Mining Linguistic Diversity as a Signal for Identifying Regions Likely to Adapt to the Changing Nature of Work

James Bain
University of Missouri
Anthropology
jcbcnc@mail.missouri.edu

Goggins, Sean
University of Missouri
Computer Science
outdoors@acm.org

1. INTRODUCTION

The experience of work is fundamentally a story of the labor of people, groups and project organizations. Work's future relies on a conquering of challenges unique to distributed, multi-disciplinary teams likely to increase in dominance. These demands are perturbed byproducts of how humans perceive their own and others performance. In this paper we first describe how our central theoretical lens of social comparison theory connects to related theories explaining the social behavior of humans in work settings. Second, we identify a few problems that arise from suboptimal social comparisons driven by our cognitive biases. Third, we describe a research project for examining these phenomena in the context of local economic health and physical well-being contrasted across populations with different bias characteristics.

2. Social Comparison, Self-Efficacy, Group Identity and Intergroup Relations

People are driven to evaluate their own abilities, and in the absence of an objective measure, will compare themselves with others rather than persist in the cognitive dissonance of not knowing the score [6]. Evading dissonance also drives people to seek comparisons with individuals most like themselves. The dissonance phobia of our humanity skews us away from objective measures altogether when the skills or abilities we are evaluating grow too complex for concise assessment [4]. Perhaps we compare our SAT scores against a normal distribution, but evaluating an individual paper is a more complex endeavor. Inspirational philosophies like the Desiderata caution us "not to compare ourselves with others, for we may become vain and bitter" [17] precisely because we love comparing ourselves with others.

Our identity is subsequently shaped by how we perceive our abilities. Our individual performance expectations will rise or fall with what we believe we are capable of accomplishing [2]. The human tilt toward social comparison is effectively hardened by reinforcement from our emergent "self-efficacy". That's how the individual operationalizes their sense of who they are and what they might accomplish. Our individual sense of ourselves and our abilities then intersects with an identity constructed through social interaction. We compare our abilities with those of others, and we learn the most from people whose abilities are not too different than our own [16].

Who we are is also defined by the social groups we identify with, and we tend to find our greatest comfort in the people we think to be most like us [13,15]. Once solidly on a team, in a group or identified as part of some human "set", we begin the process of comparison all over again [12,15]. Sometimes we wear the jerseys of our favorite athletic teams, or join online groups that reinforce our identities as scholars or parents [9]. Sometimes our biases cause us to make fatal errors in the assessment of "who is most like us". Our judgement can be short circuited to focus on identifiable, outward markers like gender and race [12–15]; and when pressed economically, these assessment failures may undermine the fate of entire communities.

In the future of work, constructed as a highly distributed, less organizationally focused endeavor, the competition between communities will become staggeringly more complex. Instead of methodical, institutionalized approaches for framing working relationships, individuals will be pressed to make judgements about collaborators on their own. The social and economic elevation of individuals, locales, subcultures and nations will be impacted by biases not connected to the suitability of individuals for specific types of work. Understanding how such a relationship between the collective bias of social collections of people and their economic and social well-being exists, and how it functions, warrants inquiry.

3. Empirical Examples

Two areas where we are presently inquiring about the role bias plays in the future of work are open source projects, where toxicity, particularly toward female contributors, is a present concern. The second is Twitter discourse, where we are exploring how anti-minority sentiment differs by locale, locale education level and the extent to which there is correspondence between racist language, linguistic diversity, and local economic, health and well-being indicators.

3.1 Open Source Project Toxicity

OSS project contributors and projects have long measured themselves using metrics, but insights into project health and resilience have proven elusive. Measuring commits was, for example, long used as a central metric for understanding individual contributions to the Linux Kernel. Kernel leaders and contributors accepted the utility "commit" a metric. The success of the Kernel popularized the use of that metric in a number of OSS ecosystems. We do not know if "commits" are a useful metric on a wide range of OSS projects or not; but we suspect more robust metrics exist.

It is not out of the question that the biases embed in commits drives negative behavior. As a detailed example, consider open source project toxicity. We know that the signals of toxicity are diverse. Sometimes they include identifiably hostile messages (i.e., an identifiable perturbation of the general project sentiment and emotional classification in communication), and other times toxicity is identifiable by a sudden "silence" in communication on primary project channels. Specifically, member identified toxicity and project communication style are the two metrics we can generate through the advancement of existing technologies and computational linguistic methods.

3.2 Public Discourse on Twitter

Recently, the United States and Europe have seen a rise in nationalist political parties who hold anti-immigrant/anti-minority sentiments [11], yet the reason for its spread is not well understood. Several theories as to why people hold such opinions toward others have been proposed [1,3] as well as evidence in support for these theories [8]. Even evolutionary explanations for in-group/outgroup sentiments have been suggested [10]. However, these relied on small samples and survey data or weren't direct measurements of anti-immigrant sentiment but, instead, political behavior [11].

