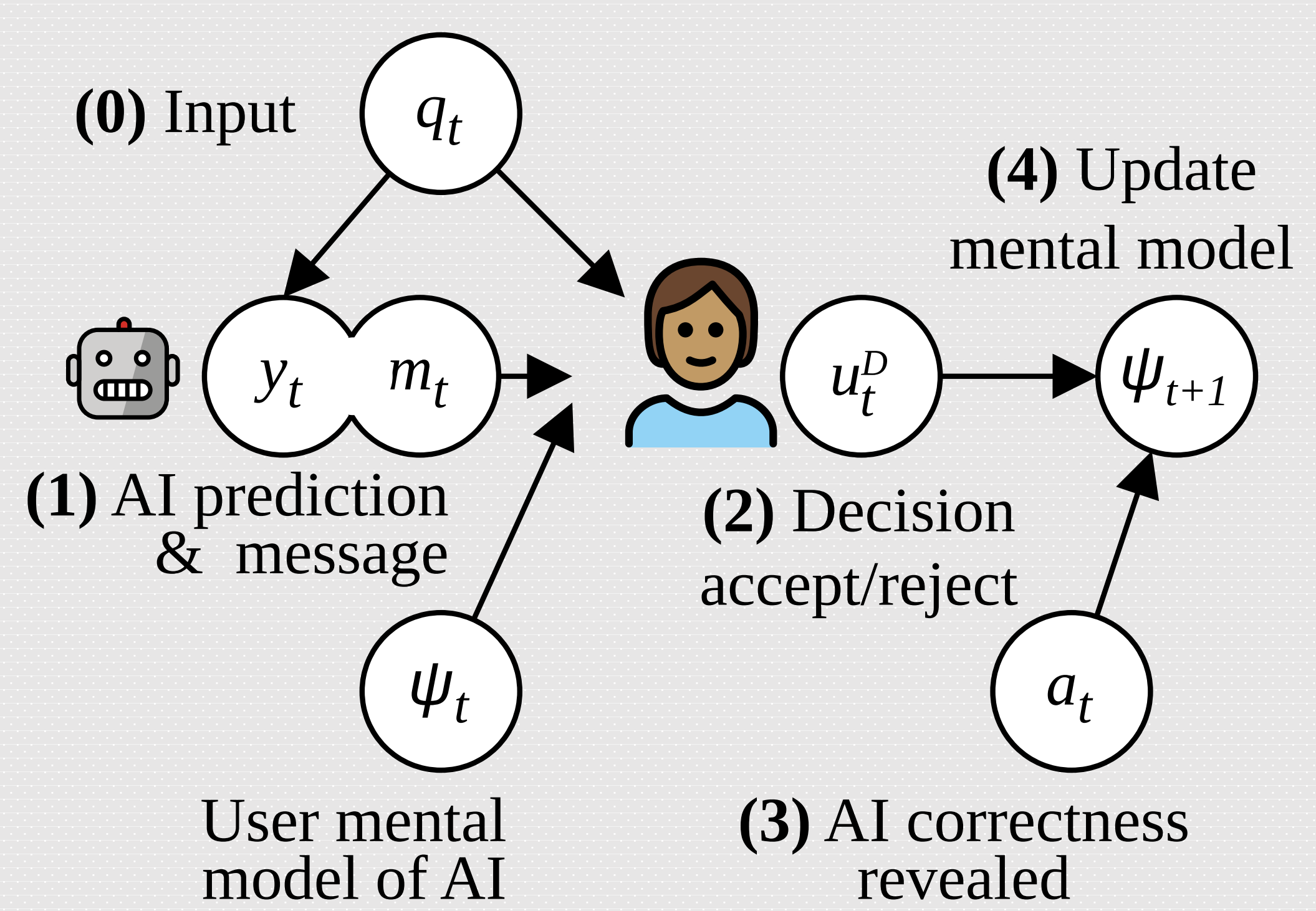
