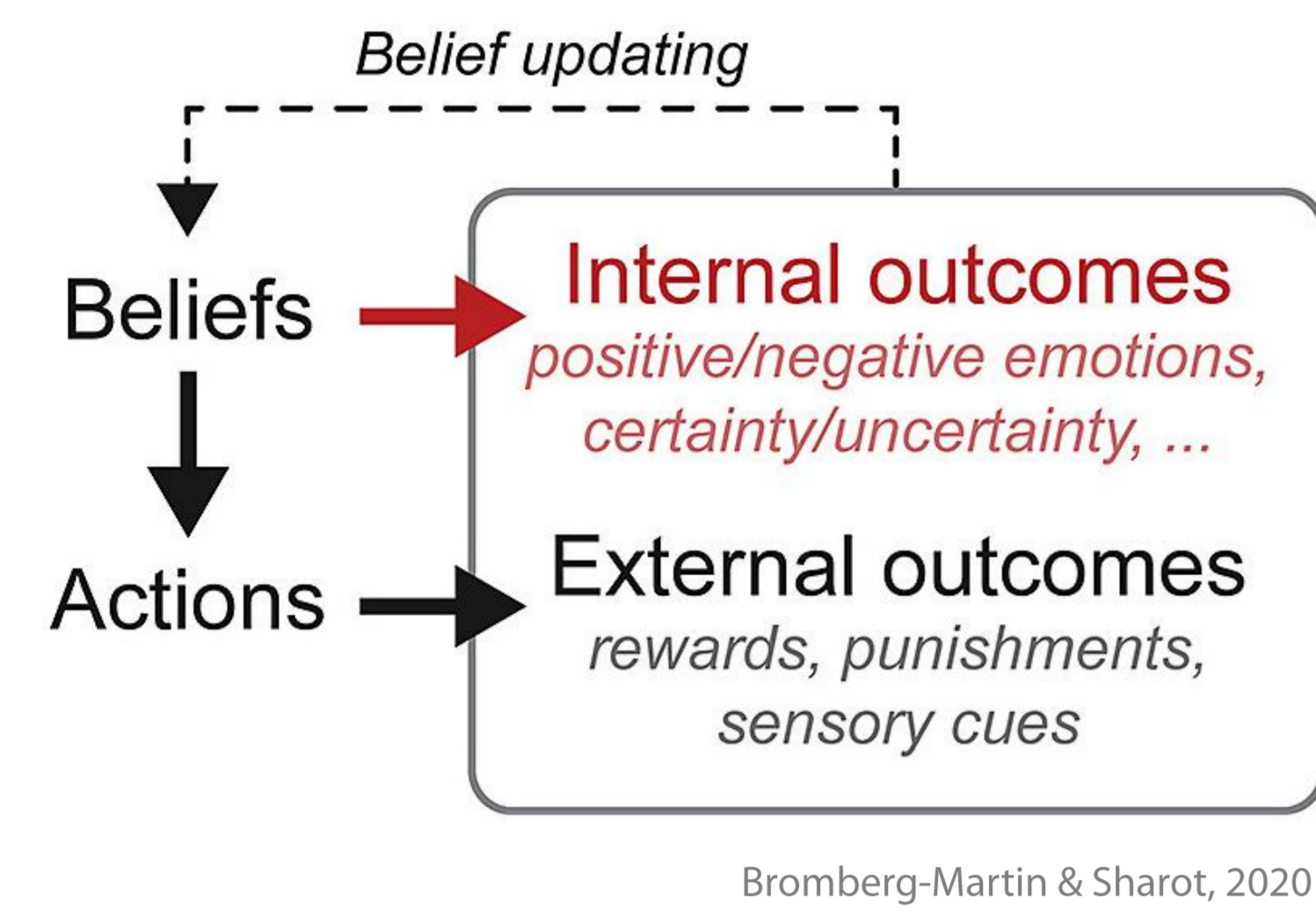


Biases in learning about the self and their association with affect

A.V. Mayer, N. Czekalla, A. Schröder, D.S. Stolz, F.M. Paulus, S. Krach, & L. Müller-Pinzler

Social Neuroscience Lab, Department of Psychiatry & Psychotherapy, Center for Brain, Behavior and Metabolism (CBBM), University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany

BACKGROUND



- Certain beliefs are more desirable than others & elicit affective experiences^{1,2}
- Desire to have certain beliefs and affects (internal outcomes) inherently biases self-related learning³

