Using Facebook Data to Examine Culture and Self-Disclosure Behaviors

Philip J. Reed, University of Washington
Emma S. Spiro, University of Washington
Carter T. Butts, University of California, Irvine

Abstract
In this work-in-progress poster, we examine the relationship between societal variables, including cultural attributes, and users' self-disclosure on Facebook. To accomplish this we use a dataset of 425,000 Facebook users who designated a national or regional network. Drawing on both standard demographic control variables and the GLOBE cultural dimensions, we execute an exhaustive model search. The best-performing model confirms our hypotheses about cultural variables, but some of our hypotheses about demographic controls are negated. Consequently, we discuss directions in which to continue our research.
Keywords: privacy, self-disclosure, online social networking, culture, demographics, large dataset, GLOBE, Hofstede, cultural dimensions.

1   Introduction

As more social interaction has moved online in recent decades, issues of privacy and self-disclosure have garnered new attention. But self-disclosure is not a new issue; sociologists such as Goffman (1959) and Altman (1975, 1977) wrote extensively about it in the 20th century.

Here we summarize our ongoing research that uses a large-scale data set of roughly a million randomly sampled Facebook users' privacy behaviors to explore the relationship between the social environment and disclosure behaviors as they relate to national cultures. Crucially, whereas extant research has linked national culture to self-reported privacy *concerns* (Bellman, Johnson, Kobrin, & Lohse, 2004; Cho, Rivera-Sánchez, & Lim, 2009), our research is the first to use online social network data to associate observed behaviors with quantitative culture measurements.

Here we report preliminary findings on disclosure and privacy that may prove beneficial to the information science community. We also hope to provoke useful methodological discussions on how large-scale, online social network data can enlighten our understanding of social behavior.

2   Theoretical Literature

Goffman (1959) saw self-disclosure as the process of managing different theatrical performances, allowing each audience to see only the performance intended for it. Online interaction is a new kind of performance--one with an invisible audience (because the individual might not know who could be observing). Online spaces also offer a great reduction in time and space constraints relative to face-to-face interaction, because someone might view online behavior from half a world away or years later (Tufekci, 2008).

Altman (1975) focused on self-disclosure as the "optimization between disclosure and withdrawal," a bidirectional process whereby a person seeks to move her actual privacy level to be in line with her desired level. Consequently, behaviors can either increase or decrease privacy.

Self-disclosure decisions are directly affected by one's social environment; individuals take cues from the context of social interaction in which they operate. Privacy behaviors may be considered the result of social norm formation within a specific cultural environment. Culture is a multi-faceted topic, but one particularly salient theoretical point for our research is the idea that culture reflects both collective values and artifacts that reflect those values (Hofstede, 2006; Javidan, House, Dorfman, Hanges, & De Luque, 2006; Schein, 2010). In this project, we examine how self-disclosure behaviors reflect one specific aspect of the social environment, namely, national cultural characteristics.

3   Measuring the social environment

Our work draws on two broad classes of explanatory variables. The first consists of standard socioeconomic demographic control variables indicators, including such variables as the Human Development Index (HDI) and the proportion of Internet users. The second, the GLOBE set of cultural indicators, may be less familiar.

Efforts to measure culture in the business literature extend back to Hofstede (1984), who developed a series of four "dimensions" (since expanded to six). Although Hofstede's approach has faced both ontological and methodological criticisms, his dimensions have continued to be the most widely used cultural indicators (Baskerville, 2003; Minkov & Hofstede, 2011; Tang & Koveos, 2008) in fields including information technology privacy research (Bellman et al., 2004; Cho et al., 2009).

However, numerous alternative sets of cultural indicators have arisen to complement Hofstede's dimensions (Inglehart & Welzel, 2010; Schwartz, 2014; Trompenaars & Hampden-Turner, 1998). For this project, we use the GLOBE dimensions (House & Javidan, 2004), the result of a rigorous survey development process involving the local expertise of 160 researchers from 62 cultures. GLOBE posits 9 basic attributes with two dimensions each, yielding 18 dimensions total. Each attribute has both a *practices* dimension, reflecting respondents' impression of how their society *is,* and a corresponding *values* dimension, reflecting how respondents state that society *should be*. Importantly, the GLOBE researchers were surprised to find that the majority of the values variables are negatively correlated to the corresponding practice variables.