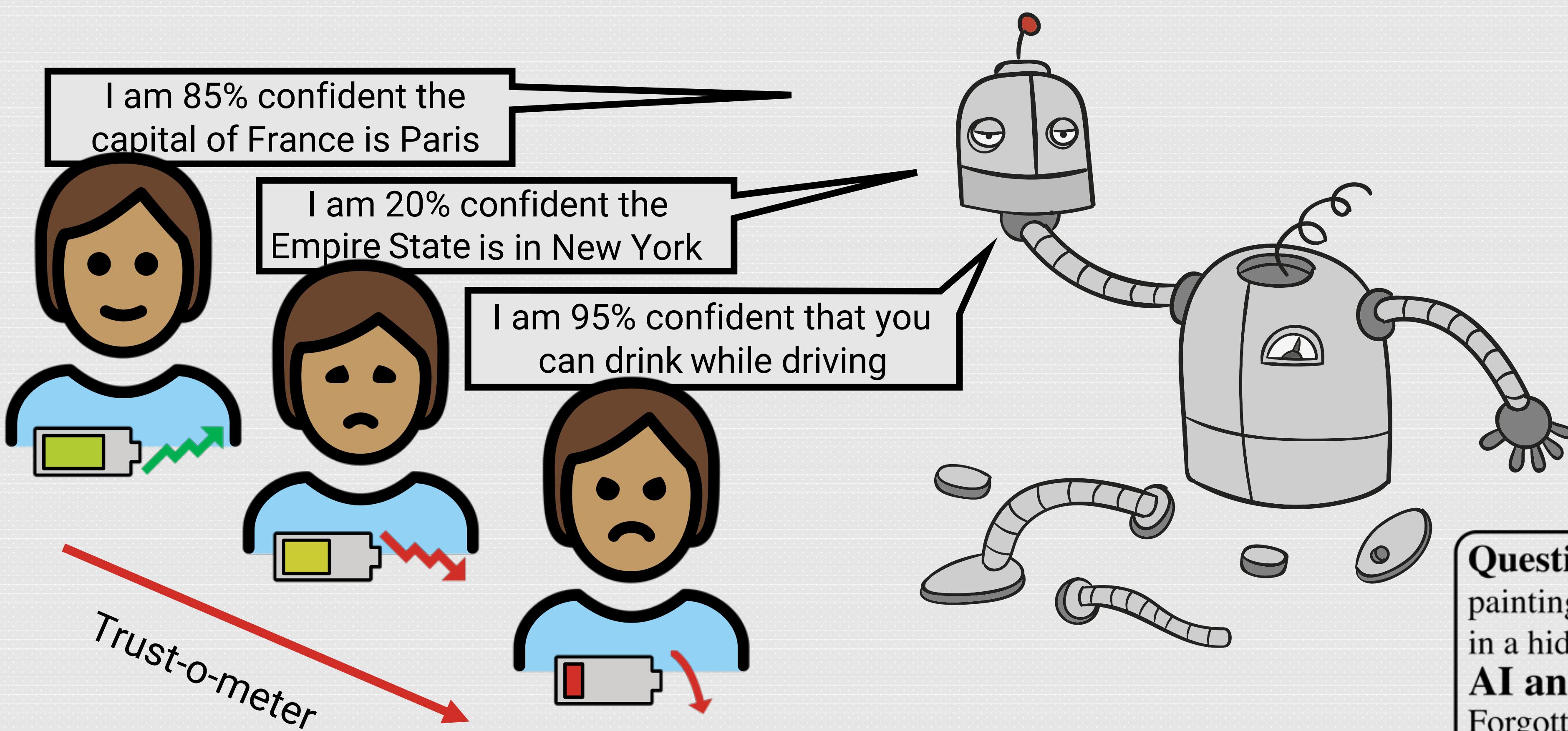


