Proposal | Predicting the Trust Radius with Machine Learning:

To What Extent Are Those "Most People" Out-Groups?

Background and aims

Generalized (social) trust (hereafter GT) or trust in "most people" is a conspicuous indicator in social survey research on intergroup relations and social cohesion. In a recent meta-analysis (Van der Meer and Tolsma 2014), forty-four out of ninety studies take GT as their outcome variable and assume that when respondents are asked to answer "would you say that most people can be trusted, or would you be careful in dealing with them?", they evaluate the trustworthiness of unknown people or even ethnic out-groups (see also Dinesen, Schaeffer, and Sønderskov forthcoming). While this prior research has debated the negative link between ethnic diversity and GT as a proxy for social cohesion, with some notable exceptions, relatively few have so far focused on systematic response bias in answers to GT.[1]

These mixed findings may well be due to the operationalization of social cohesion by GT. For example, a British study using think aloud protocols has demonstrated that the majority of respondents high in GT think "most people" refers to people they know, whereas a high proportion of those who are low in GT think about strangers (Sturgis and Smith 2010). Following the homophily principle – the tendency of individuals to associate and bond with similar others – we can expect that people known to the respondents are ethnic in-groups, and strangers are more likely to be ethnic out-groups, although formally we do not know this (cf., McPherson, Smith-Lovin, and Cook 2001, Delhey, Newton, and Welzel 2011). A faceless stranger one hypothetically meets for the first time could well be an ethnic in-group in homogenous settings as much as a person known to the respondent (friend, neighbor, family member) can be an ethnic out-group in diverse settings. In this project we, therefore, return to some of the overlooked basics in this literature: the measurement and conceptualization of GT, which is allegedly in decline by ethnic diversity. The proposed project aims a) to examine the validity of the Generalized Trust (GT) question as a measure of out-group attitudes and implicit race bias in an unprecedentedly broad manner using Supervised Machine Learning Ensembles (hereafter ML), and b) to cross-validate the results with 3 existing large-scale social surveys. The research questions are: (1) To what extent can GT be predicted by prejudice? (2) Can we cross-validate the findings with data that have less information on implicit prejudice and intergroup attitudes?

Data & Methods

This project first proposes to examine the validity of GT as a proxy for social cohesion conceptualized as intergroup attitudes and implicit prejudice by employing data from Project

[1] Although informative, most of the existing research on measurement bias assumes that within-country populations who speak the same language are homogenous in how they interpret GT (Freitag and Bauer 2013; Reeskens and Hooghe 2008; Van der Veld and Saris 2011). Moreover, there is less focus on the extent to which GT taps into feelings about ethnic out-group.

Implicit (Xu et al. 2019). Second, it proposes to test the predicted model on data from: World Values Study (WVS), the European Social Survey (ESS), and the Dutch LISS panel (LISS).

Project Implicit has collected 34 different self-reported scales (including GT and self-reported racial attitudes from intergroup relations literature) together with reaction times from the Race Implicit Association Test (Axt 2018). This data also contains detailed socio-demographic characteristics, geographic location, and political preferences of the respondent.

In the proposed project, we extend the classical test theory (Novick 1966): a set of principles that allows us to determine how successful our proxy indicator (e.g., GT) is; to predict an unobservable phenomenon (e.g., social cohesion). A first crucial step is to examine the convergent validity[2] of GT by assessing to what extent it can be explained by explicit out-group versus in-group attitudes by simultaneously testing the appropriateness of other constructs. Following the think aloud results discussed above, one can see that both in-group and out-group attitudes may have a unique effect on GT. But, second, if "most people" in the GT-question is more often interpreted to be an out-group, it should also have a reasonable correlation with an affective, more automatic, measure such as implicit prejudice (Implicit Association Test). Third, GT should be less related to socio-demographic characteristics such as age, gender, and education, if it is a good proxy for social cohesion across ethnic groups, since these variables are less directly tapping into prejudice than explicit and implicit race bias.