Table 1 includes a complete list of both the socioeconomic indicators and the GLOBE variables used.

4   Data

Our research utilizes a unique dataset of observed privacy behaviors by 976,301 randomly sampled individual Facebook users, collected at a time when Facebook infrastructure allowed querying of arbitrary IDs for certain privacy settings (Gjoka, Kurant, Butts, & Markopoulou, 2009, 2011; Spiro, Pierski, & Butts, 2012). We aggregate this data by country for those users who designated a geographical network for their country, region, or city (425,137 of 976,301 overall users, a proportion of 0.435), yielding data for precisely 100 countries. These data contain four privacy-related behaviors of interest, concerning "friend requests," users' profile photos, users' lists of friends, and private messages.

Each of these four attributes was, in 2009, set by default to allow the interaction in question, i.e.at a lower-privacy setting. We consider a user who changed any of the four attributes to a higher-privacy setting to have exhibited *privacy-awareness* (PA)*,* and a user who did not change any of the four to have shown *privacy-unawareness.*

Figure 1 shows the observed relationship between PA and GLOBE's Gender Egalitarianism (practice) variable, demonstrating how one of the cultural dimensions is empirically related to privacy behavior.

5   Methods

We model privacy awareness versus unawareness as a function of the two broad categories of variables already mentioned, basic demographic control variables and measures of culture. Because each individual has one of two mutually exclusive behaviors—either they are privacy aware (have at least one privacy setting turned on) or they are not privacy aware (have default privacy settings)—we use binomial regression to estimate the relationship between each country's explanatory variables and its users' privacy behavior.

There are some 3.4 million combinations of our 25 explanatory variables that include at least one of the 18 GLOBE cultural variables. To choose among these, we use a standard goodness of fit measure, the Akaike information criterion (AIC). We perform an exhaustive review of all 3.4 million combinations, selecting the model with the lowest AIC score.

6   Hypotheses

Proceeding from the theoretical position that sociocultural context is related to privacy behavior, we have developed hypotheses for specific relationships between our societal covariates (both demographic controls and GLOBE dimensions) and privacy awareness. Previously scholars have generated similar hypotheses for the Hofstede dimensions and privacy attitudes (Bellman et al., 2004; Cho et al., 2009), but no such precedent exists for the GLOBE dimensions and privacy behaviors. Although some of the GLOBE dimensions have similar names and in some sense derive from analogous concepts in Hofstede, the nuances of the differences between them can be treacherous (see, e.g., Venaik & Brewer, 2010).

Seven of the nine GLOBE attributes show a negative correlation between the corresponding practice and values dimensions, a topic of vigorous scholarly debate (Brewer & Venaik, 2010; Hofstede, 2006; Maseland & van Hoorn, 2010; Taras, Steel, & Kirkman, 2010). Our hypotheses focus on the practice ("is") dimensions rather than the values ("should be") ones.

We likewise strongly hypothesize that PA is positively related to both Uncertainty Avoidance and Power Distance. In addition, we hypothesize positive relationships between PA and Future Orientation and *In-group* Collectivism, but negative relationships between PA and *Institutional* Collectivism, gender egalitarianism, and humaneness. Our hypotheses are summarized in Table 1.

7   Preliminary findings

Our preliminary findings have been encouraging. We have performed an exhaustive examination of multivariate models for combinations of GLOBE dimensions, with or without any of the demographic control variables, seeking the lowest-AIC models.

The top model uses ten of the 18 GLOBE covariates and five of the seven control covariates. Six of our seven hypotheses about GLOBE practice variables were confirmed by the regression, with the seventh variable excluded from the model. All of the six coeffcients were significant (p<.05).

