Australia, Canada, Ireland, Jamaica, Malta, New Zealand, Singapore, South Africa, Trinidad and Tobago, United Kingdom, and United States. English is also an official language in for example India and Hong Kong, but number of people that uses it as a first language is low (The Government of the Hong Kong Special Administrative region's census and statistics department, 2011; Census of India, 2003). Included Spanish speaking countries are Argentina, Chile, Colombia, Costa Rica, Ecuador, Guatemala, Mexico, Panama, Peru, Spain, Uruguay, and Venezuela.

To be able to compare among different languages, different culture indicating words have to be translated. If for a given word, the spelling of British English and American English differs, both versions of that word will be included in the analysis. To compare Spanish speaking countries, these English words also had to be translated to Spanish. For each English word, the three most common Spanish translations were selected from dictionaries by the authors. Native speakers of Spanish that all followed formal university-level education in English ranked those Spanish words on how well they represented the meaning of the English word. They ranked these words on a 1 to 7 scale ('*does not compare to the word at all*', to '*can be seen as complete equivalent*'). These translators were three native speakers of Spanish from Spain (Catalonia), Colombia, and Ecuador.

3.1 Power distance
We define a country's Google based power distance score as the search volume of '*hierarchy*' + '*hierarchies*', divided by the search volume of '*equality*'. Originally, we also intended to include '*equalities*', after gathering data it turned out that search volume was so low that no data was obtainable for some countries. We therefore excluded this term from the analysis. For Spanish speaking countries we will use the keywords '*jerarquía*' + '*jerarquías*' relative to '*igualdad*'.

3.2 Individualism vs. collectivism
There are different types of words which may express individualism or collectivism. We will use three 'strategies' to estimate a country's level of individualism. We will use pronouns, the words '*give*' and '*get*', as well as some other words previously used by Twenge et al. (2012 B).

3.2.1 Pronouns
For every English speaking country that is included in Hofstede's analysis, we will calculate the ratio of the relative search volume for a singular (individualistic) pronoun (e.g. '*I*') over its plural (collectivistic) alternative (in this example: '*we*'). This is done for three groups of pronouns, '*I*' vs. '*we*', '*my*' vs. '*our*', and '*myself*' vs. '*ourselves*'. These three ratios were averaged to find the individualist over collectivist pronoun ratio for each country. The ratios of '*mine*' vs. '*ours*', and '*me*' vs. '*us*' were excluded. Because '*mine*' and '*us*' can also have different meanings, which might influence the results. '*Mine*' can mean an excavation in the earth for extracting coal or other minerals, and '*us*' can be used to refer to the United States ('US').

Unfortunately, pronouns are often dropped in sentences in Spanish. This does not mean that the Spanish language does not have pronouns, but it does mean that there are no pronouns in many Spanish sentences. Instead of using pronouns, the form of the verb is changed to indicate the subject in Spanish sentences. One might therefore try to find a significant relationship using the ratio of first person singular and the first person plural version of common verbs. We explored this option, for the verb '*ser*' ('to be')*.* A significant relationship was found between individualism and the occurrence of the first person singular version relative to the occurrence of the first person plural version of this word. For some other verbs, the search volume of the first person plural version of some of these verbs was so low that no reliable results could be obtained, so this strategy was not explored further.

3.2.2 'get' vs. 'give'
Greenfeld (2016) found that the relative use of '*get*' and '*give*' may express individualism or collectivism. We will therefore calculate a score of individualism based on the search volume of '*get*' divided by the search volume of '*give*'. For Spanish speaking countries we initially wanted to use the score of the present forms of '*dar*' relative to '*obtener*'. These words could unfortunately not be used since the search volume score for the forms of '*dar*' was found to be extremely high relative to '*obtener*', which made that the scores for '*obtener*' were very unreliable. As an alternative we therefore used the present forms of '*conseguir*' (get) vs. '*ofrecer*' (offer). For Spanish speaking countries we will calculate a score of individualism as the search volume for '*consigo*' + '*consigues*' + '*consigue*' + '*conseguimos*' + '*conseguís*' + '*consiguen*' divided by the relative score of '*ofrezco*' + '*ofreces*' + '*ofrece*' + '*ofrecemos*' + '*ofrecéis*' + '*ofrecen*'.

---

[2] Country scores on the dimensions were retrieved from Hofstede's website: http://www.geerthofstede.nl/dimension-data-matrix

3.2.3 'Self' vs. 'each other'

This strategy can be seen as a variation on the pronoun-based strategy. In fact, this strategy includes so called indefinite pronouns, which replace nouns without specifying which nouns they replace. These words were separated from the pronouns-based strategy. The reason for this is that there is no one direct opposite for one of these words, but this is possible for the words in the pronoun-based strategy (e.g. the opposite of *'self'* could be both *'each other'*, as well as *'everyone'*. The opposite of *'I'* is always *'we'*.) The words that are compared are *'self'* + *'oneself'* relative to the use of *'each other'* + *'everyone'*. Google's search volume of '*one another*' was so low that for most countries no results were presented, so it was excluded. For Spanish, we used *'yo'* + *'uno mismo'*, relative to *'El uno al otro'* + *'unos y otros'* + *'todo el mundo'* + *'todos'*.

