overlap between the cultures of two groups), and social proximity (intensity of social and kinship ties between two agents or groups). Both of the aforementioned theories identify social proximity as one of the crucial dimensions of proximity. In natural language, we use the vocabularies of social proximity and social groups seamlessly and often interchangeably: I would therefore urge Pietraszewski to develop his formal account of groups by exploring ways to upgrade proximity (understood multidimensionally) from ancillary to central status.

Finally, Pietraszewski's proposed building blocks for a computational account of social groups reveal repeatedly that the problematic of loyalty and disloyalty seems to be inextricably intertwined with how agents make inferences about groups (e.g., sect. 8.2, para. 6 and sect. 8.3, para. 2; sect. 8.4, para. 3). Which brings out the question: What would a computational theory of group loyalty itself look like? The philosophical literature on loyalty may not offer useful starting points because it often frames this topic as "an important area of the normative" (Oldenquist, 1982, p. 173). A more promising starting point seems to be appraising the dynamic relationship between social proximity and loyalty, and thereby pushing the *relational* understanding of social groups one step further.

References

Boschma, R. (2005). Proximity and innovation: A critical assessment. *Regional Studies*, 39(1), 61–74.

Malpas, J. (2012). Putting space in place: Philosophical topography and relational geography. *Environment and Planning D: Society and Space*, 30(2), 226–242.

Oldenquist, A. (1982). Loyalties. *The Journal of Philosophy*, 79(4), 173–193.

Simandan, D. (2016). Proximity, subjectivity, and space: Rethinking distance in human geography. *Geoforum*, 75, 249–252.

Simandan, D. (2018). Competition, contingency, and destabilization in urban assemblages and actor-networks. *Urban Geography*, 39(5), 655–666.

Simandan, D. (2019a). Revisiting positionality and the thesis of situated knowledge. *Dialogues in Human Geography*, 9(2), 129–149.

Simandan, D. (2019b). Competition, delays, and coevolution in markets and politics. *Geoforum*, 98, 15–24.

Simandan, D. (2019c). Iterative lagged asymmetric responses in strategic management and long-range planning. *Time & Society*, 28(4), 1363–1381.

Simandan, D. (2020). Being surprised and surprising ourselves: A geography of personal and social change. *Progress in Human Geography*, 44(1), 99–118.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, 117(2), 440–463.

Shadow banning, astroturfing, catfishing, and other online conflicts where beliefs about group membership diverge

Jordan W. Suchow

School of Business, Stevens Institute of Technology, Hoboken, NJ 07030, USA
jws@stevens.edu; http://suchow.io

Abstract

Drawing from conflicts observed in online communities (e.g., astroturfing and shadow banning), I extend Pietraszewski's theory to accommodate phenomena dependent on the intersubjectivity of groups, where representations of group membership (or beliefs about group membership) diverge. Doing so requires enriching representations to include other agents and their beliefs in a process of recursive mentalizing.

In the target article, Pietraszewski proposes a computational theory of social groups that is at its core *subjective*, defining a group in terms of a single individual's representation of it. However, social groups are not subjective: Consider that a person cannot through their own beliefs unilaterally create or destroy a group, or change an established group's membership. Rather, group membership is *intersubjective*, dependent on representations that are (at least in part) shared among members of the ingroup or outgroup (Dennen & Wieland, 2007; Eden, Jones, Sims, & Smithin, 1981; Matusov, 1996; Stahl, 2016; Zlatev, 2014). The intersubjective nature of groups gives rise to important phenomena in the context of conflict that cannot be explained by Pietraszewski's computational theory because they arise only when people's representations of group membership diverge or are believed to have diverged.

Consider the following examples often observed in conflicts within online communities:

(1) A person wrongly believes they are part of a group, whose members keep up the charade until an ultimate act of (apparent) betrayal that reveals the false belief (e.g., cyberbullying).

(2) A person becomes romantically involved with a defrauder, scammer, troll, or person with some other ulterior motive (e.g., online dating romance scams; catfishing).

(3) A person undermines a group by pretending to be a member of it while covertly acting against its interests (e.g., online sock puppetry and astroturfing; cyber espionage).

(4) Two factions of a group each reject the others' sincerely held beliefs regarding membership in that group, cleaving it in two (e.g., schisms).

(5) An aggressor creates a self-fulfilling prophecy when, failing to distinguish between groups, aggresses against them jointly, causing the groups to merge (e.g., in the emergence of some unity movements).

(6) One person does not believe a particular group exists, whereas another person cherishes their membership in that group (e.g., identity pride and erasure).

(7) A person is unaware of having been banished from a group or silenced within it (e.g., hell banning and shadow banning).

Analyzing any of these phenomena by considering only one individual's representations of group membership would fail to capture their essence. For example, in the case of astroturfing (Leiser, 2016; Sisson, 2017; Zhang, Carpenter, & Ko, 2013), where a purported grass-roots organizer is, in fact, the agent of a sponsor working against the cause, it is not enough for the group or the public to believe the agent is a member of the group. Nor is it enough for the sponsor or its agent to believe they are not a member. Rather, it is the divergence in understanding about the agent across the sponsor, the agent, the public, and

other interested parties, which gives rise to the phenomenon and the harmful consequences to the cause that are associated with it.

Here, I put forward a computational approach that extends Pietraszewski's theory to accommodate phenomena dependent on the intersubjectivity of groups.