The first innovation of the proposed project is thus as follows. Instead of relying on *ex ante* model sparsity and favoring a set of variables of interests over others, which are dictated by simplified theoretical models of human behavior, we can make use of quantifying complexity. In other words, we can "have an explicit numeric measure of model complexity, [and] …view it as a parameter that can be "tuned" to produce the best out of sample predictions" (Varian 2014, p. 7). The advantage of ML over conventional statistical analyses (Samii, Paler, and Daly 2016; Varian 2014) lies, first, in its flexibility to select variables when many potential predictors are available. Second, we can model nonlinear relationships. Third, there are less limits to the number of datasets, observed cases, interactions between variables, and modelling strategies. Finally, the results are less tainted by researcher degrees of freedom in preferring a scale or measure over another. Our goal with prediction, however, remains in line with what social science generally attempts to do, to get good out-of-sample predictions, and to avoid overfitting. Therefore, we train, validate, and test the model again (holdout) with differently sized random subsamples of the data. In short, we perform (a sort of) meta-analysis by predicting validity from characteristics of the individual and other survey items.

In addition, this project has another innovation by cross-validating the predicted model on survey data where less variables are available. Out of sample prediction will hence be extended to other social survey data: WVS, ESS, and the LISS panel. Since social science research is costly and survey practitioners often face the problem of overburdening participants with many questions, it

[2] The degree to which two measures that theoretically should be related, are in fact related.

is crucial to be able to rely on fewer questions. As we know a person's GT score in WVS, ESS, and LISS, we can test the model with data that has less indicators, and assess the quality of our model. The results of this project will benefit researchers and practitioners interested in the quality of social cohesion indicators.

References

Axt, Jordan R. 2018. "The Best Way to Measure Explicit Racial Attitudes Is to Ask About Them." *Social Psychological and Personality Science* 9(8): 896–906.

Delhey, Jan, Kenneth Newton, and Christian Welzel. 2011. "How General Is Trust in 'Most People'? Solving the Radius of Trust Problem." *American Sociological Review* 76(5): 786–807.

Dinesen, Peter Thisted, Merlin Schaeffer, and Kim Mannemar Sønderskov. "Ethnic Diversity and Social Trust: A Narrative and Meta-Analytical Review Ethnic Diversity and Social Cohesion View Project Perceived and Actual Discrimination View Project." *Preprint*.

Freitag, Markus, and Paul C Bauer. 2013. "Testing for Measurement Equivalence in Surveys Dimensions of Social Trust across Cultural Contexts." *Public Opinion Quarterly* 77(S1): 24–44.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook. 2001. "Birds Of A Feather: Homophily in Social Networks." *Annual Review of Sociology* 27: 415–44.

Van der Meer, Tom, and Jochem Tolsma. 2014. "Ethnic Diversity and Its Effects on Social Cohesion." *Annual Review of Sociology* 40(1): 459–78.

Novick, Melvin R. 1966. "The Axioms and Principal Results of Classical Test Theory." *Journal of Mathematical Psychology* 3(1): 1–18.

Reeskens, Tim, and Marc Hooghe. 2008. "Cross-Cultural Measurement Equivalence of Generalized Trust. Evidence from the European Social Survey (2002 and 2004)." *Social Indicators Research* 85(3): 515–32.

Samii, Cyrus, Laura Paler, and Sarah Zukerman Daly. 2016. "Retrospective Causal Inference with Machine Learning Ensembles: An Application to Anti-Recidivism Policies in Colombia." *Political Analysis* 24(4): 434–56.

Sturgis, Patrick, and Patten Smith. 2010. "Assessing the Validity of Generalized Trust Questions: What Kind of Trust Are We Measuring?" *International Journal of Public Opinion Research* 22(1): 74–92.

Varian, Hal R. 2014. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives* 28(2): 3–28.

Van der Veld, William M, and Willem E Saris. 2011. "Causes of Generalized Social Trust." *European association for methodology series*: 207–47.

Xu, Kaiyuan, Nicole Lofaro, Brian A. Nosek, Anthony G. Greenwald, and Jordan Axt. 2019. "Race IAT 2002-2018." *OSF*. March 28. osf.io/52qxl.