3.2.4 Other individualistic and collectivist words by Twenge et al. (2012 B)

Twenge, Campbell & Gentile (2012 B) mentioned 40 words that were found to represent either individualism or collectivism in American books. We will calculate a score of individualism by dividing the search volume score of all the individualistic words combined by the volume score of the collectivistic words combined.

This means that we will divide the relative search frequency of the words *'independent'* + *'individual'* + *'individually'* + *'unique'* + *'uniqueness'* + *'self'* + *'independence'* + *'oneself'* + *'soloist'* + *'identity'* + *'personalized'* + *'solo'* + *'solitary'* + *'personalize'* + *'loner'* + *'standout'* + *'single'* + *'personal'* + *'sole'* + *'singularity'* by the relative search frequency of the words *'communal'* + *'community'* + *'commune'* + *'unity'* + *'communitarian'* + *'united'* + *'teamwork'* + *'team'* + *'collective'* + *'village'* + *'tribe'* + *'collectivization'* + *'group'* + *'collectivism'* + *'everyone'* + *'family'* + *'share'* + *'socialism'* + *'tribal'* + *'union'*.

For the Spanish speaking countries we will use *'independiente* + *individual* + *'individualmente'* + *'único'* + *'unicidad'* + *'yo'* + *'independencia'* + *'uno mismo'* + *'solista'* + *'identidad'* + *'personalizado'* + *'solo'* + *'solitario'* + *'personalizar'* + *'solitario'* + *'destacar'* + *'soltero'* + *'personal'* + *'único'* + *'singularidad'* , devided by the search volume of *'comunal'* + *'comunidad'* + *'comuna'* + *'unidad'* + *'comunitaria'* + *'unido'* + *'trabajo en equipo'* + *'equipo'* + *'colectivo'* + *'pueblo'* + *'tribu'* + *'colectivización'* + *'grupo'* + *'colectivismo'* + *'todo el mundo'* + *'familia'* + *'compartir'* + *'socialismo'* + *'tribal'* + *'Unión'*.

3.3 Long-term orientation

3.3.1 Moral values

A score for long-term orientation will be calculated by dividing the search frequency of the words '*persistence', 'perseverance', 'shame'*, divided by the search frequency of the words '*tradition, dignity, reputation'*. The last two words will be used as an alternative for *'face'*. For the Spanish speaking countries we will use the search volume of *'persistencia'* + *'perseverancia'* + *'vergüenza',* devided by the search volume of *'tradicion'* + *'tradición'* + *'dignidad'* + *'reputación'*.

3.3.2 'why' vs. 'what'

A second score of long-term orientation will be based on the difference between holistic vs. analytic thinking style. To calculate this ratio, we will divide the search frequency of '*why'* by the search frequency of '*what'*. For the Spanish speaking countries, we will use *'qué'*, and *'por qué'*.

3.4 Combining scores of Spanish and English speaking countries

Even though there are enough countries in the analysis in order to make reliable judgements for the included countries, an increase of the number of countries that were analyzed may improve the generalizability of the results to other countries. Therefore, we will not just analyze the two sets of countries by themselves alone, but also analyze the combined results by analyzing both the Spanish and English speaking countries at the same time. We excluded Spanish pronouns from the analysis because pronouns are frequently dropped in Spanish natural sentences. As alternative we will therefore analyze the use of pronouns in English speaking countries while including the languages German, French and Dutch. We included these three languages because, unlike many other languages, these do not have the pronoun drop possibility in natural sentences. The included countries (or regions) are Germany, Austria, Switzerland (only German speaking), France, Quebec (Canada), Wallonia (Belgium), The Netherlands, and Flanders (Belgium). A limitation of including these countries is that these countries are all Western. Still, the level of individualism differed noticeably from a moderate 55 (Austria), to high 80 (The Netherlands)