Twitter, a micro-blogging, social media website where users posts short texts of no more than 140 characters[1] called tweets, provides a complimentary method to evaluating peoples' feelings toward topics. Tweets are often personal in nature where a user describes their plans, their opinions, or what they are doing [5], although media (i.e. images, videos, gifs) are often shared as well. Although only 1% of tweets are geotagged with exact coordinates, the majority of twitter users do add a location to their profile indicating where they live [7]. This allows me to assess aggregated sentiments toward minority and immigrant populations at a geographic level. When combined with city-level Census data, this will allow for me to assess how demographic, economic and social variables influence anti-minority sentiment at the city-level.

4. Research Proposal

In the interest of space, we are focusing our proposal on Twitter data, but are prepared to discuss detailed proposals for each of our empirical examples. The following hypotheses about minority/immigrant sentiment will be tested:

H1: Greater levels of integration of minority and immigrant populations within a city will mitigate anti-minority sentiment.

Individuals from city neighborhoods with a proportion of minority populations have more positive feelings toward minority populations (McDermott 2015). Using Census-block data on ethnic and racial diversity, I will evaluate the relationship between the levels of city integration with the aggregated sentiment towards minority populations.

H2: Increase unemployment in combination with lower levels of education will increase anti-minority sentiment.

Employment rates have shown to have an inverse effect on anti-immigrant sentiment. Education has also shown to share a complex relationship with unemployment levels with regards to anti-immigrant sentiment. Using Census data to measure the rate of unemployment and education levels of a city, I will compare this against aggregated anti-minority/anti-immigrant sentiment.

H3: Anti-minority rhetoric will demonstrate greater levels of group vs. group speak as opposed to individual vs. group speak.

Anti-minority sentiment is hypothesized to be more of a product of perceived threat against a group rather than individual experiences with minority populations. If this is the case, then I would expect tweets marked as more anti-minority will have a greater presence of first-person plural pronouns (*we* and *us*) and fewer first person singular pronouns (*I* and *me*) to demonstrate threats to group rather than threats to self.

5. ACKNOWLEDGMENTS

Our thanks to Syracuse University, the University of Missouri Data Science and Analytics Program, The University of Missouri Anthropology Department, Matt Blackwood, Derek Howard, Christian Cmhil-Warn, Chi-Ren Shyu the Sloan Foundation and the National Science Foundation for their Support.

6. REFERENCES

1. GW Allport. 1954. The nature of prejudice: Boston. The nature of prejudice: Boston.
2. A Bandura. 1977. Social Learning Theory. Prentice-Hall, Englewood Cliffs, NJ.
3. Herbert Blumer. 1958. Race Prejudice as a Sense of Group Position. The Pacific Sociological Review 1, 1: 3–7. https://doi.org/10.2307/1388607
4. Justin T. Buckingham and Mark D. Alicke. 2002. The influence of individual versus aggregate social comparison and the presence of others on self-evaluations. Journal of Personality and Social Psychology 83, 5: 1117–1130. https://doi.org/10.1037//0022-3514.83.5.1117
5. Edward Crook. 2012. The Twitter Landscape. A Brand watch social insights report.
6. L. Festinger. 1954. A Theory of Social Comparison Processes. Human Relations 7: 117–140.
7. Jacob Goldenberg and Moshe Levy. 2009. Distance is not dead: Social interaction and geographical distance in the internet era. arXiv preprint arXiv:0906.3202.
8. Jens Hainmueller and Dominik Hangartner. 2013. Who Gets a Swiss Passport? A Natural Experiment in Immigrant Discrimination. American Political Science Review 107, 01: 159–187. https://doi.org/10.1017/S0003055412000494
9. Y. Ren, R. Kraut, and S. Kiesler. 2007. Applying common identity and bond theory to design of online communities. Organization Studies 28, 3: 377–408.
10. Patricia A Roos. 2009. Interconnecting work and family: Race and class differences in women's work status and attitudes. WSQ: Women's Studies Quarterly 37, 2: 103–120.
11. Per Strömblad and Bo Malmberg. 2016. Ethnic segregation and xenophobic party preference: Exploring the influence of the presence of visible minorities on local electoral support for the Sweden Democrats. Journal of Urban Affairs 38, 4: 530–545.
12. H. Tajfel. 1974. Social Identity and Intergroup Behavior. Social Science Information 13: 65–93.
13. Henri Tajfel. 1978. Differentiation Between Social Groups. Academic Press, New York.
14. Henri Tajfel. 1982. Social Identity and Intergroup Relations. Cambridge University Press, London.
15. J.C. Turner, R.J. Brown, and H. Tajfel. 1979. Social Comparison and Group Interest in Ingroup Favouritism. European Journal of Social Psychology 9: 187–204.
16. L. Vygotsky. 1962. Thought and Language. MIT Press, Cambridge, MA.
17. Wikipedia. 2017. Max Ehrmann. Wikipedia. Retrieved from http://en.wikipedia.org/wiki/Max_Ehrmann

[1] Twitter did introduce a 280 character limit in October of 2017 and it is currently in beta being tested on a small subset of users.