# Shared Intentionality and the Representation of Groups; or, How to Build a Socially Adept Robot

Ben Phillips
Arizona State University
bsphilli@asu.edu

(Commentary, forthcoming in *Behavioral and Brain Sciences*)[1]

**Abstract**. Pietraszewski provides a compelling case that representations of certain interaction-types are the "cognitive primitives" that allow all tokens of *group-in-conflict* to be represented within the mind. Here, I argue that the folk concept GROUP encodes shared intentions and goals as more central than these interaction-types, and that providing a computational theory of social groups will be more difficult than Pietraszewski envisages.

In defending his stimulating proposal, Pietraszewski does not focus on the role that theory of mind (or "mindreading") plays in guiding applications of the folk concept GROUP. Importantly, though, there are good reasons for thinking that attributions of shared intentionality have a central role to play.

There is an ongoing debate as to what exactly shared intentionality involves. At a minimum, it requires more than the just the possession of common goals and intentions. For example, people simultaneously jogging along a trail may have the same goal of getting fit, but they are not

[1] Commentary on Pietraszewski, "Towards a computational theory of social groups: A finite set of cognitive primitives for representing any and all social groups in the context of conflict" (*Behavioral and Brain Sciences*: https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/towards-a-computational-theory-of-social-groups-a-finite-set-of-cognitive-primitives-for-representing-any-and-all-social-groups-in-the-context-of-conflict/1C28965AE6F94B9060A20C1DB8EA2A80)

engaging in shared intentionality unless they each harbor an intention of the form, "I intend that *we* jog together" (see Bratman, 1999, chapter 8; Tomasello et al., 2005).

There is direct evidence that shared intentionality drives costly decisions to help ingroup members over outgroup members (McClung et al., 2017). There is also evidence that applications of the folk concept GROUP are guided by attributions of shared intentionality. A number of studies have found that entitativity perception—the tendency to regard some aggregates of people as more "groupish" than others—is mediated by judgments concerning shared intentionality (for some recent discussions, see Phillips, 2021a, 2021b). For example, people's impression that an aggregate of individuals constitutes a genuine group is enhanced when they observe these individuals moving in synchrony (Ip et al., 2006; Lakens & Stel, 2011; Wilson & Gos, 2019). Importantly, these studies suggest that people only tend to regard synchronous movement as a cue for groupishness when they see it as resulting from shared intentionality.

Research into entitativity perception therefore suggests that the folk encode shared intentions and goals as central to the concept GROUP; whereas, they encode certain visible cues, such as synchronous movement, as relatively peripheral (or "ancillary" to use Pietraszewski's term). There are various models of conceptual centrality. The core insight is that a feature, *F*, is more central to a given concept than feature, *G*, if *G* is represented as depending on *F* more than *F* depends on *G* (see Sloman et al., 1998). Thus, the mediational effects outlined above suggest that the folk encode shared intentions and goals as more central to GROUP than synchronous movements, because they represent the latter as causally depending on the former, but not vice versa.

Pietraszewski notes that detecting group-based intentions helps us to predict whether an agent will participate in one of the triadic interactions with certain others (note 10, p. 48). Arguably, though, just as the folk encode shared intentions as more central to GROUP than coordinated movements, they also encode shared intentions as more central than the triadic interaction-types identified by Pietraszewski. In contexts of conflict and cooperation, agents participate in these sorts of interactions precisely because they share certain intentions and goals with fellow group members. For example, when Germany invaded Belgium in 1914, Britain's declaration of war was predictable, in part, because Britain and Belgium shared a suite of goals and intentions (all enshrined in a treaty). By the same token, consider a case in which we are *not* willing to categorize an aggregate as a genuine group. In Pietraszewski's example, some people are waiting for a bus when a motorist throws a stone at one of them. Suppose each person at the bus stop simultaneously hurls a stone back at the driver. The studies outlined above suggest that the folk will not categorize this aggregate of people as a genuine group unless they see them as sharing an intention of the form, "I intend that *we* attack the driver together."

