1. **Target Article:** "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" by David Pietraszewski

2. **Word count:**
   - Abstract: 58
   - Main text: 991
   - References: 654
   - Total (=all of the above + addresses etc): 1.912

3. **Commentary title:**
   A neuroscientific perspective on the computational theory of social groups

4. **Authors of this commentary:** Marco K. Wittmann[1], Nadira S. Faber[2,3], Claus Lamm[4,5]

5. **Institutions:**
   1 Wellcome Centre for Integrative Neuroimaging (WIN), Department of Experimental Psychology, University of Oxford, Oxford, UK
   2 Department of Psychology, University of Exeter, Exeter, UK
   3 Uehiro Centre for Practical Ethics, University of Oxford, Oxford, UK
   4 Faculty of Psychology, University of Vienna, Austria
   5 Vienna Cognitive Science Hub, University of Vienna, Austria

6. **Full institutional mailing address**
   - MW: Department of Experimental Psychology, University of Oxford, Tinsley Building, Mansfield Road, OX1 3SR, Oxford, United Kingdom
   - NF: College of Life and Environmental Sciences, University of Exeter, Washington Singer Building, Exeter, EX4 4QG, United Kingdom
   - CL: Department of Cognition. Emotion, and Methods in Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria

7. **Institutional telephone numbers**
   - MW: +44 7522513021
   - NF: +44 1392 724633
   - CL: +43 1427747130

8. **Emails:**
   - MW: marco.k.wittmann@gmail.com
   - NF: nadira.faber@gmail.com
   - CL: claus.lamm@univie.ac.at

9. **Homepages:**
   - MW: https://www.psy.ox.ac.uk/team/marco-wittmann
   - NF: http://nadirafaber.com/
   - CL: scan.psy.univie.ac.at

10. **Abstract: (58 words)**

We welcome a computational theory on social groups, yet we argue it would benefit from a broader scope. A neuroscientific perspective offers the possibility to disentangle which computations employed in a group context are genuinely social in nature. Concurrently, we emphasize that a unifying theory of social groups needs to additionally consider higher-level processes like motivations and emotions.

11. **Main text (991 words)**

Social groups are studied in a variety of fields in the behavioural and brain sciences, mirroring their central role in human and non-human societies. As three researchers studying groups and their individuals from different perspectives (e.g., Faber et al., 2017; Lockwood et al., 2018; Rütgen et al., 2015), we very much welcome an attempt for a unifying framework on groups. A shared conceptualisation of groups would indeed allow researchers to bridge gaps between disciplines and could be extended to core group topics beyond conflict, like cooperation and coordination problems. However, to really be unifying, we argue a broader scope than the one presented in the target article would be needed. Such a broader scope should in particular consider i) a neuroscientific angle that ii) incorporates motivations and emotions.

The target article dissects the cognitive mechanisms allowing individuals to navigate social groups. It argues that group computations arise from considering *costs* that one agent imposes on another. Representing costs in group settings presses basic *relational primitives* into action. Relational primitives are the computational scaffold that enables us to think of the relationships between individuals in a group. However, we know that the computational architecture proposed, should it exist, must arise from neural processes. In neuroscience, we routinely consider learning from rewards and losses using computational approaches (Sutton & Barto, 2018). Dr. Pietraszewski's framework suggests that this quantitative framework of learning and decision-making might in fact be an adequate route to understanding the relational primitives at the heart of social group membership computations.

Neural networks in prefrontal cortex and interconnected subcortical regions have been identified that determine how people weigh costs and benefits in non-social settings (Basten et al., 2010; Klein-Flugge et al., 2016). One intriguing implication from Dr. Pietraszewski's framework is that the same machinery that has evolutionary arisen to guide cost-benefit decisions in non-social domains may be used to instantiate relational primitives in group contexts. Should this be the case, then this raises a fundamental question about the social nature of group representations. This question is: to what degree are the component processes underlying our ability to navigate social groups *specifically social*? Neuroscientifically, it is likely that many component processes underlying social cognition are shared between social and non-social domains (Wittmann et al., 2018). For instance, the amygdala and the lateral orbitofrontal cortex are important for learning from rewards and associating them with specific stimuli (Murray & Rudebeck, 2018). These computations might underlie some of their contributions to social cognition (Munuera et al., 2018; Sliwa & Freiwald, 2017). However, other brain regions have been more specifically linked to our ability to think about other people and infer their beliefs like the temporoparietal junction and dorsomedial prefrontal cortex (Lamm et al., 2016; Saxe, 2006) and it is possible that at least some of the computations performed in these regions are specifically needed for navigating social environments.