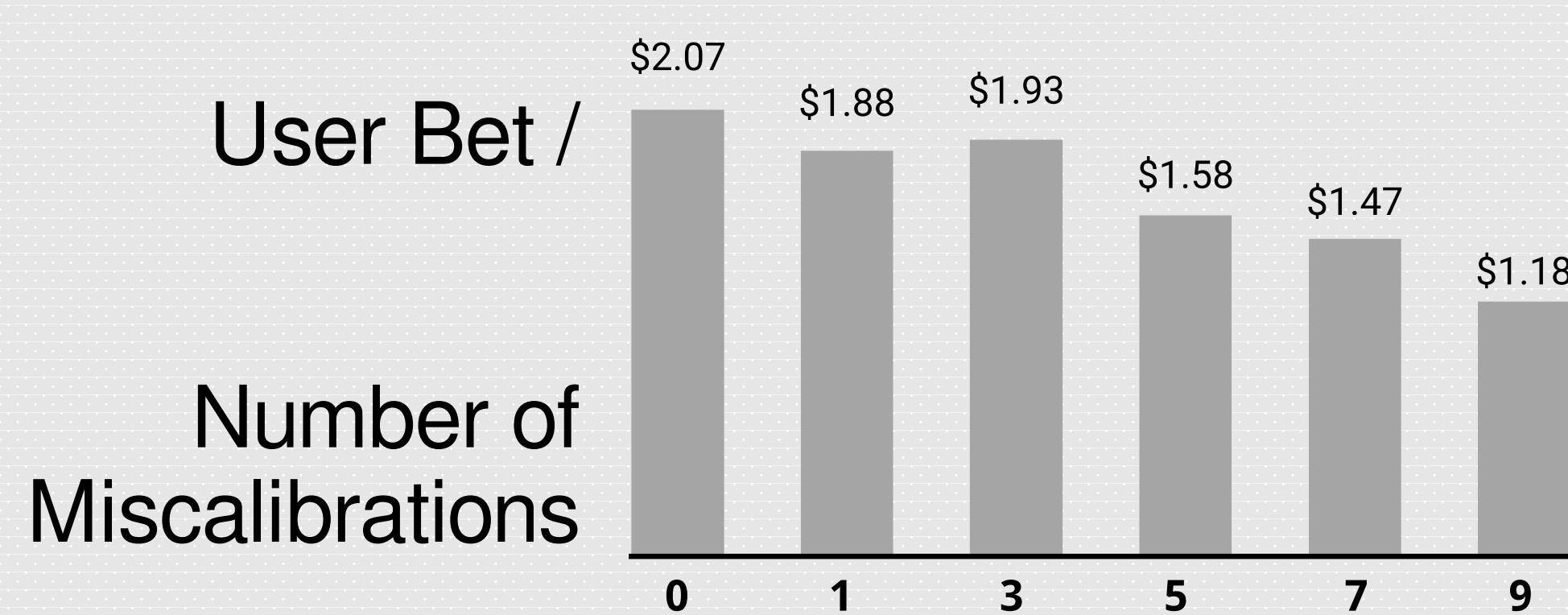
# WHEN TO NOT TRUST AI?



We simulate a set-up with users collaborating with AI, and measure how miscalibrations affects user trust and team performance