3.5 Future orientation

GLOBE's future orientation dimension measures can be seen as the level to which a nation is concerned with the future. Countries with a high score on this dimension may therefore relatively search more often for future year numbers than countries with a low score on this dimension. Because of the use of year numbers in Arabic numerals, language issues are not relevant and a much larger number of countries can be compared at the same time. Preis et al. (2012) calculated a so-called 'future-orientation ratio' for 45 countries by dividing the relative search volume for the next year number (in this example '*2011*') by the relative search volume for the previous year number ('*2009*' in this example. We will take a similar approach, but instead of dividing the search volume for every 'next year number' by the volume of 'previous year number', we will divide the search volume of 'next year number' by the search volume of 'present year number'. We will do this because the concept of future orientation as described by House et al. (2002) can be seen as the degree to which societies are engaged in future-oriented behavior relative to present-oriented behavior. We will calculate this ratio for three years (2013, 2014, and 2015). We will average these results which leads to a 'online future orientation ratio'. This means we will calculate a measure of a country's future orientation as following:

$$\text{Online Future Orientation ratio} = \frac{\sum_{n=2013}^{2015} \frac{\text{search volume score year } n+1}{\text{search volume score year } n}}{3}$$

4. RESULTS

Results are presented in the table below. Significant relationships are **bold**. Data were gathered for eleven English-speaking, and twelve Spanish speaking countries. If the search volume for a

Table 1. Correlations of relative search volume of culture expressing words with Hofstede's cultural dimensions

	Countries					
	English-speaking countries			Spanish-speaking countries		
Dimension	r	$p*$	N	r	$p*$	N
Power distance	.333	.180	10[A]	**.592**	**.028**	**9 [B]**
Individualism vs. collectivism						
Strategy						
Pronouns	**.600**	**.010**	**11**	N.A.	N.A.	N.A.
'get' vs. 'give' ratio	.345	.139	11	.394	.075	12
'self' vs. 'each other' ratio	.147	.532	11	-.412	.063	12
Other words by Twenge et al. (2012 B)	-.418	.073	11	-.182	.411	12
Long-term vs. short-term orientation						
Strategy						
'persistence' etc. vs. 'tradition' etc.	-.222	.404	9[A]	**.571**	**.048**	**8**
'why' vs. 'what' ratio	**.629**	**.012**	**10**	.400	.170	8

NOTE: Significant correlations ($p < .05$) are printed **bold**. For long-term orientation, less countries have been studied by Hofstede, therefore Jamaica, Costa Rica, Ecuador, Guatemala, and Panama are missing.. [a] = excluding Malta due to a lack of data. [b] = excluding Malta and Trinidad due to a lack of data. [B] = excluding Costa Rica, Panama, and Uruguay due to a lack of data.

* = Kendall's tau correlation coefficient, two-tailed

certain keyword is too low, no data is obtainable via Google trends. Therefore, for a few countries not all data was gathered for all dimensions. This was the case for Malta (for the dimensions *power distance,* and the pronoun-based strategy for *individualism vs. collectivism*), and Costa Rica, Panama, and Uruguay (for the dimension *power distance*)

4.1 Power distance

A significant relationship was found between the relative search volume of words that may express power distance and Hofstede's measure of power distance. This relationship was found for Spanish speaking countries (Kendall's tau correlation coefficient $r = .592$, N = 9, $p = .028$), but not for English speaking countries (Kendall's tau correlation coefficient $r = .333$, N = 10, $p = .180$).

4.2 Collectivism vs. individualism

4.2.1 Pronouns
In all English-speaking countries, singular (I-oriented) pronouns were used more often than plural (we-oriented) pronouns. Still, the ratio between the use of both forms differed considerably per country. People in the United Kingdom (individualist) search for I-oriented pronouns about 14 times more often than we-oriented pronouns, compared to about 8 time more for Singaporeans (collectivist). There is a significant relationship between individualist/ collectivist pronoun ratio and Hofstede's measure for individualism. (Kendall's tau correlation coefficient $r = .600$, N = 11, $p = .010$).

4.2.2 'get' vs. 'give'
For all English speaking countries, 'get' occurred more frequently than 'give'. Jamaica (collectivist) searched for 'get' four times more frequently then 'give'. The United Kingdom searched about nine times more frequently for 'get'. Still, not enough data was available to say that this relationship was significant (Kendall's tau correlation coefficient $r = .345$, N = 11, $p = .139$). The same can be said for Spanish speaking countries (Kendall's tau correlation coefficient $r = .394$, N = 12, $p = .075$). The forms of 'get' occur more frequently than 'give' in less collectivistic countries like Argentina, whereas forms of 'give' occurred more frequently than forms of 'get' in countries that were more collectivist, like Ecuador and Panama

4.2.3 'self' + 'oneself', 'each other' + 'one another' + 'everyone'
In all English speaking countries, the 'self' and 'oneself' occurred more frequently than 'each other' and 'everyone'. The degree to which these individualistic words occurred more frequently differed however from about 3 times in Trinidad (collectivist), to 9 in South Africa (individualist). Overall, the relationship between 'self' vs. 'each other' and individualism was not significant (Kendall's tau correlation coefficient $r = .147$, N = 11, $p = .532$). A negative, but insignificant relationship was found for Spanish versions of the ratio between 'self' over 'each other' (Kendall's tau correlation coefficient $r = -.412$, N = 12, $p = .063$).