Applying this perspective to group computations, we might speculate that group cognition relies similarly on a mixture of social and non-social mechanisms. Following the rationale of the target article, it might draw particularly on the ability to compute rewards and costs that ensue from the actions of ourselves and others. In addition, an ability that seems particularly pertinent for group cognition may be the ability to infer relationships between multiple agents. Recent studies have explored the specific computations via which the brain computes relationships between objects and even abstract concepts (Behrens et al., 2018). This might be central for instantiating the relational primitives underpinning our ability to think about social groups. Nevertheless, despite the potential existence of such a domain general mechanism for forming relationships, it is possible that in particular dorsomedial prefrontal cortex, one of the most prominent regions in social cognition research, is specifically important for forming relationships between social agents (Izuma & Adolphs, 2013). Dorsomedial prefrontal cortex represents self and others in an interdependent frame of reference and appears to be causally important to separate out self and other related information (Wittmann et al., 2016, 2021).

Therefore, by employing a neurocomputational perspective, we may gain more precise information on which aspects of group representations may be genuinely social. However, by no means do we suggest being "reductionist" in a computational theory of groups. In fact, there are specific component mechanisms involved in social processes that may not fit in the categories proposed in the target article. Specifically, *motivations* and *emotions* are crucial for different aspects of group functioning – generally, and when it comes to conflict within and between groups. For example, a key social motivation for group functioning is the desire to build a positive reputation in the eyes of other people (cf. Faber et al., 2016). This motivation critically shapes prosocial behaviour (Ariely et al., 2009; Nowak & Sigmund, 2005) and also group-decision making (De Dreu et al., 2008; Faulmüller et al., 2012). Regarding emotions, empathy is an exemplary social emotion that is crucial. Empathy allows us to understand each other from a first-person experiential perspective (e.g. Lamm et al., 2019). While this can have beneficial effects on prosocial behaviours, such as intergroup cooperation, there is also a potential "dark side" to empathy. This may come out in competitive contexts (when we use our understanding of others to better compete against them), as well as when considering that empathy and the ensuing behaviour is prone to in-group biases (Bloom, 2017). While we have only started to understand the neurocomputational processes that underpin motivations and emotions on an individual level, even less is known on how a group context may alter or amplify these processes – or vice versa, how these processes determine membership in groups.

In summary, neuroscience provides a complementary approach that may enrich the proposed computational theory of social groups. It may help determine the precise – social and non-social - component mechanisms underlying group computations and provide a scaffold to incorporate additional component processes relating to motivations and emotions as well as their interaction into a computational theory of social groups.

12. **Conflict of interest statement:** The authors declare no competing interests.

13. **Funding Statement:** This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

14. **Alphabetical reference list**

Ariely, D., Bracha, A., & Meier, S. (2009). Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *American Economic Review*, *99*(1), 544–555. https://doi.org/10.1257/aer.99.1.544

Basten, U., Biele, G., Heekeren, H. R., & Fiebach, C. J. (2010). How the brain integrates costs and benefits during decision making. *Proc Natl Acad Sci U S A*, *107*(50), 21767–21772. https://doi.org/10.1073/pnas.0908104107

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron*, *100*(2), 490–509. https://doi.org/10.1016/j.neuron.2018.10.002

Bloom, P. (2017). Empathy and Its Discontents. *Trends in Cognitive Sciences*, *21*(1), 24–31.