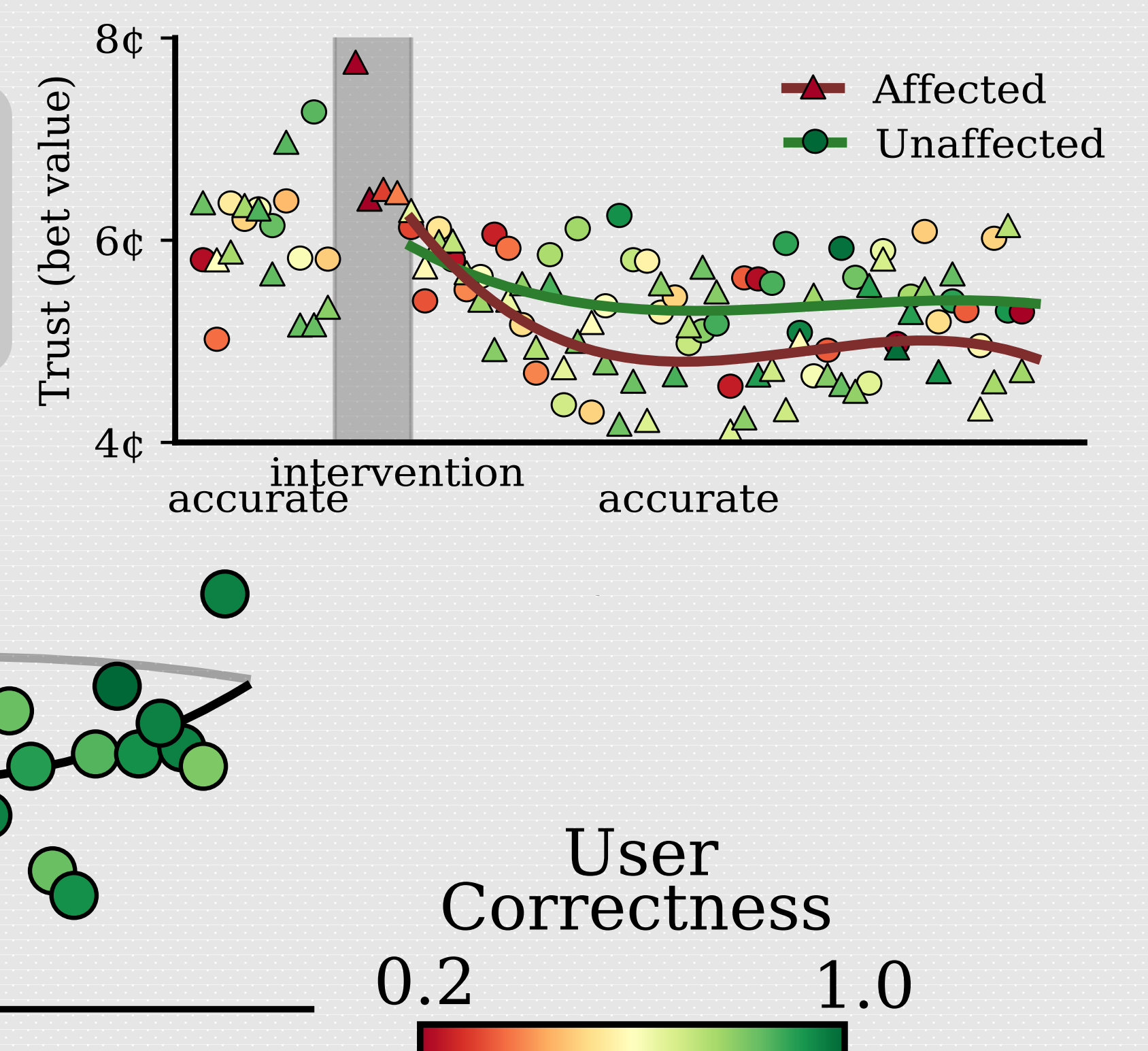
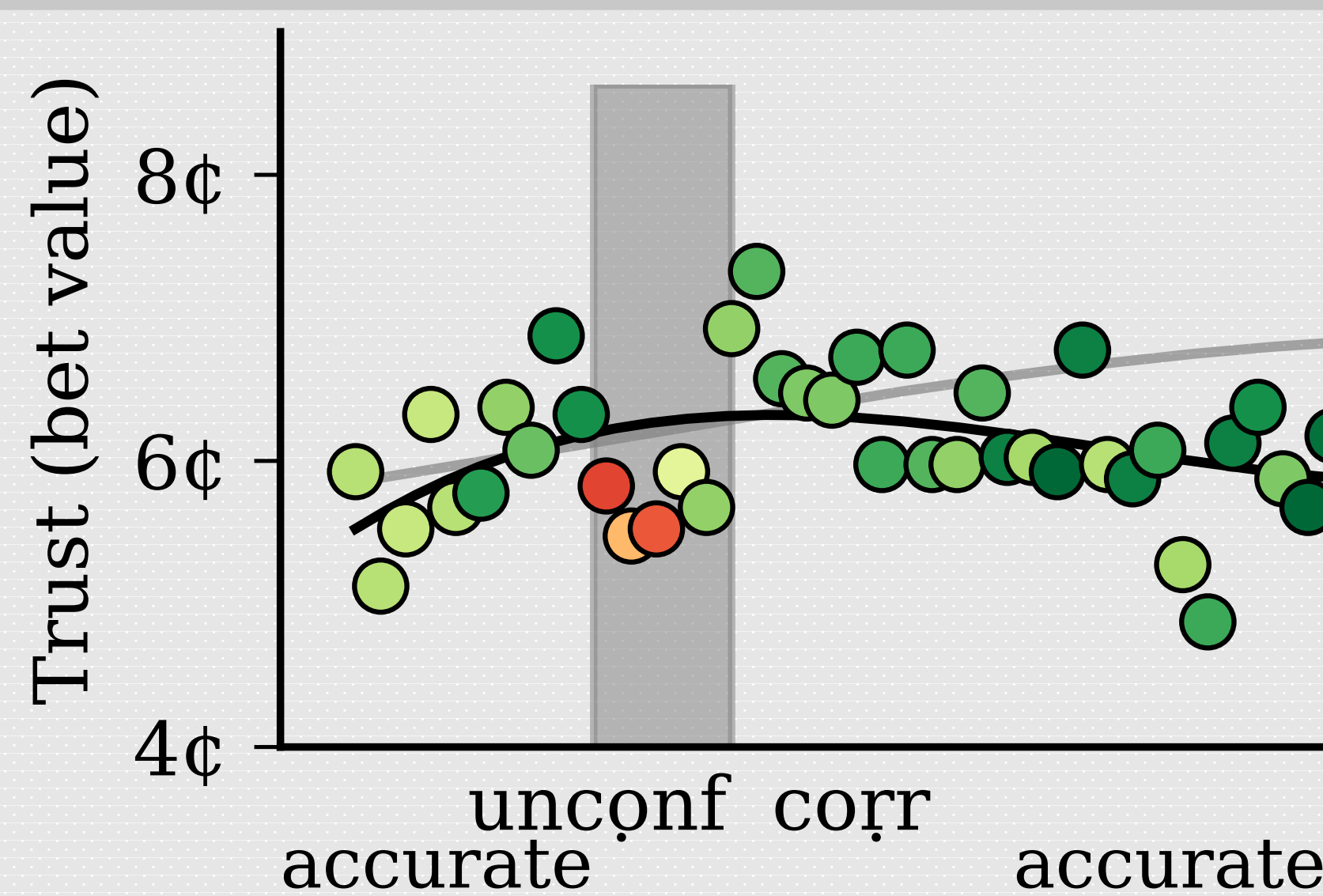
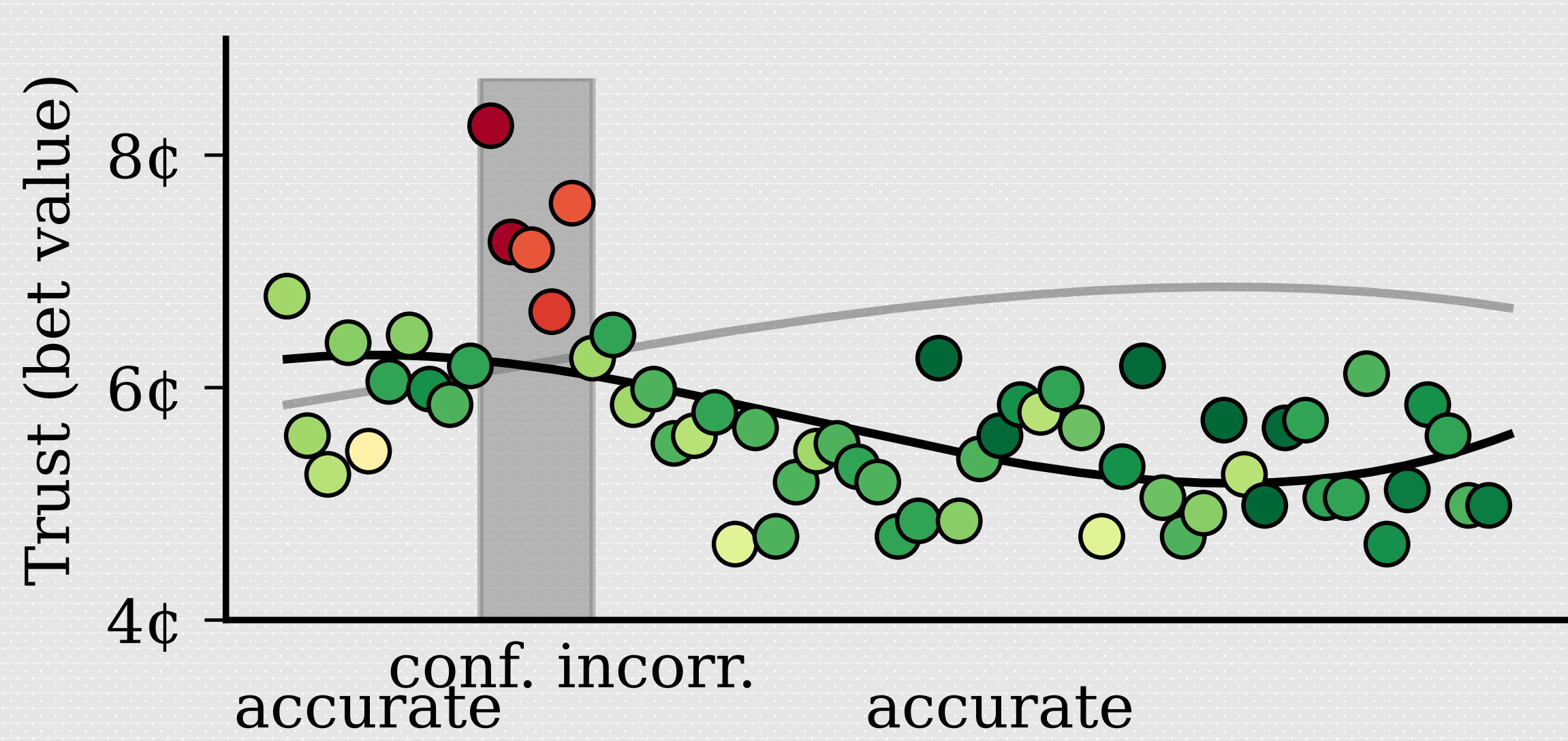
## MISCALIBRATIONS AFFECT TRUST

**Confidently incorrect** miscalibrations