The Identity Labeling Problem

Kenneth Joseph

Network Science Institute, Northeastern University

Institute for Quantitative Social Science, Harvard University

ACT asserts that the way we label ourselves and others impacts our behavior in social settings. What the theory has never demanded, however, is a rigorous model of how one decides which identities to apply in the first place. Roughly speaking, Affect Control Theorists, in particular Heise and MacKinnon (2010), have purported a heuristic two-step model for how such decisions are made. First, an individual determines the most appropriate institution for a given social situation. This decision limits the scope of possible identities that could be applied - for example, an interaction in the "family" institution cannot have a participant with the identity "professor". After this institutional decision is made, affective information and partial definitions of the social situation at hand are then used to "fill in the blanks" for the identity of any unlabeled individuals.

In my recent work, I have suggested and provided initial remedies for at least two problems with this conceptualization of how individuals are labeled:

First, institutions simply do not exist within (Bayes)ACT's mathematical model; they are merely hard and hardly-ever used constraints. It is for this reason that artificial agents in BayesACT simulations can revert to situations like the following, quoted from Schröder et al. (2017): "...both agents have developed the shared belief that one of them (agent A) is an 'executioner' while the other (agent B) is a 'great grandmother' ". While *affectively*, this made sense given initial constraints, from an institutional perspective this identity pairing is unlikely. To this end, the first part of my talk would introduce recent work on a statistical model of ACT that jointly considers institutional and affective information (Joseph et al., 2017). I will give a high-level explanation of how I learn the parameters of this model (including EPA profiles for identities not in existing dictionaries) from text data. Figure 1 shows joint affective/institutional structures learned by the model that might be used in to provide labels for individuals in social situations.

Second, there is limited evidence that a two-step model of identity selection (i.e. first choose an institution, then rely on affective information) is actually how people make identity labeling decisions in the real world. Indeed, recent evidence from "cognitive" social psychologists suggests that affective and semantic information work *together*, at the same time, to produce identity labelings (Ehret et al., 2014). To this end, I will give a high level overview of both published (Joseph and Carley, 2016) and more recent unpublished work (in collaboration Jonathan Morgan) showing that, indeed, a two-step process may not be appropriate to describe how people define social situations with identities. Instead, affective and semantic (institutional) information combine in interesting ways to produce definitions of situation. This evidence is drawn from survey experiments that ask respondents questions like the one displayed in Figure 2, where respondents much choose which identity best "fits" an individual in a hypothetical social situation.

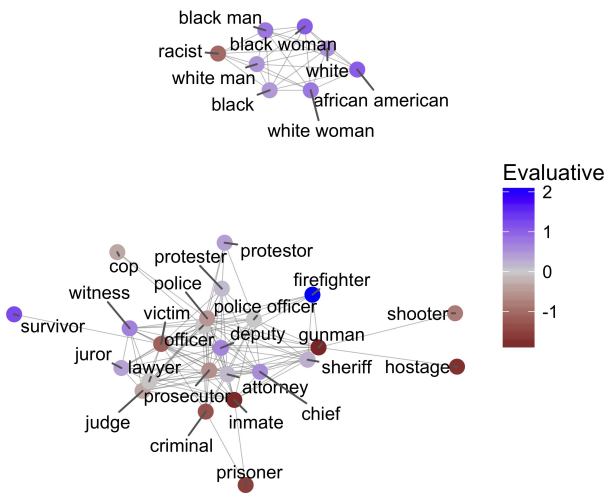

Figure 1: Network views of two institutional structures (a "legal/protest" institution and a "race" institution) learned by the model I have developed on Twitter data relevant to the deaths of Michael Brown and Eric Garner. Links represent strong semantic relationships (and therefore, clusters can be considered institutions) and identities are colored by their evaluative meaning.

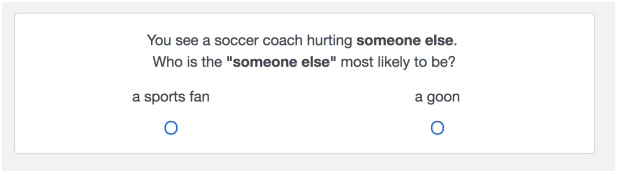

Figure 2: An example of the type of questions we asked survey participants to understand how institutional (semantic) and affective structures combine to generate particular labelings of individuals

References

Ehret, P. J., Monroe, B. M., and Read, S. J. (2014). Modeling the Dynamics of Evaluation A Multilevel Neural Network Implementation of the Iterative Reprocessing Model. *Personality and Social Psychology Review*, page 1088868314544221.

Heise, D. R. and MacKinnon, N. J. (2010). *Self, Identity, and Social Institutions*. Palgrave Macmillan.

Joseph, K. and Carley, K. M. (2016). Relating semantic similarity and semantic association to how humans label other people. *NLP+ CSS 2016*, page 1.

Joseph, K., Wei, W., and Carley, K. M. (2017). Girls rule, boys drool: Extracting semantic and affective stereotypes from Twitter. In *2017 ACM Conference on Computer Supported Cooperative Work.(CSCW)*.

Schröder, T., Hoey, J., and Rogers, K. B. (2017). Modeling dynamic identities and uncertainty in social interactions: Bayesian affect control theory. *Am. Soc. Rev.*