Social Motivations - An Overlooked Topic in the Intrinsic Motivation Landscape?

Frank Förster
Robotics Research Group
School of Physics Engineering and Computer Science
University of Hertfordshire
f.foerster@herts.ac.uk

Abstract

The following abstract is somewhat of a provocation piece where I describe what I perceive to be an important omission from the intrinsic motivation research landscape: the modelling of social motivations. I will shortly describe what social motivations are, point out the difference between intrinsic motivations as modelled in machine learning and social motivation, and will hint how the integration of social motivation in interactive systems might help to address important issues in human-oriented machine learning[1].

1 Social Motivations in Psychology and Cognitive Science

Social motivations appear to be somewhat of an overlooked topic not just in the technical[2] literature on intrinsic motivation, but also in the psychological literature on this topic. Morris et al. (2022), for example, giving an otherwise detailed overview of the psychological research of intrinsic and extrinsic motivations, only mention them in the context of social play. *Social motivation* is an umbrella term for number of affective and cognitive factors that make social experiences in general and joint action in particular *rewarding in their own right*, that is, independent of any goals that the participants of that interaction may have had, either individually or jointly, at the outset of the interaction (Godman, 2013). Godman cites three factors falling under this category, namely, (1) the hormone Oxytocin, that she suggests is involved in rewarding non-aggressive forms of human joint action; (2) social emotions such as love and trust; and (3) empathy. However, for the purpose of this abstract the details are not overly important. What is important is the fact that there is an intrinsic motivation driving human conduct, that is not goal-oriented, and that shapes many of our social interactions. Experimental support for the existence - and effect - of social motivations is provided by Carr and Walton (2014). Carr and Walton show how mere cues of being involved in some form of joint engagement, even if not physically co-present, renders the involved more interested in challenging tasks, increases the joy of performing the task as well as their motivation to take on such tasks in the future.

2 Intrinsic Motivations in Machine Learning

Intrinsic motivations are frequently located in one of three classes (Aubret et al., 2023) . The first class is based on prediction, frequently termed *surprise*, an emotion that captures the mismatch between an agent's expectation of what is ought to occur and what actually occurs (Barto et al., 2013). The second class is based on *novelty*, which arises when an item of knowledge is either new (= not present in the agent's memory) or significantly different from the other knowledge in its memory (ibid.). The third class is based on the acquisition of *competence* or *skill* (Baldassarre and Mirolli, 2012).

Noticeably, all of these three classes of intrinsic motivation "aim", more or less directly, for the agent to learn something, either by learning something new in terms of knowledge or competence, or by learning to improve its capability in predicting future states of its environment. This is hardly surprising given that their technical formulations arose from the the field of machine *learning*. It is nevertheless important to point out this particular aspect of these types of intrinsic motivations as it is not obvious whether social motivations have this orientation towards learning - at least not in the direct manner that the other three types have. Their

---

[1]With 'human-oriented machine learning' I mean to refer to the application of machine learning in human-robot and human-computer interaction, including its application in dialogue systems.

[2]'technical' as in machine learning

purpose appears to be the enabling and maintaining of joint engagements in general, and joint actions in particular, which are of particular importance in human-robot interaction.

3 Current Issues in Human-Oriented Machine Learning

Belhassein et al. (2022) provide an overview of the current challenges in human-robot joint action, which probably encompasses the majority of interactions in human-robot interaction (HRI), including speech-centered ones. As they attest, many interactions with robots are still frustrating or off-putting and end in failure (cf. Förster et al. 2023; Förster et al. 2023). Reason for such failure include a high degree of uncertainty with regard to the common ground between robot and human interactor and poor coordination during joint action, including the inability to recognize interactional failure. Belhassein et al. argue that communicating the commitment to joint act on part of the robot and the expression of (shared) emotion may contribute to tackling these issues.

There are a good number of communication strategies, including emotional displays, that could be deployed for this purpose (see Belhassein et al. 2022 for an overview). However, rather than hand-crafting such strategies, not an uncommon practice in HRI, it would be beneficial to design a more principled method for deploying communicative acts as well as for keeping track of communicative expression on the human interactant's part.

The question I would like to raise with this abstract then is, whether (1) a formal formulation of social motivation could or should be the kind of principled method alluded to here, and (2) how this formulation may look like in view of the differences between social and other types of intrinsic motivations, especially the difference in terms of orientation towards learning.

If such a formulation should be achieved, I suggest that I could be used alongside other intrinsic motivations in frameworks such as multi-object reinforcement learning (cf. Palm and Schwenker 2019).

Acknowledgments

References

Arthur Aubret, Laetitia Matignon, and Salima Hassas. 2023. An Information-Theoretic Perspective on Intrinsic Motivation in Reinforcement Learning: A Survey. *Entropy*, 25(2):327.

Gianluca Baldassarre and Marco Mirolli. 2012. Intrinsically Motivated Learning Systems: An Overview. In *Intrinsically Motivated Learning in Natural and Artificial Systems*, pages 1–14. Springer.

Andrew Barto, Marco Mirolli, and Gianluca Baldassarre. 2013. Novelty or Surprise? *Frontiers in Psychology*, 4:907.

Kathleen Belhassein, Víctor Fernández-Castro, Amandine Mayima, Aurélie Clodic, Elisabeth Pacherie, Michèle Guidetti, Rachid Alami, and Hélène Cochet. 2022. Addressing joint action challenges in hri: Insights from psychology and philosophy. *Acta Psychologica*, 222:103476.

Priyanka B. Carr and Gregory M. Walton. 2014. Cues of working together fuel intrinsic motivation. *Journal of Experimental Social Psychology*, 53:169–184.

Frank Förster, Marta Romeo, Patrick Holthaus, Birthe Nesset, Maria J. Galvez Trigo, Christian Dondrup, and Joel E. Fischer. 2023. Working with troubles and failures in conversation between humans and robots. In *Proceedings of the 5th International Conference on Conversational User Interfaces*, CUI '23, New York, NY, USA. Association for Computing Machinery.

Frank Förster, Marta Romeo, Patrick Holthaus, Luke J. Wood, Christian Dondrup, Joel E. Fischer, Farhana Ferdousi Liza, Sara Kaszuba, Julian Hough, Birthe Nesset, Daniel Hernández García, Dimosthenis Kontogiorgos, Jennifer Williams, Elif Ecem Özkan, Pepita Barnard, Gustavo Berumen, Dominic Price, Sue Cobb, Martina Wiltschko, and 11 others. 2023. Working with troubles and failures in conversation between humans and robots: workshop report. *Frontiers in Robotics and AI*, Volume 10 - 2023.

Marion Godman. 2013. Why we do things together: The social motivation for joint action. *Philosophical Psychology*, 26(4):588–603.

Laurel S Morris, Mora M Grehl, Sarah B Rutter, Marishka Mehta, and Margaret L Westwater. 2022. On what motivates us: a detailed review of intrinsic v. extrinsic motivation. *Psychological Medicine*, 52(10):1801–1816.

Günther Palm and Friedhelm Schwenker. 2019. Artificial Development by Reinforcement Learning Can Benefit From Multiple Motivations. *Frontiers in Robotics and AI*, 6:6.