

Abstract

Robots are increasingly expected to act as social partners beyond functional assistance to shaping behavior, motivation, and relational dynamics. While current human–robot interaction (HRI) research has largely emphasized positive communication strategies such as praise, encouragement, and empathy, human societies frequently rely on negative affective strategies—criticism, irony, and admonishment—as constructive tools to motivate improvement and maintain order. Without the ability to employ such strategies appropriately, robots risk being limited to overly positive roles, reducing their effectiveness and acceptance in real-world social contexts.

This thesis addresses this gap by investigating how robots can employ negative affective strategies in ways that are positive, context-sensitive, and socially appropriate. We argue that subtle, appropriate negative affective strategies—rather than overt or indiscriminate negativity—can encourage effort, preserve rapport, and help balance hierarchical dynamics. Through three studies, we systematically examine the effects of negative affective strategies, formalize design rules for their appropriate use, and implement these insights into a teleoperation support system for hierarchical interactions.

In our first study (Chapter 3), we investigate how diverse strategies, including praise, encouragement, empathy, irony, and disappointed remarks affect users in an exercise partner scenario. By analyzing 80 robot utterances across multiple affective categories, we find that the impact of negative affective strategies is highly context-dependent: irony can energize users in low-arousal positive states but discourage them under fatigue. This study highlights that not all negativity is harmful and shows that, under specific conditions, mild negative affective expressions can even outperform positive ones in motivating effort and sustaining engagement.

In our second study (Chapter 4), we focus on irony as a form of negative affective feedback. Drawing on interviews with 25 irony receivers, we identify factors that distinguish constructive from harmful uses, such as user mood, energy level, and task efficiency. Based on these insights, we design context-sensitive rules for irony use in robots and validate them in an experiment with 22 participants. Results show that the rule-based irony system significantly improves exercise duration, user enjoyment, and perceptions of being understood, compared to indiscriminate irony. This study establishes concrete guidelines for how robots can apply irony effectively while safeguarding users' psychological well-being.

In our third study (Chapter 5), we extend these insights to the broader challenge of power-forward relationships in remote interactions. Situating our research in sports-club contexts, which naturally provide senior–junior hierarchies, we design a teleoperation support system that assists high-power individuals in guiding juniors via avatar robots. The system integrates operator intent recognition with exerciser state monitoring to implement three core policies: proactive guidance during silence or neglect, early conflict risk management, and autonomy-preserving design. In an experiment with 17 real senior–junior pairs, the system reduced operator workload, improved communication strategies, and enhanced juniors’ acceptance and enjoyment, while exercise performance remained comparable across conditions.

By addressing the constructive use of negative affective strategies from multiple perspectives—effects, rules, and real-world applications—this thesis presents a unified framework for enabling robots to use negativity in socially appropriate ways. These insights contribute to advancing HRI from one-dimensional positivity toward more nuanced, context-aware interaction strategies, supporting not only motivation and engagement but also the ethical management of power in hierarchical relationships.