4.2.4 Other individualistic and collectivist words by Twenge et al. (2012 B)
No relationship was found between the relative word use of individualistic and collectivist words with Hofstede's index for individualism. This was the case for both English-speaking countries (Kendall's tau correlation coefficient $r = .-418$, N = 11, $p = .073$), as well as for Spanish-speaking countries. (Kendall's tau correlation coefficient $r = .-182$, N = 12, $p = .411$)

4.3 Long-term orientation

4.3.1 Moral values
A significant relationship was found for words that may express long-term orientation and Hofstede's index of long-term orientation for Spanish speaking countries (Kendall's tau correlation coefficient $r = .571$, N = 8, $p = .048$). The same cannot be said for English-speaking countries (Kendall's tau correlation coefficient $r = .-222$, N = 9, $p = .404$).

4.3.2 'why' vs. 'what'
A relationship was found between the ratio of 'why' vs. 'what' and Hofstede's index of long-term orientation for English speaking countries (Kendall's tau correlation coefficient $r = .629$, N = 10, $p = .012$). No significant correlation was found however for the Spanish version of this measure (Kendall's tau correlation coefficient $r = .400$, N = 8, $p = .170$).

4.4 Combining scores of Spanish and English speaking countries

We do not just analyze the two sets of countries by themselves alone, but also look at the combined results by analyzing both the Spanish and English speaking countries at the same time. A potential problem that arises when combining two sets of samples is that the samples are no longer entirely independent. As a result, the countries are usually clustered together by language and the resulting correlation coefficient cannot be seen as completely valid. An example can be found below. This graph represents a country's level of individualism and the relative occurrence of forms of '*get*' relative to '*give*' (Kendall's tau correlation coefficient .602 p < .001 N = 23). For two cases, the clustering of the two languages was much less clear: for the, '*why*' vs. '*what*' ratio (Kendall's tau correlation coefficient .497 p = .004 N = 18), and the '*self*' vs. '*each other*' ratio (Kendall's tau correlation coefficient .350 p < .020 N = 23). Even though this clustering of countries among language is less problematic for these data, we still can question the independence of samples.
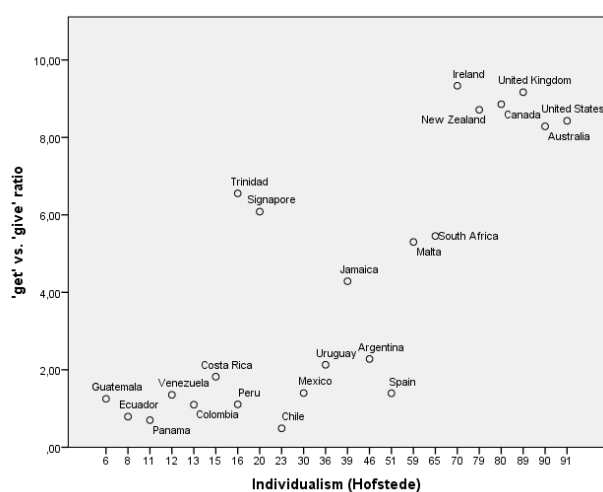

Figure 1. Individualism and 'get' vs. 'give' ratio for Spanish and English speaking countries

For analyzing the relationship between individualism and the relative use of pronouns, we included four languages which makes the likelihood that results can be generalized across other languages much larger than when only including Spanish and English. The inclusion of these languages decreased the strength of the relationship between relative pronoun use and individualism slightly, but the relationship was still significant (Kendall's tau correlation coefficient $r = .322$, N = 20, $p = .041$).

4.5 Future orientation

As hypothesized, the newly created future orientation measure correlated significantly with GLOBE's practices measure of future orientation (Kendall's tau correlation coefficient $r = .405$, N = 42, $p < .001$). It is suggested by Preis et al. (2012) that countries with a high GDP per capita may search more often for year numbers in the future because a country's GDP per capita may affect the type of information sought online, possibly due to economic influences on available Internet infrastructure. But even after controlling for GDP per capita, the relationship between GLOBE's index of future orientation and the relative number of searches for year numbers in the future was still significant (Pearson correlation coefficient $r = .411$, N = 42, $p = .008$).

5. DISCUSSION AND LIMITATIONS
5.1 Discussion of results

Even though we could not find a relationship between search query data and culture for every cultural dimension for every country, the overall results indicate that online search behavior is a reflection of a country's national culture. Combining just two languages (Spanish and English) turned out to be problematic because the languages clustered together, which indicates that the independence of samples cannot be assumed. Possibly, this is because even the best translation for an English word in a given language is not necessarily an exact equivalent of that English word in terms of day-to-day use. As a possible way of dealing with this problem, future research choose more simple words for which the same meaning in different languages can be assumed, for example pronouns.

