Modeling Dynamic Identities and Uncertainty in Social Interactions: Bayesian Affect Control Theory

Tobias Schröder[1], Jesse Hoey[2], and Kimberly B. Rogers[3]

Identity is a core concept in social science, thought to link individual minds with the social order and collective patterns of sense-making. Identities are paradoxical things. They emerge from individuals' subjective interpretations of interactions and yet they are part of a cultural consensus on social structures. They are properties of the person as much as they result from the affordances of the situation. Most identity meanings are highly stable in society and yet they are subject to constant renegotiations and sometimes dramatic shifts. Here, we propose a novel theory of social interaction that accounts for the paradoxical nature of human identities. The theory – a generalization of David Heise's (2007) affect control theory (ACT) based on Bayesian probability theory – is implemented as a computational model, which we have called BayesACT (Hoey et al. 2013).

Affect Control Theory (ACT; Heise, 2007)

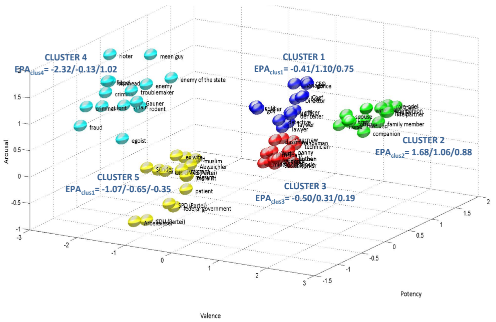


Figure 1: Social concepts in affective space.

Based on Osgood's (1962) finding that affective dimensions of Evaluation, Potency, and Activity (EPA) organize semantic relations of concepts (Fig. 1), ACT's empirical base consists of measured fundamental sentiments **f** attached to words denoting identities and behaviors. These "affective dictionaries" model collective representations of society, which have been shown in many studies (some of them with representative samples) to be highly consensual and stable within cultures (Fig. 1; Ambrasat et al. 2014).

Situated events create transient impressions $\boldsymbol{\tau}$ of identities that may more or less deviate from the fundamental sentiments. E.g., a grandfather who yells at a child seems less nice than a prototypical grandfather. This impression formation is modeled with a nonlinear set of equations: $\boldsymbol{\tau}=\boldsymbol{M}\mathbf{f}$ (eq.1) where the prediction coefficients $\boldsymbol{M}$ come from empirical rating studies. ACT postulates that humans are motivated to confirm identity meanings when interacting. ACT's computational model INTERACT (Heise 1997) generates behavior predictions by minimizing the "deflection" $\boldsymbol{D}=\Sigma(\boldsymbol{\tau}-\mathbf{f})^2$ (eq.2) summed over EPA dimensions, identities, and behaviors relevant to the situation.

BayesACT: A Probabilistic Generalization

ACT has good empirical support, but its mathematical models do not suitably capture the uncertain, dynamic, and paradoxical nature of identity. Therefore, ACT was generalized probabilistically and formulated as a partially observable Markov decision process (POMDP; Hoey et al. 2013). Identities are represented as probability distributions in EPA space, not as points, which is more compatible with sociological theories of identities as uncertain, fluid and subject to constant change. Dynamic changes of identities during interactions are modeled as changes in the shape of the distribution.

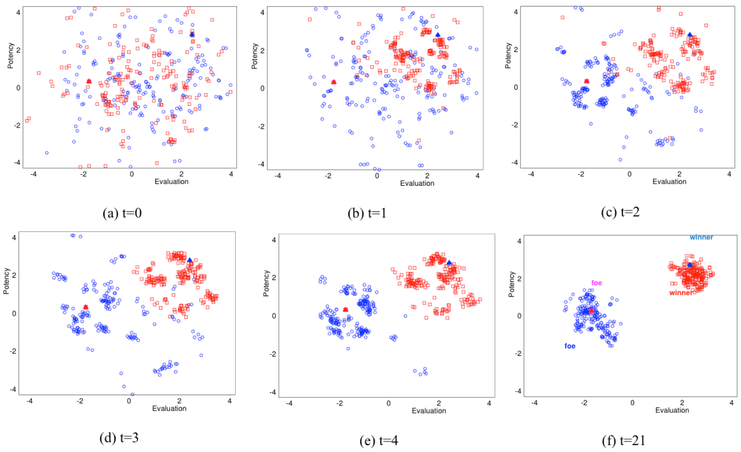
BayesACT inherits from ACT the fundamental identity verification (i.e., deflection minimization) mechanism as the motivating force behind people's social interactions. Deflection (eq. 2) is proposed to be