However, several of our hypotheses about the demographic control variables were not confirmed. The following three variables had coefficients in the top model with a sign contrary to our expectation: trust ($p<.01$), population 65 and over ($p<.001$), and female to male ratio in primary and secondary education ($p<.001$). However, our hypothesized sign for HDI was confirmed ($p<.001$). We suspect these findings result from unexpected interactions between demographic and GLOBE variables, for example between the gender-related variables in each, but we continue to investigate.

8   Discussion

Our work to this point indicates the existence of a relationship between most of the cultural variables, both Hofstede and GLOBE, and individuals' privacy behaviors on Facebook. However, some open questions continue to demand further analysis, particularly regarding the interaction of our demographic control variables with the GLOBE cultural dimensions and the impact of that interaction on the coefficients of the control variables. We plan to dig deeper into understanding these relationships among these variables.

In our work-in-progress poster presentation, we contribute to the scholarly discussion on theoretical, empirical, and methodological levels. Theoretically, we suggest some ways in which culture, particularly the attributes highlighted by the Hofstede and GLOBE data, interacts with online privacy. Empirically, we affirm theoretical predictions of relationship between these attributes and online privacy behavior, as distinct from self-reported beliefs. Methodologically, we demonstrate how large-scale social network data can be used to test privacy-and culture-related hypotheses. We hope that these contributions are only the start of an extensive conversation among information scientists about privacy and culture.

References

Altman, I. (1975). The environment and social behavior: privacy, personal space, territory, and crowding. Retrieved from http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED131515

Altman, I. (1977). Privacy regulation: culturally universal or culturally specific? *Journal of Social Issues*, *33*(3), 66–84.

Baskerville, R. F. (2003). Hofstede never studied culture. *Accounting, Organizations and Society*, *28*(1), 1–14.

Bellman, S., Johnson, Eric, Kobrin, Stephen, Lohse, Gerald, Johnson, E., Kobrin, S., & Lohse, G. (2004). International differences in information privacy concerns: A global survey of consumers. *The Information Society*, *20*(5), 313–324.

Brewer, P., & Venaik, S. (2010). GLOBE practices and values: A case of diminishing marginal utility? *Journal of International Business Studies*, *41*(8), 1316–1324.

Cho, H., Rivera-Sánchez, M., & Lim, S. S. (2009). A multinational study on online privacy: global concerns and local responses. *New Media & Society*, *11*(3), 395–416. doi:10.1177/1461444808101618

Gjoka, M., Kurant, M., Butts, C. T., & Markopoulou, A. (2009). A walk in facebook: Uniform sampling of users in online social networks. *arXiv Preprint arXiv:0906.0060*. Retrieved from http://arxiv.org/abs/0906.0060

Gjoka, M., Kurant, M., Butts, C. T., & Markopoulou, A. (2011). Practical recommendations on crawling online social networks. *Selected Areas in Communications, IEEE Journal on*, *29*(9), 1872–1892.

Goffman, E. (1959). *The presentation of self in everyday life.* Garden City, NY: Anchor Doubleday Books.

Hofstede, G. (1984). *Culture's consequences: International differences in work-related values* (2nd ed., Vol. 5). Beverly Hills, CA: Sage Publications.

Hofstede, G. (2006). What did GLOBE really measure? Researchers' minds versus respondents' minds. *Journal of International Business Studies*, *37*(6), 882–896.

House, R. J., & Javidan, M. (2004). Overview of GLOBE. In R. J. House, P. J. Hanges, M. Javidan, P. W. Dorfman, & V. Gupta (Eds.), *Culture, leadership, and organizations: the GLOBE study of 62 societies*. Thousand Oaks, Calif.: Sage Publications.

Inglehart, R., & Welzel, C. (2010). Changing mass priorities: The link between modernization and democracy. *Perspectives on Politics*, *8*(02), 551–567. doi:10.1017/S1537592710001258

Javidan, M., House, R. J., Dorfman, P. W., Hanges, P. J., & De Luque, M. S. (2006). Conceptualizing and measuring cultures and their consequences: a comparative review of GLOBE's and Hofstede's approaches. *Journal of International Business Studies*, *37*(6), 897–914.