Twenge et al. (2012 B) were able to identify a significant relationship between individualism and word use for American books over the period 1960-2008 (a period for which an increase in individualism was expected). With data from Google searches however, no clear relationship was found for the same words. A possible explanation is that the words used in books are different from language used in search queries. Another explanation why there was no clear relationship found between the words suggested by Twenge et al. and individualism may be because there is no underlying theory behind the selection of word. Instead, Twenge et al. asked a group of untrained people to come up with individualistic and collectivist words, and another group to rate those words on individualism and collectivism. Because these words are not based on theory, it may be that they only do not (or only partly) represent Hofstede's conceptualization of individualism and collectivism. A third explanation might be that the relatively, some of these words are much more often used in search queries than other words which biases the results because just a few keywords make up most of the search volume.

5.2 Limitations

Even though the results support the hypothesis that online searches reflect national culture, several potential problems with measurement were identified from the literature. These potential measurement problems may lead to wrong results. These potential problems are: unrepresentativeness of Google users (Mellon, 2011), external validity (Campbell and Stanley, 1963 p. 5), content validity, and reliability (Carmines and Zeller, 1980). We will reflect upon the importance of these potential threats.

5.2.1 Unrepresentativeness of Google Users

Internet users may differ demographically from a country's general population in age and socio-economic status. The population on Google is self-selecting and not necessarily an accurate reflection of a country's entire population. It is likely that the people that use Google are younger and somewhat wealthier than non-users. Because of this, Mellon (2011) notes this potentially creates a bias for using internet data as a reflection of a country's entire population.

Even though internet searchers themselves may not be fully representative of a country's entire population, this does not necessarily mean that their search behavior is unrepresentative of the entire population. Cultural values may be so widespread and strong among citizens of a county that the behavior of internet users is comparable to those of the general population. According to Hofstede (2001 p.23), it is common to measure culture among smaller groups, as long as the groups that are compared share a common characteristic (e.g. comparing Swedish nurses with Spanish nurses). Following this argumentation, comparing internet users within different countries should not be

problematic, as long as we assume that a Google user in country A can be compared to one in country B.

Currently, there is no literature available about how a country's culture is reflected by its Google searches. In politics however, Mellon (2011) used Google Trends to study dynamics in issue salience (i.e. what people consider the most important societal problem) in the US. He found large similarities between results of representative surveys and Google search data. He concluded that, at least in the US, data from Google Trends may well reflect what is prioritized among the entire population. Still, there is currently no foolproof method of testing whether the this generalization can be made for countries with a lower internet penetration as well.

5.2.2 External validity

According to Campbell & Stanley (1963 p. 21), external validity is about generalizability: "whether a causal relationship holds over variations in persons, settings, treatments, and outcomes". In this research we mainly looked at countries that speak either Spanish or English. We can question if these results can also be generalized across countries that do not speak one of these languages. According to Franke & Richney (2010) one prerequisite is that we need to compare at least 7-10 countries to make credible international generalizations. This condition was satisfied for both sets countries. By combining both countries into one analysis, we found that countries might cluster together with other countries speaking the same language, which suggests that language differences makes comparing multiple countries hard. This language problem is not an issue for the dimension future orientation, since we compared year numbers in the future rather than words. The fact that we also found a significant relationship between culture and Google search behavior suggests that we can make this generalization at least for the future orientation dimension.

5.2.3 Content validity

Content validity is defined as "the extent to which an empirical measurement reflects a specific domain of content" (Carmines and Zeller, 1980). Mellon (2011) proposes a qualitative assessment of content validity for Google Trends data. Even though this assessment was originally used for assessing issue salience, we can use the assessment as well to validate our search data. This method is based on the idea that people have multiple motivations to use certain keywords, and that the use of a keyword may not always reflect the construct what the researcher intends to measure. Mellon (2013) mentioned an example: a researcher studying the labor market might be interested in the search volume for 'jobs'. The relative search volume of 'jobs' may however be biased by people that search for former Apple CEO 'Steve Jobs'.

Mellon's method of assessing content validity consists of checking the 50 so called 'top searches' that include the search term. These top searches are the terms that are most frequently searched in combination with the term of which content validity is to be checked. It is possible that some of these 'top searches' are unrelated to the term that we are interested in. Mellon (2011) argues that we can question a term's content validity if too many of the top searches are unrelated to the term under interest. For all dimensions for which we found significant relationships between Google search data and culture, we checked the top searches for their content validity.

We analyzed the 'top 50 related searches'. For example, for the search term '*I*' one of the most frequently used 'top searches' term is '*i phone*', a common misspelling of Apple's popular iPhone. It is possible to exclude such an unrelated term from the results by using a '-' sign. After excluding the term 'phone' from the list however, no substantial change in search volume was found, suggesting that, based on this method, there is no reason to question the content validity of the term '*I*'.