reduce trust more than **unconfidently correct** ones

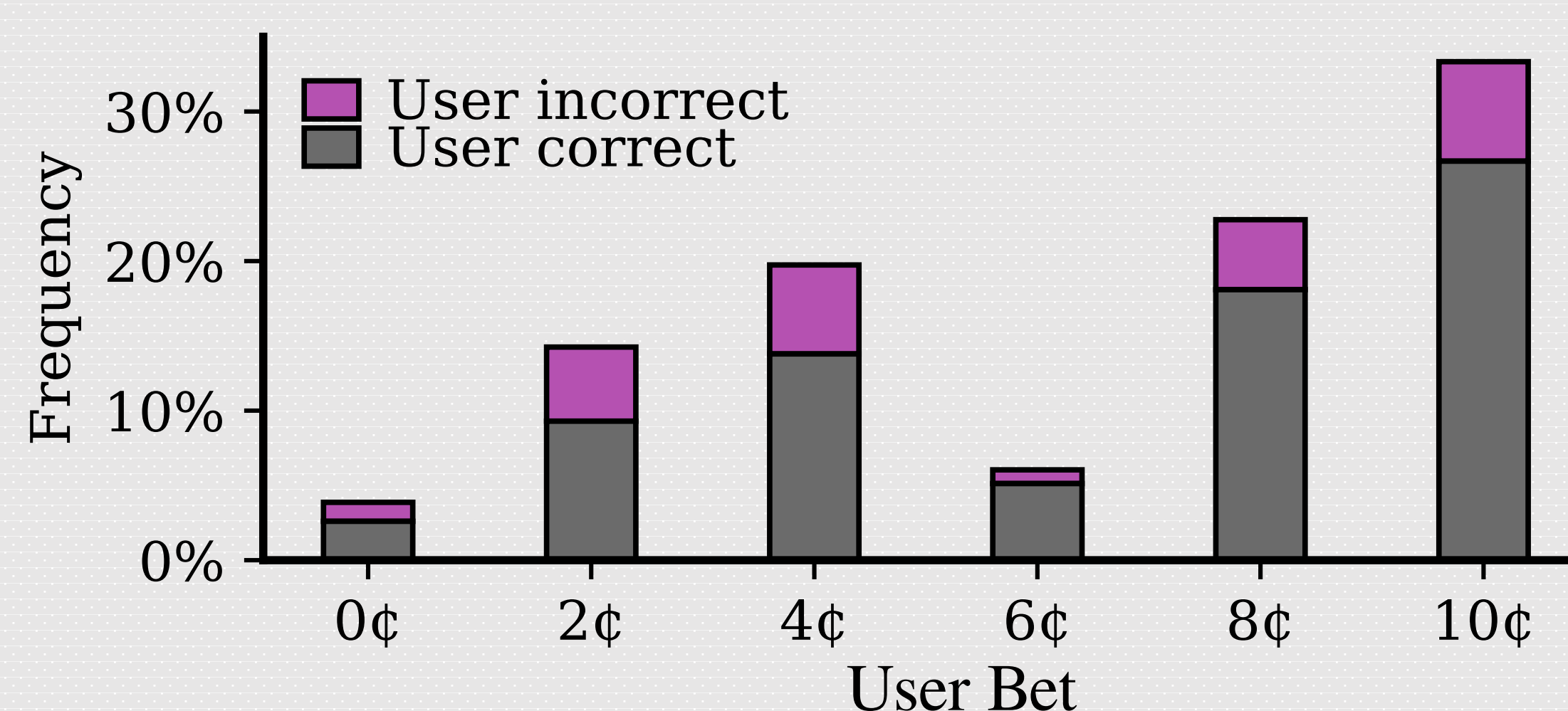
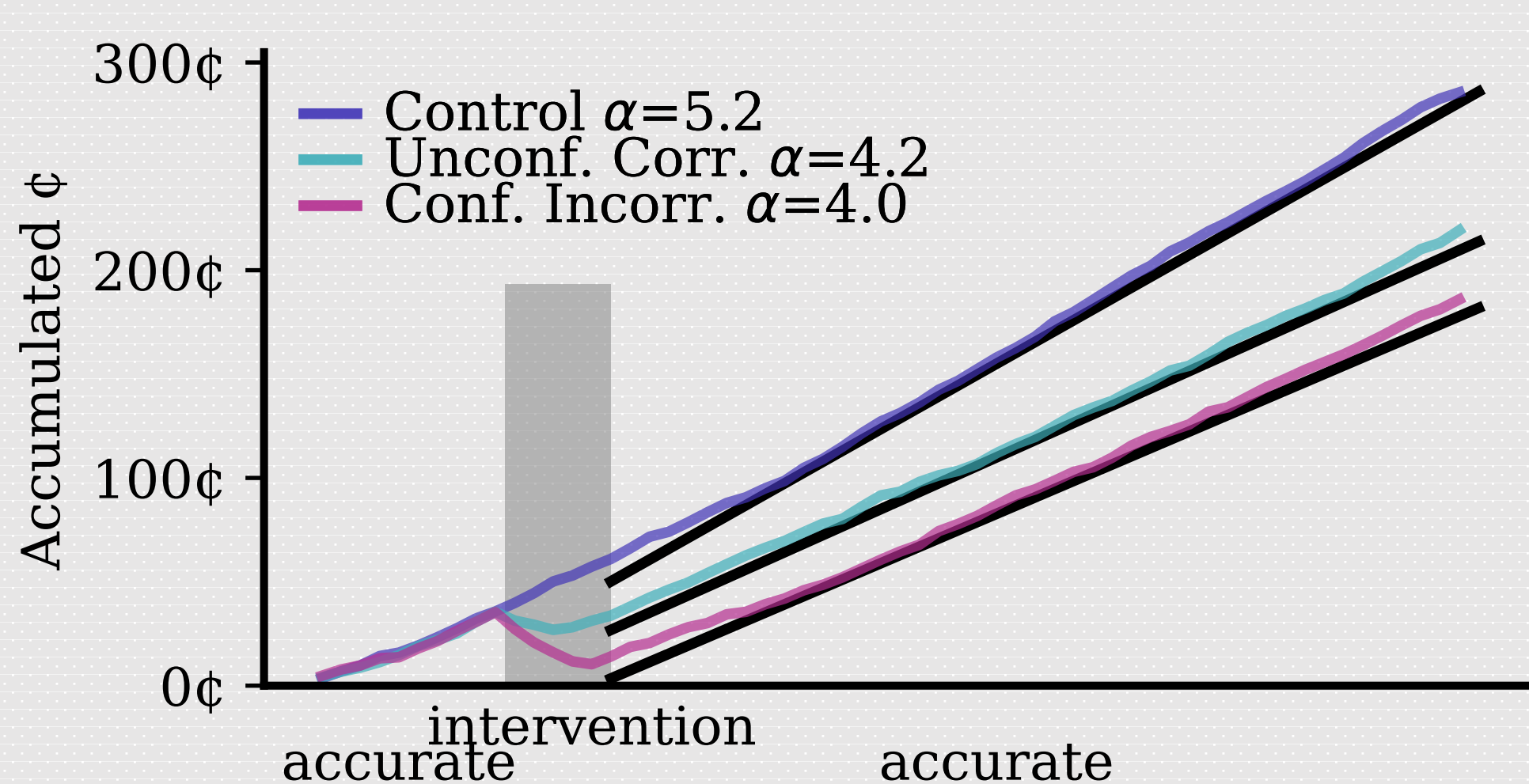


Even 5 miscalibrations are enough to affect user trust long-term