De Dreu, C. K. W., Nijstad, B. A., & van Knippenberg, D. (2008). Motivated Information Processing in Group Judgment and Decision Making. *Personality and Social Psychology Review*, *12*(1), 22–49. https://doi.org/10.1177/1088868307304092

Faber, N. S., Häusser, J. A., & Kerr, N. L. (2017). Sleep Deprivation Impairs and Caffeine Enhances My Performance, but Not Always Our Performance: How Acting in a Group Can Change the Effects of Impairments and Enhancements. *Personality and Social Psychology Review*, *21*(1), 3–28. https://doi.org/10.1177/1088868315609487

Faber, N. S., Savulescu, J., & Van Lange, P. A. (2016). Reputational concerns as a general determinant of group functioning. *Behavioral and Brain Sciences*, *39*. https://doi.org/10.1017/S0140525X15001363

Faulmüller, N., Mojzisch, A., Kerschreiter, R., & Schulz-Hardt, S. (2012). Do you want to convince me or to be understood? Preference-consistent information sharing and its motivational determinants. *Personality and Social Psychology Bulletin*, *38*(12), 1684–1696. https://doi.org/10.1177/0146167212458707

Izuma, K., & Adolphs, R. (2013). Social manipulation of preference in the human brain. *Neuron*, *78*(3), 563–573. https://doi.org/10.1016/j.neuron.2013.03.023

Klein-Flugge, M. C., Kennerley, S. W., Friston, K., & Bestmann, S. (2016). Neural Signatures of Value Comparison in Human Cingulate Cortex during Decisions Requiring an Effort-Reward Trade-off. *Journal of Neuroscience*, *36*(39), 10002–10015. https://doi.org/10.1523/JNEUROSCI.0292-16.2016

Lamm, C., Bukowski, H., & Silani, G. (2016). From shared to distinct self–other representations in empathy: Evidence from neurotypical function and socio-cognitive disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150083. https://doi.org/10.1098/rstb.2015.0083

Lamm, C., Rütgen, M., & Wagner, I. C. (2019). Imaging empathy and prosocial emotions. *Neuroscience Letters*, *693*, 49–53. https://doi.org/10.1016/j.neulet.2017.06.054

Lockwood, P. L., Wittmann, M. K., Apps, M. A. J., Klein-Flügge, M. C., Crockett, M. J., Humphreys, G. W., & Rushworth, M. F. S. (2018). Neural mechanisms for learning self and other ownership. *Nature Communications*, *9*(1), 4747. https://doi.org/10.1038/s41467-018-07231-9

Munuera, J., Rigotti, M., & Salzman, C. D. (2018). Shared neural coding for social hierarchy and reward value in primate amygdala. *Nature Neuroscience*. https://doi.org/10.1038/s41593-018-0082-8

Murray, E. A., & Rudebeck, P. H. (2018). Specializations for reward-guided decision-making in the primate ventral prefrontal cortex. *Nature Reviews Neuroscience*, *19*(7), 404–417. https://doi.org/10.1038/s41583-018-0013-4

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*(7063), 1291–1298. https://doi.org/10.1038/nature04131

Rütgen, M., Seidel, E.-M., Silani, G., Riečanský, I., Hummer, A., Windischberger, C., Petrovic, P., & Lamm, C. (2015). Placebo analgesia and its opioidergic regulation suggest that empathy for pain is grounded in self pain. *Proceedings of the National Academy of Sciences*, *112*(41), E5638–E5646. https://doi.org/10.1073/pnas.1511269112

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, *16*(2), 235–239. https://doi.org/10.1016/j.conb.2006.03.001

Sliwa, J., & Freiwald, W. A. (2017). A dedicated network for social interaction processing in the primate brain. *Science*, *356*(6339), 745–749. https://doi.org/10.1126/science.aam6383

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An introduction*. MIT Press.

Wittmann, M. K., Kolling, N., Faber, N. S., Scholl, J., Nelissen, N., & Rushworth, M. F. (2016). Self-Other Mergence in the Frontal Cortex during Cooperation and Competition. *Neuron*, *91*(2), 482–493. https://doi.org/10.1016/j.neuron.2016.06.022

Wittmann, M. K., Lockwood, P. L., & Rushworth, M. F. S. (2018). Neural Mechanisms of Social Cognition in Primates. *Annual Review of Neuroscience*, *41*, 99–118. https://doi.org/10.1146/annurev-neuro-080317-061450

Wittmann, M. K., Trudel, N., Trier, H. A., Klein-Flügge, M. C., Sel, A., Verhagen, L., & Rushworth, M. F. S. (2021). Causal manipulation of self-other mergence in the dorsomedial prefrontal cortex. *Neuron*, S0896627321003809. https://doi.org/10.1016/j.neuron.2021.05.027