Maseland, R., & van Hoorn, A. (2010). Values and marginal preferences in international business. *Journal of International Business Studies*, *41*(8), 1325–1329.

Minkov, M., & Hofstede, G. (2011). The evolution of Hofstede's doctrine. *Cross Cultural Management: An International Journal*, *18*(1), 10–20.

Schein, E. H. (2010). *Organizational Culture and Leadership* (4th ed.). Hoboken, NJ: Wiley. Retrieved from http://orbis.eblib.com/patron/FullRecord.aspx?p=588878

Schwartz, S. H. (2014). National culture as value orientations: Consequences of value differences and cultural distance-Chapter 20. In V. A. Ginsburgh & D. Throsby (Eds.), *Handbook of the Economics of Art and Culture* (Vol. 2, pp. 547–586).

Spiro, E. S., Pierski, N. M., & Butts, C. T. (2012). Mapping an ecology of privacy: A Cross-national comparison of control of self-presentation online. Retrieved from http://citation.allacademic.com/meta/p564099_index.html

Tang, L., & Koveos, P. E. (2008). A framework to update Hofstede's cultural value indices: economic dynamics and institutional stability. *Journal of International Business Studies*, *39*(6), 1045–1063.

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.

Trompenaars, F., & Hampden-Turner, C. (1998). *Riding the waves of culture*. McGraw-
Hill New York. Retrieved from http://ocan.yasar.edu.tr/wp-
content/uploads/2013/09/Riding-the-waves_Part-1.pdf

Tufekci, Z. (2008). Can you see me now? Audience and disclosure regulation in online
social network sites. *Bulletin of Science, Technology & Society*, *28*(1), 20–36.

Venaik, S., & Brewer, P. (2010). Avoiding uncertainty in Hofstede and GLOBE. *Journal
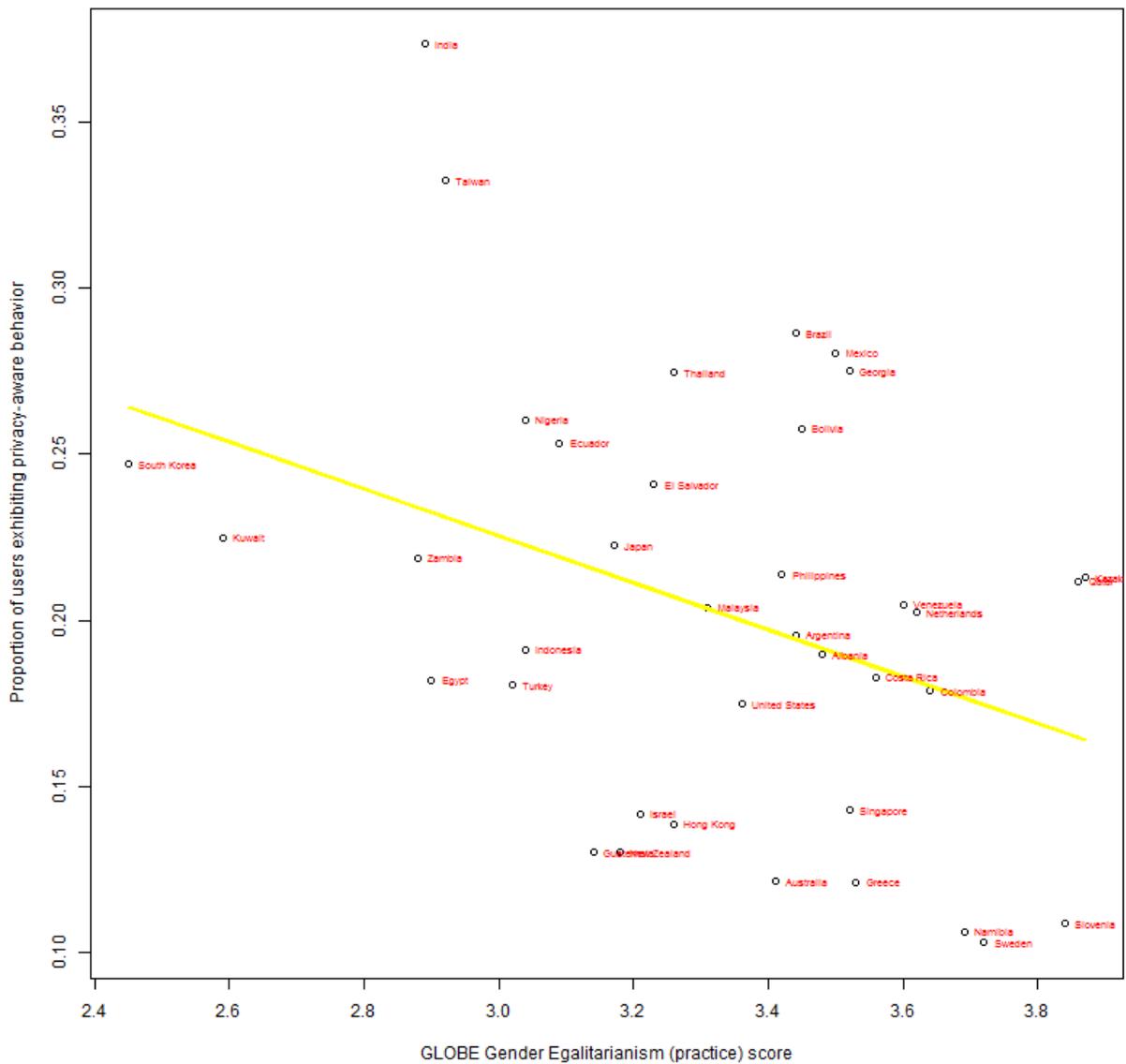of International Business Studies*, *41*(8), 1294–1315.