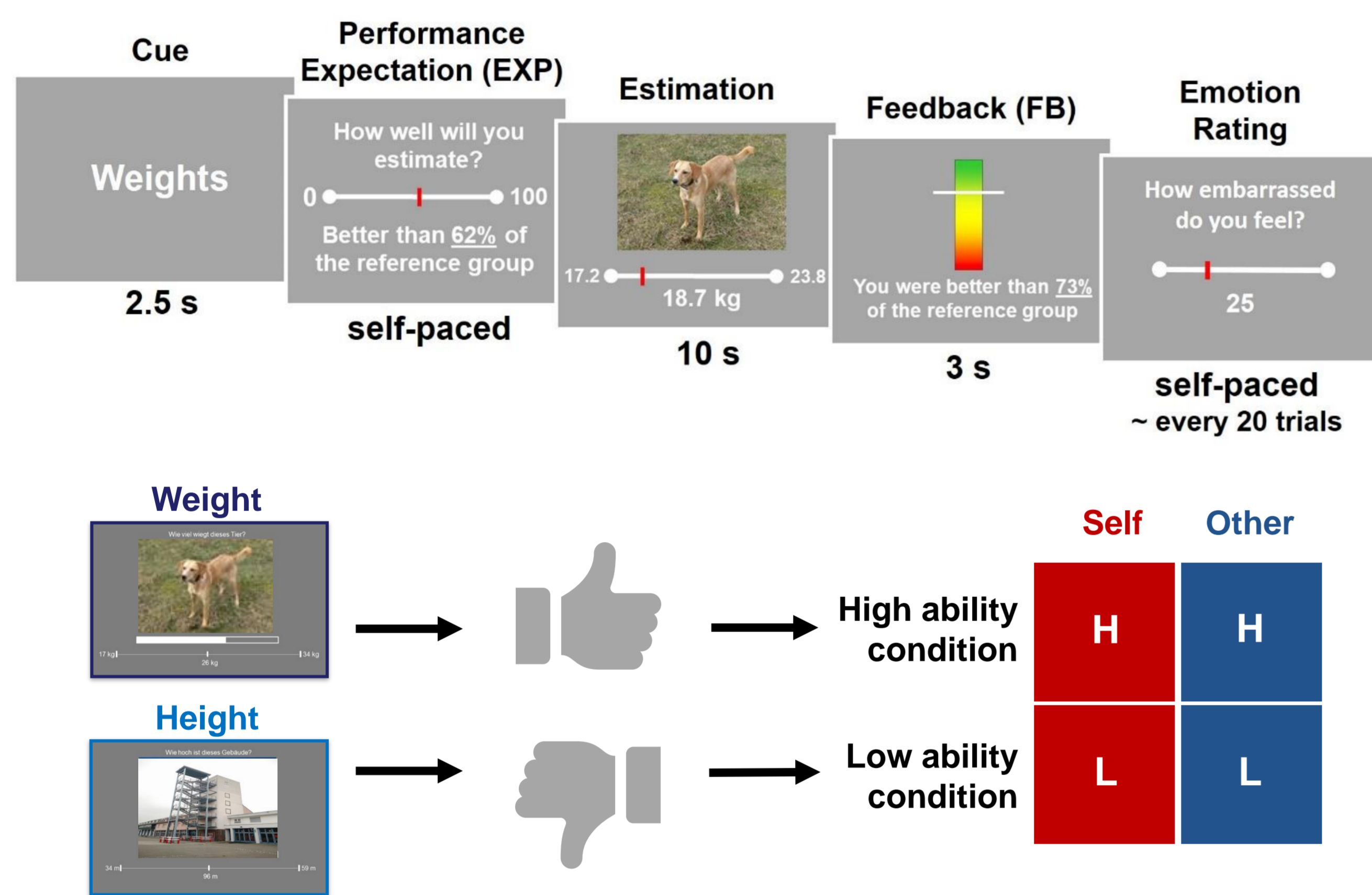
Bromberg-Martin & Sharot, 2020

AIMS

- To examine how people learn about their own abilities using an established estimation task^{4,5} and assess potential learning biases
- To examine whether these biases are directly linked to self-conscious affect (embarrassment and pride)
- To examine how affective states are integrated with self-related learning processes in the brain using fMRI

