

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Confidence phenotypes: a unified computational account of value and decision certainty in reinforcement learning

Permalink

<https://escholarship.org/uc/item/2k75s2qr>

Journal

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

Authors

Comay, Nicolas A
Solovey, Guillermo
Barttfeld, Pablo

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

Confidence phenotypes: a unified computational account of value and decision certainty in reinforcement learning

Nicolás A. Comay

Instituto de Investigaciones Psicológicas, UNC-CONICET, Argentina.

Guillermo Solovey^{1,2,3}

¹Instituto de Cálculo, UBA-CONICET, Argentina.

²Laboratorio de Neurociencia, Universidad Torcuato Di Tella, Argentina.

³Escuela de Negocios, Universidad Torcuato Di Tella, Argentina.

Pablo Barttfeld

Instituto de Investigaciones Psicológicas, UNC-CONICET, Argentina.

Abstract

Prior work in human reinforcement learning has distinguished between value confidence (certainty in value estimates) and decision confidence (certainty that a choice is correct), but how these signals are computed and interact during learning has not been directly tested. Here, we combine two new experiments with previously published datasets to evaluate competing computational hypotheses. We show that value confidence follows a Bayesian computation reflecting the precision of value estimates and adaptively regulates the exploration-exploitation trade-off. In contrast, decision confidence systematically departs from Bayesian predictions, particularly on incorrect trials. A hybrid model combining Bayesian probability of being correct with the precision of the decision variable provides a better account of decision confidence. Importantly, individual differences in the weighting of these components predict both task performance and metacognitive accuracy. Together, our results offer a unified computational account of how distinct forms of confidence shape learning and decision-making under uncertainty.