Table of Figures

Table of Tables

Figure 1. Scatter plot of privacy-aware behavior versus societal gender egalitarianism

Socio-economic indicators (7 variables)		
Human Development Index, 2010	negative	HDI correlates very highly with political openness and democracy—the reason why we did not use any of these political measures—and in more open societies people have less to fear from self-disclosure.
GDP per capita		
Female-to-male ratio, primary and secondary school	negative	More gender-equal societies carry fewer threats to self-disclosure.
Population age 65 and over	negative	Older users are less likely to have experience changing settings to control self-disclosure.
Urban population		
Trust	negative	In more trusting societies, fewer threats are perceived.
Internet users as a proportion of population		
GLOBE cultural dimensions (18 variables)		
Assertiveness (practice)		
Assertiveness (value)		
Institutional Collectivism (practice)	negative	Where large institutions are trusted to provide for individuals, the latter feel less need to take privacy into their own hands.
Institutional Collectivism (value)		
In-group Collectivism (practice)	positive	Where group identity is strong, competing groups can pose threats to each other.
In-group Collectivism (value)		
Future Orientation (practice)	positive	Privacy threats involve a future state, so future-oriented cultures consider them more intensively.
Future Orientation (value)		
Gender Egalitarianism (practice)	negative	More gender-equal societies carry fewer threats to self-disclosure.
Gender Egalitarianism (value)		
Humane Orientation (practice)	negative	Humane societies carry fewer threats.
Humane Orientation (value)		
Performance Orientation (practice)		
Performance Orientation (value)		
Power Distance (practice)	positive	Where PD is great, users perceive more risk.
Power Distance (value)		
Uncertainty Avoidance (practice)	positive	Privacy threats are a form of uncertainty, so higher-UA cultures feel more need to avoid them.
Uncertainty Avoidance (value)		

Table 1. List of explanatory variables with hypothesized direction of relationship.

Sources: World Bank; Trust as compiled by ASEP/JDS based on World Values Survey and other analogous surveys; GLOBE dimensions as compiled at Harzing.com.

9