

Minha Lee  
Lily Frank  
Wijnand IJsselsteijn

Title:

**Whose emotions are second-order artificial emotions?**

**Intimate feelings fostered with and by technology**

In affective computing, machines can detect people's emotions through computer vision or speech processing, but they can also convey or mimic human emotions. A robot can frown, a smart speaker can sound happy, and a chatbot can send emojis. Yet as these various AI systems become more technically capable, simple expressions like smiles or emojis are likely to evolve into complex emotional displays, e.g., gratitude or grief. Even if a machine never truly *feels* grief like a person can, humans may still attribute complex emotional experiences to artificial agents. **Artificial emotions (AE)** are then an extension of or projections of our own emotions.

Emotional responses, such as reactive attitudes, are central to human relationships. When we feel grateful to or resentful of someone, we often praise them or assign them blame (Strawson, 1974); emotions like resentment often signal that one person has experienced a violation of expected mutual respect (Strawson 1974; Darwall, 2006). Further, humans employ first and second order emotions. Second-order emotions are emotions about emotions (Alfano, 2017): One might feel surprised at oneself (second-order emotion) by one's own resentment (first-order emotion) towards someone.

However, future AE can influence people's real emotions, e.g., sympathy (second-order emotion) towards expressed grief (first-order artificial emotion) if artificial agents express a full range of emotions from happiness through a smiling emoji to grief with virtual tears. AE then challenge assumptions about who *has* an emotion. Blurred emotional boundaries arise when we interact with emotionally intimate machines of the future, with many ethical implications.

Bibliography:

Alfano, Mark. *Moral psychology: An introduction*. John Wiley & Sons, 2016.

Darwall, Stephen L. *The second-person standpoint: Morality, respect, and accountability*. Harvard University Press, 2006.

Strawson, Peter Frederick. *Freedom and resentment and other essays*. Routledge, 2008 (1974).