The extension begins by zooming out from the focal individual studied by Pietraszewski to consider the representations of all three individuals in the triad (or more generally, any interested parties). Minimally, this is accomplished by endowing each individual with their own representation of the kind put forward by Pietraszewski in terms of which agents will tend to fill the group-constitutive roles. Doing so requires no new computational machinery and permits analysis of diverse phenomena where these representations diverge. For example, in the case of shadow banning, a person is exiled from an online community without their knowledge by a moderator who causes the exiled person's communication to be invisible to other community members (Cole, 2018; waxpancake, 2009). It is common for the shadow-banned individual, the moderator, and other community members to each have their own understanding: The shadow-banned individual believes they are part of the group, the moderator believes they are not, and other members of the group may variously believe the individual is a member, a non-member, or does not even exist. A meaningful description of a group must, therefore, allow expression of divergent representations of group membership.

The extension proceeds by enriching the content of each individual's representation to include the representations of other individuals via a process of recursive mentalizing. Although in the previous step, we endowed each agent with a representation that included other agents filling (or not) group-constitutive roles, but not those other agents' beliefs, in the current step we recurse, enabling each agent to represent other agents' beliefs (Frith & Frith, 2005). At infinite recursion depth, this produces effects of common knowledge (de Freitas, Thomas, DeScioli, & Pinker, 2019; Platow, Foddy, Yamagishi, Lim, & Chow, 2012; Thomas et al., 2016). New computational machinery in the form of recursive mentalizing must be brought to the table, bringing with it the power to model complex social phenomena that depend on misrepresentation and deception, where actions are taken because they are expected to validate another person's wrongly held beliefs or cause them to misinterpret which agents fill group-constitutive roles. Returning to the example of astroturfing, consider that divergence in representations alone is not enough to fully capture its essence – being mistaken as a member of a group is not astroturfing. Rather, the agent must also take an action because they believe it will cause a certain impression in the minds of the public with respect to group membership. A representation of a group in the context of conflict must, therefore, enable individuals to represent the beliefs of others.

Enriching the representation put forward by Pietraszewski to include other agents and their beliefs in a process of recursive mentalizing permits analysis of complex social phenomena that arise from the intersubjectivity of groups.

Conflict of interest. None.

References

Cole, S. (2018). Where did the concept of "shadow banning" come from? Motherboard: Tech by Vice. https://www.vice.com/en/article/a3q744/where-did-shadow-banning-come-from-trump-republicans-shadowbanned.

de Freitas, J., Thomas, K., DeScioli, P., & Pinker, S. (2019). Common knowledge, coordination, and strategic mentalizing in human social life. *Proceedings of the National Academy of Sciences*, 116(28), 13751–13758.

Dennen, V. P., & Wieland, K. (2007). From interaction to intersubjectivity: Facilitating online group discourse processes. *Distance Education*, 28(3), 281–297.

Eden, C., Jones, S., Sims, D., & Smithin, T. (1981). The intersubjectivity of issues and issues of intersubjectivity. *Journal of Management Studies*, 18(1), 37–47.

Frith, C., & Frith, U. (2005). Theory of mind. *Current Biology*, 15(17), R644–R645.

Leiser, M. (2016). AstroTurfing, "CyberTurfing" and other online persuasion campaigns. *European Journal of Law and Technology*, 7(1), 1–27.

Matusov, E. (1996). Intersubjectivity without agreement. *Mind, Culture, and Activity*, 3(1), 25–45.

Platow, M. J., Foddy, M., Yamagishi, T., Lim, L. I., & Chow, A. (2012). Two experimental tests of trust in in-group strangers: The moderating role of common knowledge of group membership. *European Journal of Social Psychology*, 42(1), 30–35.

Sisson, D. C. (2017). Inauthentic communication, organization-public relationships, and trust: A content analysis of online astroturfing news coverage. *Public Relations Review*, 43(4), 788–795.

Stahl, G. (2016). From intersubjectivity to group cognition. *Computer Supported Cooperative Work* (CSCW), 25(4), 355–384.

Thomas, K. A., De Freitas, J., DeScioli, P., & Pinker, S. (2016). Recursive mentalizing and common knowledge in the bystander effect. *Journal of Experimental Psychology: General*, 145(5), 621–629.

waxpancake. (2009). What was the first website to hide troll's activity to everyone but the troll himself? Ask MetaFilter. https://ask.metafilter.com/117775/.

Zhang, J., Carpenter, D., & Ko, M. (2013). Online astroturfing: A theoretical perspective. *AMCIS* 2013.

Zlatev, J. (2014). The co-evolution of human intersubjectivity, morality and language. In *The social origins of language* (pp. 249–266). Oxford University Press.

More than one way to skin a cat: Addressing the arbitration problem in developmental science

Denis Tatone

Department of Cognitive Science, Central European University, 1100 Vienna, Austria
TatoneD@ceu.edu; denis.tatone@gmail.com

Abstract

David Pietraszewski's theory of social groups offers a developmentally plausible account of how we reason about group membership, as it delineates clear boundaries to the hypothesis space that children must navigate. Merits notwithstanding, the account remains silent with respect to the arbitration problem: It does not explain how children can appropriately select among competing frames when interpreting social interactions.

From a developmental standpoint, the main virtue of David Pietraszewski's theory is its ability to deal effectively with the reduction problem. At its core is the claim that the gamut of multi-agent conflict can be decomposed into four types of triadic interactions, each specifying a distinct way by which third parties can be drawn into conflict. By constraining the hypothesis space to a finite repertoire of coalitional schemata, the theory offers a developmentally plausible way in which young learners may infer role assignments across multi-agent configurations. Once the appropriate stance is adopted, attributing group membership becomes an eminently tractable task (cf. Thomsen & Carey, 2013).

Nothing, however, guarantees that third-party interactions end up being interpreted through coalitional lenses. Any given social