Deleting specific words form the term has its disadvantages. Excluding these words may lead to the perceived idea that we have grip on how we define the query for which we withdraw data. In fact, because only the 50 top searches are included, which is only a small part of all queries, we can never know for sure how well the search term relates to the concept under interest. There is another disadvantage of excluding certain words from the query. It is possible that the search volume for the term that is excluded is very low in countries with less search data. In such a case, instead of neglecting the exclusion of this term, Google does not provide any data at all and the country has to be excluded completely.

This was the case for the dimension power distance. For the term '*hierarchy*', one of the top searches was '*maslow*', the creator of a psychological theory that is often referred to as 'Maslow's hierarchy of needs'. The term '*maslow*' can be seen as totally unrelated to the concept of power distance. Search frequency for the combination of '*hierarchy*' and '*maslow*' turned out to be so low for three of the eleven English-speaking countries that no data for those countries could be presented by Google Trends. After excluding these three countries from the analysis, the relationship between the '*hierarchy*' vs. '*equality*' ratio turned out not to be significant anymore.

Based on this, we can say that excluding some keywords from the search query is more helpful for some countries than others. We can however, use the 50 top related searches to give a subjective judgement about to what extent the search volume of a term is biased by unrelated searches. We must however take into account that we do not know how well this top 50 is a representation of all search queries.

5.2.4 Reliability

Reliability is defined as the extent to which an test a measurement procedure yields the same results on repeated trials (Carmines and Zeller, 1980). Culture only evolves slowly, but does not change from week to week. Small short-term deviations for culture expressing search terms via Google Trends can therefore be seen as random error, and thus a potential threat to the reliability of data. To check reliability, we analyze the weekly deviation of search volume for the 'I' vs. 'we'-ratio for every English speaking country between January 2004 and January 2016. We will do this mainly as an example to reflect upon the potential problem of unreliability of data, rather than an in-depth reliability analysis for every country. Because we gather data on a country level, reliability may differ per country. Eventually, the goal is to show that errors in weekly points of data become smaller if the number of internet users in a country increases.

We observed a small trend in the plots that represent the weekly observations for many countries. At first sight this may seem undesirable because culture is expected to change only very slowly and Hofstede did not update his country scores for more than a decade. Still, this does not have to be problematic and may be explained in two ways. First, culture may indeed change slightly over time, but Hofstede (2001, p. 60) notes that many indicators of culture move into the same direction for many countries. This means that even though the absolute value of a cultural indicator of a country may change, this does not mean that this change is also observable relatively to other countries. Second, Google Trends only provides relative search frequency data, no absolute data. This means that an increase (or decrease) in relative search frequency for a keyword does not necessarily indicate an increase (or decrease) in absolute number, or a decline in interest of searches for that keyword (Choi, 2012).

Because of the reasons described above, we have some underlying trend that we can explain. We define an error as the absolute deviation of the weekly ratio from the trend line. The trend line represents the average score of the weekly 'I' vs. 'we'-ratio over a two-year period. An example of how these weekly observations differ from the trend line can be seen below for US 'I' vs. 'we'-ratio.
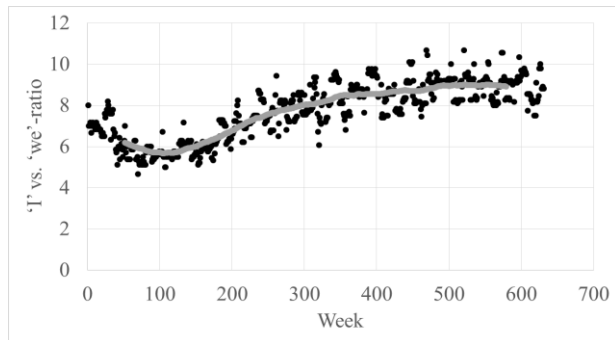

Figure 2. US weekly 'I' vs. 'we'-ratio between 2004 and 2016 and underlying trend line (grey)

The graph above shows that the weekly observations of the 'I' vs. 'we'-ratio scatter quite well around the trend line, even though there is some variation. The weekly variation of the 'I' vs. 'we'-ratio differs considerably per country, which can be seen in the table below. The variance of the error was significantly higher for those countries in which the absolute number of internet users is lower (Kendall's tau -.673 p=.004). This figure may indicate that Google Trends search results become more reliable if the number of internet users (and therefore potential Google users) increases.

