Unveiling the Influence of Automation on Morality: A Multi-modal Examination of the Trolley Problem

Federico Cassioli[1,2, 3*], Davide Crivelli[3], Michela Balconi[3]

[1.] Service de Psychologie cognitive et Neuropsychologie, University of Mons, Mons, 7000, Belgium
[2] Centre for Interdisciplinarity research in psychophysiology and electrophysiology of cognition, University of Mons, Mons, 7000, Belgium
[3] International research center for Applied Cognitive Neuroscience (IrcCAN), Department of Psychology, Università Cattolica del Sacro Cuore Milan, Italy

*Federico Cassioli, Ph.D.,
federico.cassioli@umons.ac.be
Place du Parc, 18, University of Mons, Mons, 7000, Belgium

Incorporating automated technology into decision-making processes raises ethical concerns. This study assessed participants' (n=34; M_{age} = 24.6±3, 21–35; n_{female}= 64.7%) evaluations of morality, consciousness, responsibility, intentionality, and emotional aspects regarding agents (human, automated-machine) and behaviors (action, inaction, usually labeled as “utilitarian” “deontological”) involved in a modified version of the Trolley Problem (12-trials). Reaction times (RTs), electroencephalography frequencies (10–10 IS; 32-channel SynAmps system, Scan 4.2, Compumedics Neuroscan Inc.), autonomic (heart-rate-variability, HRV), and the Balanced-Emotional-Empathy-Scale (BEES) data were collected.

The linear-mixed-models (LMMs) analysis revealed that participants applied different moral norms, perceiving humans as possessing higher levels of morality, responsibility, and consciousness, regardless of the performed behavior. The BEES positively correlated with the emotional impact’s evaluation in human conditions and not in machines. Regarding brain data, when assessing the agent’s consciousness, we observed increased F7 and F8 power within the beta frequency window (13-30 Hz) for machines compared to humans in the inaction conditions. The same pattern was detected in human dilemmas, with higher activity in the action condition. This pattern potentially reflects ventrolateral-prefrontal-cortex (VLPFC) activity, a region associated with conflict control and uncertainty. Furthermore, when evaluating the agent’s intentionality and the derived emotional impact we observed a general gamma (30.5-50 Hz) synchronization, often occurring in response to negative emotional content. Lastly, reduced variability in machine action and human inaction was found. Decreased HRV during decision-making is frequently associated with uncertainty and challenges in emotion regulation.

These findings suggest an imbalance in judgments influenced by the agent involved in moral dilemmas.