METHODS

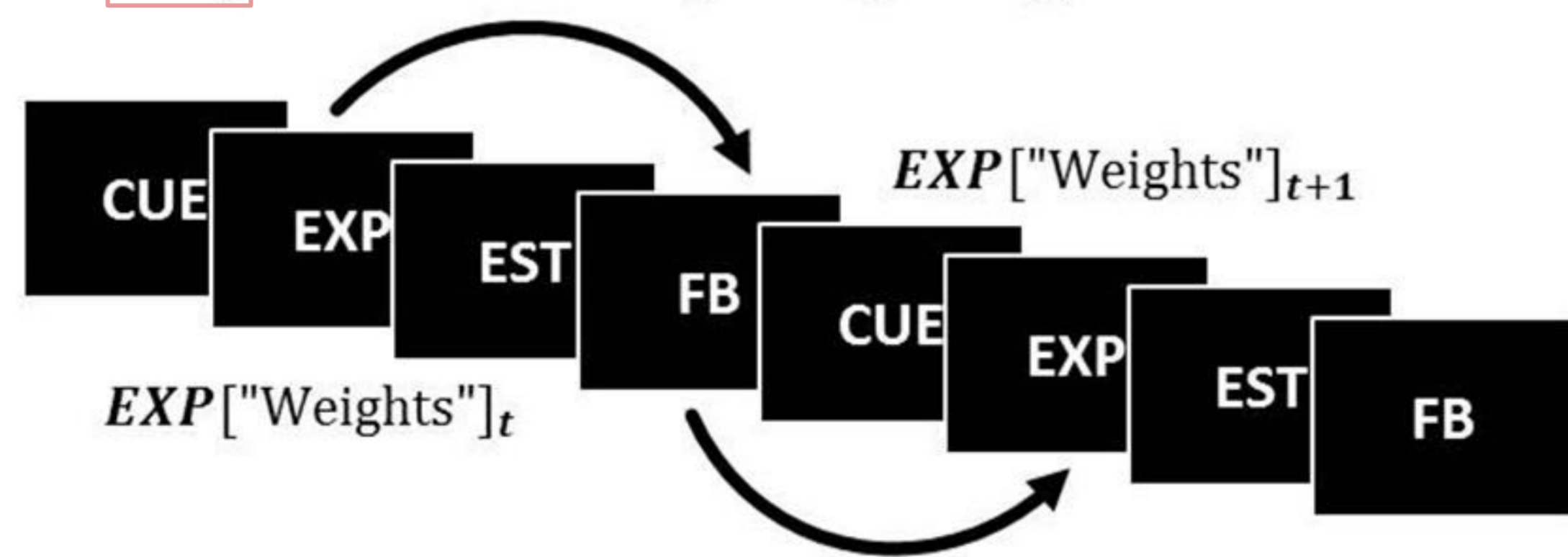
LOOP (Learning Of Own Performance) task
MRI: n = 39; behavior only: n = 30



An adapted Rescorla-Wagner model was used to model dynamic changes in performance expectations, that is, ability beliefs