Table 2. No. internet users and variance of error per country

Country	No. internet users[3]	Variance of error
United States	286,942,000	.381
United Kingdom	54,027,000	.611
Canada	29,660,000	.442
South Africa	23,766,000	1.300
Australia	18,478,000	.660
Singapore	3,986,000	.293
Ireland	3,737,000	1.402
New Zealand	3,613,000	.958
Jamaica	1,100,000	3.812
Trinidad	782,000	4.081
Malta	283,000	4.680

6. CONCLUSION

This paper started with the question how (and if) a country's relative search volume of pronouns, year numbers, and other culture expressing keywords in Google search queries reflect its national culture. Overall results confirm this hypothesis. Collectivist countries are more '*we*' and '*give*' oriented than individualistic countries in their Google search queries. Cultures that are said to be future oriented search relatively more often for year numbers in the future than countries that are less future

---

[3] Calculated using population data from the US Census Bureau (2013), and internet penetration data from the International Telecommunication Union (2013)

oriented, even after controlling for GDP. Analytic, short-term oriented thinking countries were found to search more frequently for '*what*', whereas holistic, long-term thinking countries search more often for '*why*'. Still, some potential measurement problems were identified, related to unrepresentativeness of Google users, external validity, content validity, and reliability.

Future research may try to find combinations of words which relative use may reflect a country's score on cultural dimensions for those dimensions for which we were not able to find keywords which use may explain culture, notably uncertainty avoidance, and masculinity vs. femininity. Future research may also focus on finding ways to compare countries with different languages, and if results can be generalized across other languages and countries as well. Future research may also focus on creating models to predict a country's or region's culture based on Google search data. These models may even be used to track changes within culture or regional differences. At this moment, Google Trends only provides data for searches since 2004. Previous research that studied changes in language in books usually takes a longer time span, usually several decades (e.g. Twenge et al. 2012 A; Twenge et al. 2012 B; Hamamura & Xu, 2015; Greenfeld, 2016). This means that in the (very) long term, it may be possible to use search engine data to track changes in culture.

7. ACKNOWLEDGEMENTS

8. REFERENCES

The World Telecommunication/ICT Indicators Database. (2013). Retrieved from: http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx

Ahmed, T., Mouratidis, H., & Preston, D. (2009). Website Design Guidelines: High Power Distance and High Context Culture. International Journal of Cyber Society and Education, 2(1), 47-60.

Aula, A., Khan, R. M., & Guan, Z. (2010). How does search behavior change as search becomes more difficult? Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Benthaus, J., & Neufeld, F. (2015). Responding to Customer Demand: Investigating Customer Agility of Financial Institutions.

Bureau, U. S. C. (2013). International Data Base Country Rankings. Retrieved from: http://www.census.gov/population/international/data/idb/rank.php

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for generalized causal inference: Houghton Mifflin.

Carmines, E. G., & Zeller, R. A. (1979). Reliability and validity assessment (Vol. 17): Sage publications.

Carter, D. W., Crosson, S., & Liese, C. (2015). Nowcasting Intraseasonal Recreational Fishing Harvest with Internet Search Volume. PloS one, 10(9), e0137752.

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. Economic Record, 88(s1), 2-9.

Christenson, J. A. (1984). Gemeinschaft and gesellschaft: Testing the spatial and communal hypotheses. Social Forces, 63(1), 160-168.

Cyr, D., & Head, M. (2013). Website design in an international context: The role of gender in masculine versus feminine oriented countries. Computers in Human Behavior, 29(4), 1358-1367.

De Mooij, M., & Hofstede, G. (2010). The Hofstede model: Applications to global branding and advertising strategy and research. International Journal of Advertising, 29(1), 85-110.

department, T. G. o. t. H. K. S. A. r. s. c. a. s. (2011). Population (1) Aged 5 and Over by Place of Birth, Usual Language and Ability to Speak Other Language/Dialect, 2011

Fischer, R. (2006). Congruence and functions of personal and cultural values: Do my values reflect my culture's values? Personality and Social Psychology Bulletin, 32(11), 1419-1431.

Franke, G. R., & Richey, R. G. (2010). Improving generalizations from multi-country comparisons in international business research. Journal of international business studies, 41(8), 1275-1293.

Gardner, W. L., Gabriel, S., & Lee, A. Y. (1999). "I" value freedom, but "we" value relationships: Self-construal priming mirrors cultural differences in judgment. Psychological Science, 10(4), 321-326.

Gerecke, G. A., & House, G. (2013). Update: TMT national culture, demographic heterogeneity, and profit change using LTO-WVS. Research in Business and Economics Journal, 8, 1.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. Nature, 457(7232), 1012-1014.

Greenfield, P. M. (2009). Linking social change and developmental change: shifting pathways of human development. Developmental psychology, 45(2), 401.

Greenfield, P. M. (2016). Social change, cultural evolution, and human development. Current Opinion in Psychology, 8, 84-92.

Hamamura, T., & Xu, Y. (2015). Changes in Chinese culture as examined through changes in personal pronoun usage. Journal of Cross-Cultural Psychology, 46(7), 930-941.