Miscalibration on one domain affect trust also on other domain but mostly on the original one → evidence for domain being represented in user mental model

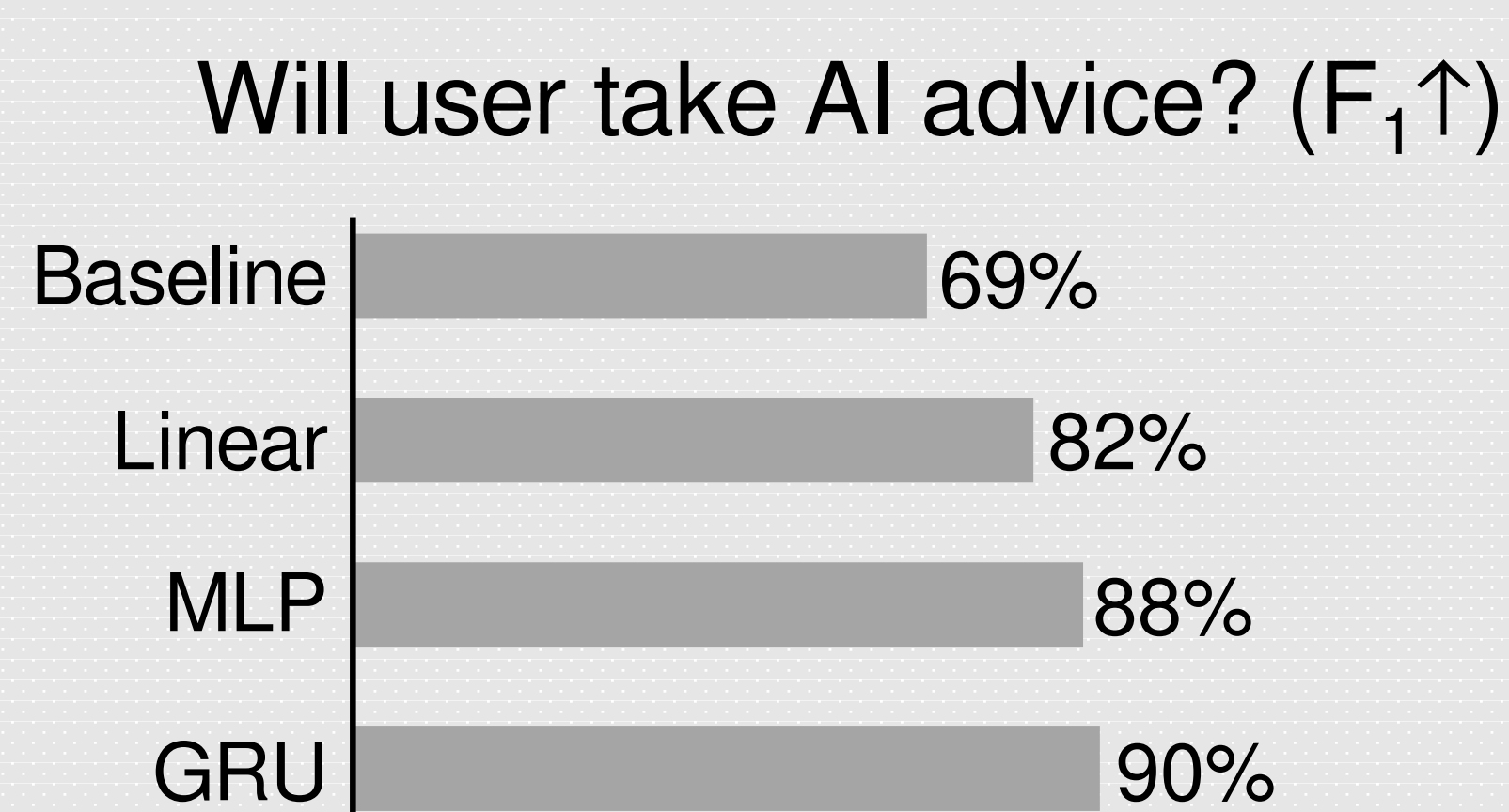
