

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

A Drift Diffusion Model of Trust in Human-Robot Interaction

Permalink

<https://escholarship.org/uc/item/9918p1wg>

Journal

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

Authors

Saad, Laura

Roesler, Eileen

Phillips, Elizabeth

et al.

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 Drift Diffusion Model of Trust in Human-Robot Interaction

Laura Saad

NRC Postdoctoral Associate
U.S. Naval Research Laboratory, Washington, DC, USA

Eileen Roesler

George Mason University, Fairfax, VA, USA

Elizabeth Phillips

George Mason University, Fairfax, VA, USA

J. Gregory Trafton

U.S. Naval Research Laboratory, Washington, DC, USA

Abstract

Trust in human-robot interaction (HRI) is widely studied, yet the cognitive mechanisms underlying it are unclear. We propose the Expectation-Trust Model (ETM), a process-level account of trust grounded in expectations and formalized within a drift diffusion modeling framework. In ETM, trust is represented as an evidence accumulation process in which existing knowledge about robots determines the starting point, while deviations from expectations influence the drift rate. Positive drift reflects behaviors that exceed expectations, such as excellent performance, whereas negative drift reflects expectation violations, such as robot errors. We apply ETM to simulate trust dynamics from two published datasets, capturing empirical patterns including trust formation, violation, and repair. Differences in trust were explained by variation in starting point and drift rate corresponding to robot design, performance, and trust repair strategies. ETM provides a parsimonious mechanistic explanation for trust in HRI and generates precise, testable predictions about how expectations shape trust.