Hofstede, G. (1994). The business of international business is culture. International business review, 3(1), 1-14.

Hofstede, G. (2006). What did GLOBE really measure? Researchers' minds versus respondents' minds. Journal of international business studies, 882-896.

Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. Online readings in psychology and culture, 2(1), 8.

House, R., Javidan, M., Hanges, P., & Dorfman, P. (2002). Understanding cultures and implicit leadership theories across the globe: an introduction to project GLOBE. Journal of world business, 37(1), 3-10.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004). Culture, leadership, and organizations: The GLOBE study of 62 societies: Sage publications.

Maseland, R., & Van Hoorn, A. (2009). Explaining the negative correlation between values and practices: A note on the Hofstede–GLOBE debate. Journal of international business studies, 40(3), 527-532.

Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: comparing the context sensitivity of Japanese and Americans. Journal of personality and social psychology, 81(5), 922.

McCrae, R. R., Terracciano, A., Realo, A., & Allik, J. (2008). Interpreting GLOBE societal practices scales. Journal of Cross-Cultural Psychology, 39(6), 805-810.

Mellon, J. (2011). Search Indices and Issue Salience: the properties of Google Trends as a measure of issue salience. Sociology Working Papers, 1.

Mellon, J. (2013). Where and when can we use Google Trends to measure issue salience? PS: Political Science & Politics, 46(02), 280-290.

Minkov, M., & Blagoev, V. (2012). What do Project GLOBE's cultural dimensions reflect? An empirical perspective. Asia pacific business review, 18(1), 27-43.

Morgan, M. F., Moni, K. B., & Jobling, A. (2009). Who? Where? What? When? Why? How? Question words–What do they mean? British Journal of Learning Disabilities, 37(3), 178-185.

Na, J., & Choi, I. (2009). Culture and first-person pronouns. Personality and Social Psychology Bulletin, 35(11), 1492-1499.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: holistic versus analytic cognition. Psychological review, 108(2), 291.

Park, J., Baek, Y. M., & Cha, M. (2014). Cross-Cultural Comparison of Nonverbal Cues in Emoticons on Twitter: Evidence from Big Data Analysis. Journal of Communication, 64(2), 333-354.

Preis, T., Moat, H. S., Stanley, H. E., & Bishop, S. R. (2012). Quantifying the advantage of looking forward. Scientific reports, 2.

Preis, T., Reith, D., & Stanley, H. E. (2010). Complex dynamics of our economic life on different scales: insights from search engine query data. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 368(1933), 5707-5719.

Qi, H., Manrique, P., Johnson, D., Restrepo, E., & Johnson, N. F. (2016). Open source data reveals connection between online and on-street protest activity. EPJ Data Science, 5(1), 1.

Straub, D., Boudreau, M.-C., & Gefen, D. (2004). Validation guidelines for IS positivist research. The Communications of the Association for Information Systems, 13(1), 63.

Taras, V., Steel, P., & Kirkman, B. L. (2010). Negative practice–value correlations in the GLOBE data: Unexpected findings, questionnaire limitations and research directions. Journal of international business studies, 41(8), 1330-1338.

Troelstra, S., Bosdriesz, J., de Boer, M., & Kunst, A. (2014). Effect of tobacco control policies on information seeking for smoking cessation in the Netherlands: A Google Trends study. The European Journal of Public Health, 24(suppl 2), cku164. 043.

Twenge, J. M., Campbell, W. K., & Gentile, B. (2012 A). Changes in pronoun use in American books and the rise of individualism, 1960-2008. Journal of Cross-Cultural Psychology, 0022022112455100.

Twenge, J. M., Campbell, W. K., & Gentile, B. (2012 B). Increases in individualistic words and phrases in American books, 1960–2008. PloS one, 7(7), e40181.

Uz, I. (2014). Individualism and first person pronoun use in written texts across languages. Journal of Cross-Cultural Psychology, 45(10), 1671-1678.

Vaughan, L., & Romero-Frías, E. (2014). Web search volume as a predictor of academic fame: An exploration of Google trends. Journal of the Association for Information Science and Technology, 65(4), 707-720.

Venaik, S., & Brewer, P. (2008). Contradictions in national culture: Hofstede vs GLOBE. Paper presented at the Academy of International Business 2008 Annual Meeting, Milan, Italy.

Walasek, L., & Brown, G. D. Income Inequality, Income, and Internet Searches for Status Goods: A Cross-National Study of the Association Between Inequality and Well-Being. Social Indicators Research, 1-14.

Yu, F., Peng, T., Peng, K., Tang, S., Chen, C. S., Qian, X., . . . Chai, F. (2016). Cultural Value Shifting in Pronoun Use. Journal of Cross-Cultural Psychology, 47(2), 310-316.