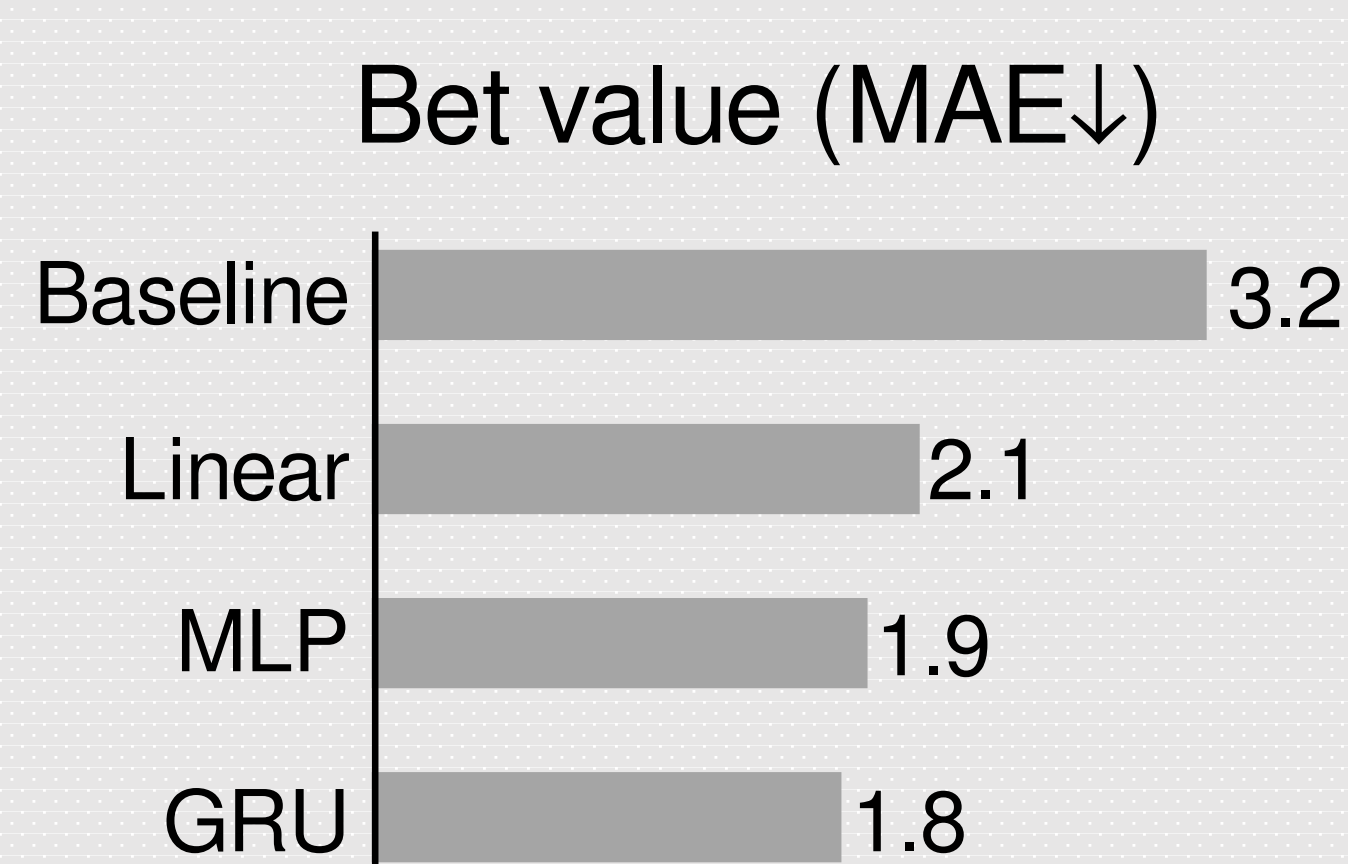


## MISCALIBRATIONS AFFECT HUMAN-AI COLLABORATION



With lower trust users do not accept AI predictions and lower Human-AI team success

## PREDICTING USER DECISIONS



Predicting user trust and decisions allows AI to adapt its output to improve collaboration; GRU learns how user trust evolves with respect to trust-eroding events.