---

[1] Potsdam University of Applied Sciences, Germany, Institute for Urban Futures. Corresponding author: post@tobiasschroeder.de
[2] University of Waterloo, Canada, Cheriton School of Computer Science
[3] Mount Holyoke College, U.S.A., Department of Sociology

the logarithm of a probabilistic potential, which is proportional to a multidimensional Gaussian probability distribution contingent on the distance between fundamental sentiments $\boldsymbol{f'}$ and transient impressions $\boldsymbol{\tau'}$ at t+1: $\varphi(\mathbf{f}',\boldsymbol{\tau}') \propto e^{-(\mathbf{f}'-\boldsymbol{\tau}')^T \Sigma^{-1} (\mathbf{f}'-\boldsymbol{\tau}')}$ (eq.3). I.e., actions that allow a BayesACT agent to keep transient impressions close to fundamental identity meanings appear more likely.

Simulation of Social Interaction



(a) t=0 (b) t=1 (c) t=2

(d) t=3 (e) t=4 (f) t=21

Figure 2: Simulation of two agents learning their mutual identities.

We have used BayesACT to simulate phenomena of theoretical importance in sociology and social psychology. The underlying ACT has received much empirical support (e.g., Schröder & Scholl, 2009), which also lends validity to BayesACT. Fig. 2 shows a simulation where two BayesACT agents quickly learn their mutual affective identities through interacting with each other. Other simulations show that agents can also learn their own previously unknown identity suggested through the actions of other persons – we believe a possible mechanism of socialization of a child, or simply a new member of an existing group. BayesACT also handles cases well where people‘s actions are governed by multiple, sometimes competing identities. A much-discussed example is a female executive who might struggle with balancing society‘s contradictory expectations towards women versus business leaders.

Discussion and Outlook

We combined POMDP models from artificial intelligence with an established social psychological theory to create BayesACT, an intelligent and adaptive computational model of social interaction. Our hope is to provide a sound theoretical basis to exciting current developments in computational social science – given the popularity of physics metaphors in the field (Pentland, 2014), we might say a “particle physics” approach to social science, to complement current “statistical mechanics” approaches.

References

Ambrasat, J., von Scheve, C., Schauenburg, G., Conrad, M., & Schröder, T. (2014). Consensus and stratification in the affective meaning of human sociality. *Proceedings of the National Academy of Sciences of the U.S.A., 111*, 8001-8006.

Heise, D. R. (1997). *INTERACT: Software and manual*. Available at the Affect Control Theory website: http://www.indiana.edu/~socpsy/ACT/

Heise, D. R. (2007). *Expressive order*. New York: Springer.

Hoey, J., Schröder, T., & Alhothali, A. (2013). Bayesian affect control theory. *Proceedings of the 5th Biannual Humaine Conference on Affective Computing and Intelligent Interaction (ACII-2013)* (pp. 166-172). IEEE Explore.

Osgood, C. E. (1962). Studies on the generality of affective meaning systems. *American Psychologist, 17*, 10-28.

Pentland, A. (2014). *Social physics*. New York: Penguin Press.

Schröder, T. & Scholl, W. (2009). Affective dynamics of leadership: An experimental test of affect control theory. *Social Psychology Quarterly, 72*, 180-197