Trust Resilience in Pedagogical Agents: Will Anthropomorphism Help Against Trust Decline?

Fransisca Hapsari
Technische Universität Darmstadt, Darmstadt, Germany

Ronja Stemler
Technische Universität Darmstadt, Darmstadt, Germany

Stephanie Pieschl
Technische Universität Darmstadt, Darmstadt, Germany

Abstract

Trust is an important factor in interaction with automated agents. This study tracks users' trust calibration to automated agents in a vocabulary learning task. We hypothesize that trust declines as agent reliability declines and that anthropomorphism should buffer against this decline. Replicating de Visser et al. (2016), 60 participants guessed the meaning of 96 foreign words in a 4x4x2 mixed experiment. In each trial, they guessed alone, then got an agent's recommendation and gave trust judgments, and made a final decision. Four pedagogical agents varying in anthropomorphism (within-subject: human, robot, smart speaker, computer) recommended answers with decreasing reliability (within-subject: 100%, 67.5%, 50%, 0%). Furthermore, participants either did or did not watch an introductory video about the agents (between-subject). Behavioral and judgment data were analysed via mixed-effects models and ANOVAs. Two-way interaction shows that trust declined differently in various agents, but there is little evidence supporting trust resilience in any agent.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*