If what I am suggesting is right, this puts pressure on Pietraszewski's claim that representations of triadic interaction-types are the "cognitive primitives" that constitute the folk concept of a group-in-conflict. A number of studies have found that when one feature is encoded as more central to a given concept than another feature, the former is a stronger determinant of categorization decisions (e.g., see Ahn et al., 2000). Thus, according to the alternative hypothesis on offer, applications of GROUP—in contexts of both cooperation and conflict—are guided by representations of certain triadic interactions, but only insofar as these interactions serve as cues for shared intentionality. This suggests that the task of constructing a computational theory of group cognition is more difficult than Pietraszewski envisages, for it will require no less than a

computational theory of mindreading. To put it another way, suppose we were to build Pietraszewski's robot, which can navigate the social world by deploying representations of groups. If our robot cannot attribute shared intentionality to others, it will be left in the dust by its (socially adept) human counterparts.

To conclude, it is worth noting that a full-blown capacity for shared intentionality takes time to develop in humans (Tomasello et al., 2005)—presumably, the capacity to *attribute* shared intentionality takes considerably longer. Nonetheless, there is evidence that infants can track some of the interaction-types identified by Pietraszewski (e.g., see Ting et al., 2019). Similarly, the capacity for shared intentionality appears to be largely absent in nonhuman primates (Tomasello et al., 2005). Regardless, some nonhuman primates, such as baboons, are able to track groups as they fluctuate across episodes of conflict (e.g., see Cheney & Seyfarth, 2007). It is possible that young children, as well as some nonhuman primates, track groups-in-conflict by detecting triadic interactions of the sort identified by Pietraszewski. If so, Pietraszewski's account may describe an early-developing, phylogenetically ancient, system for detecting groups-in-conflict. This system may output a relatively "thin" concept of groups the possession of which does not require an understanding of shared intentionality. Instead, possessing the thin concept might only require an agent to detect triadic interactions by utilizing low-level perceptual cues (e.g. visible instances of hitting, chasing, etc.). In contrast, the "thick" concept of a group that adults deploy in central cognition appears to encode shared intentions as central, and triadic interactions as more peripheral.

**References**


Ahn, W., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology*, 361–416.

Bratman, M. (1999). *Faces of intention: Selected essays on intention and agency*, Cambridge: Cambridge University Press.

Cheney, D. L., & Seyfarth, R. M. (2007). *Baboon metaphysics: The evolution of a social mind*. The University of Chicago Press.

Ip, G. W. M., Chiu, C. Y., & Wan, C. (2006). Birds of a feather and birds flocking together: Physical versus behavioral cues may lead to trait- versus goal-based group perception. *Journal of Personality and Social Psychology,* 90, 368–381.

Lakens, D., & Stel, M. (2011). If they move in sync, they must feel in sync: Movement synchrony leads to attributions of rapport and entitativity. *Social Cognition*, 29, 1–14.

McClung, J., Placì, S., Bangerter, A., Clément, F., & Bshary, R. (2017). The language of cooperation: shared intentionality drives variation in helping as a function of group membership. *Proceedings of the Royal Society B: Biological Sciences, 284*.

Phillips, B. (2021a). The roots of racial categorization. *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-021-00525-w

Phillips, B. (2021b). Entitativity and implicit measures of social cognition. *Mind & Language*. https://doi.org/10.1111/mila.12350

Sloman, S. A., Love, B. C., & Ahn, W-K. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, 22(2), 189–228.

Ting, F., He, Z., & Baillargeon, R. (2019). Toddlers and infants expect individuals to refrain from helping an ingroup victim's aggressor. *PNAS*, 116(13), 6025–6034.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences*, 28(5), 675–691. doi: 10.1017/S0140525X05000129.

Wilson, S., & Gos, C. (2019). Perceiving social cohesion: Movement synchrony and task demands both matter. *Perception*, 48(4), 316–329.