

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

A computational theory of learning moral weights

Permalink

<https://escholarship.org/uc/item/08q3s787>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Spiteri, Glen W

Maier, Maximilian

Lieder, Falk

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

A Computational Theory of Learning Moral Weights

Glen Spiteri^{*1}, Maximilian Maier², and Falk Lieder¹

¹Department of Psychology, University of California, Los Angeles, 7525 Pritzker Hall, Los Angeles, CA 90095

²Warwick Business School, University of Warwick, Scarman Rd, Coventry, CV4 7AL

Abstract

What determines whose welfare people consider in their moral decisions? We propose that people learn entity-specific moral weights through reinforcement learning (RL), where decision outcomes provide the learning signal. We formalize this in a computational model in which agents update their moral weights for different stakeholders based on whether considering those stakeholders' welfare led to better- or worse-than-expected outcomes. To test this model, we simulate agents learning in market environments (which reward cooperation with strangers) versus non-market environments (which reward exploitation). We show that this mechanism is sufficient to explain two empirical phenomena linking market integration to prosociality: (1) cross-cultural variation in dictator game offers across small-scale societies, and (2) within-cultural variation in lost-letter return rates across 188 Italian municipalities. Together, our simulation results suggest that updating moral weights via RL may be an important mechanism of moral change at both the individual and the societal level.

Keywords: cooperation; impartial prosociality; morality; RL; moral learning; social decision-making; cultural differences

*Correspondence: glenwspiteri@g.ucla.edu