PE: prediction error

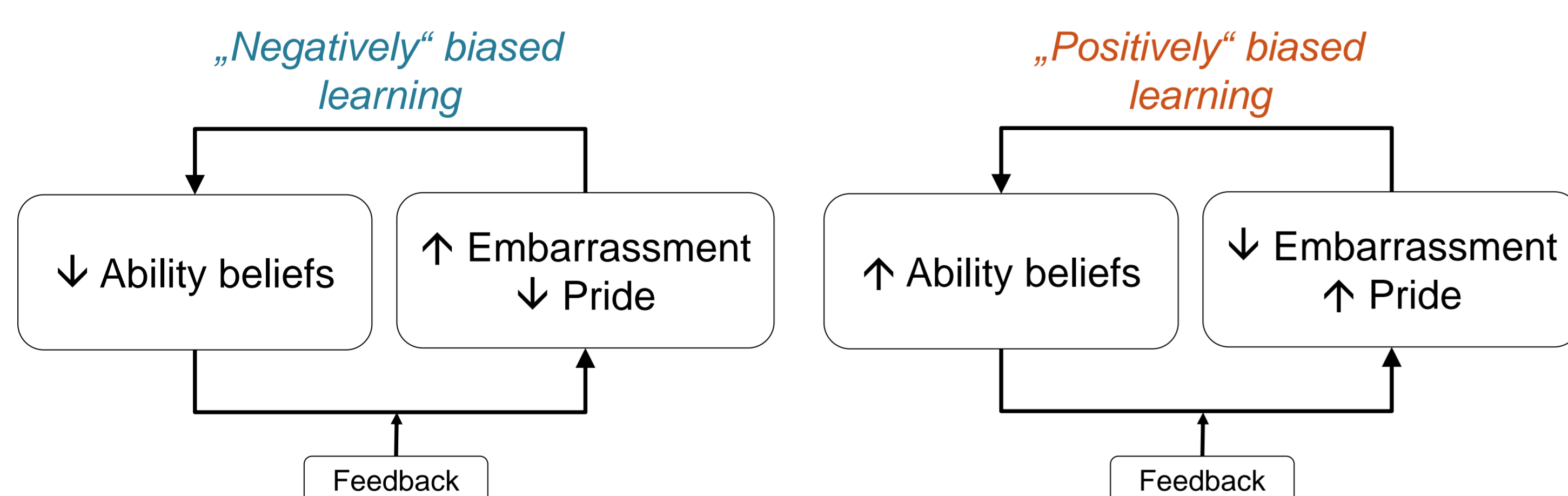
$$PE_t = FB - EXP["Weights"]_t$$



$$EXP["Weights"]_{t+1} = EXP["Weights"]_t + \alpha \cdot PE_t$$

α : learning rate

DISCUSSION

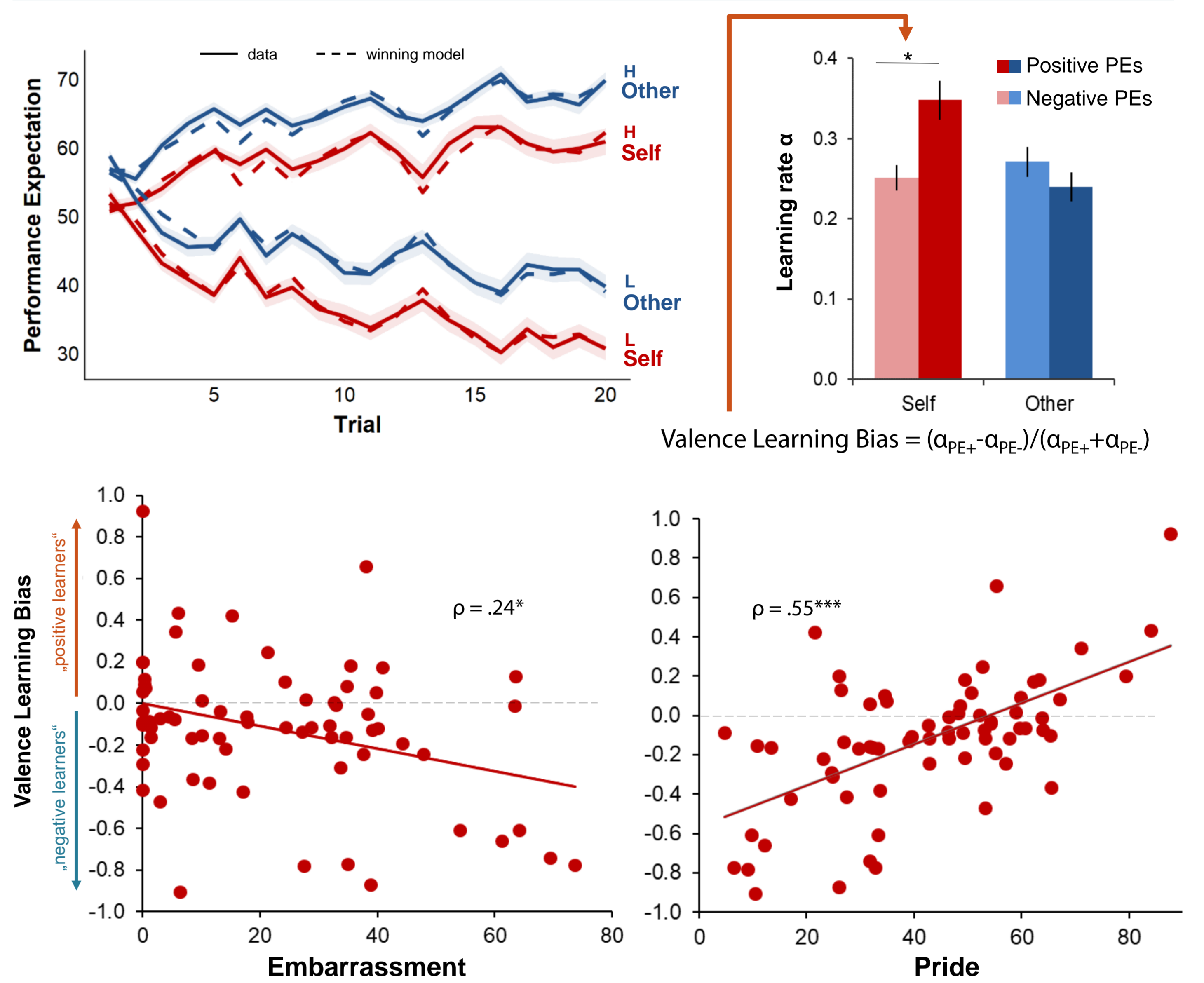


Affective states potentially shape self-related learning via:

- affective tagging of stimuli, altered attention allocation and stimulus processing (amygdala, anterior insula⁶)
- shift of dopaminergic responses to PEs (VTA/SN⁷)
- altered value representation (mPFC⁸)

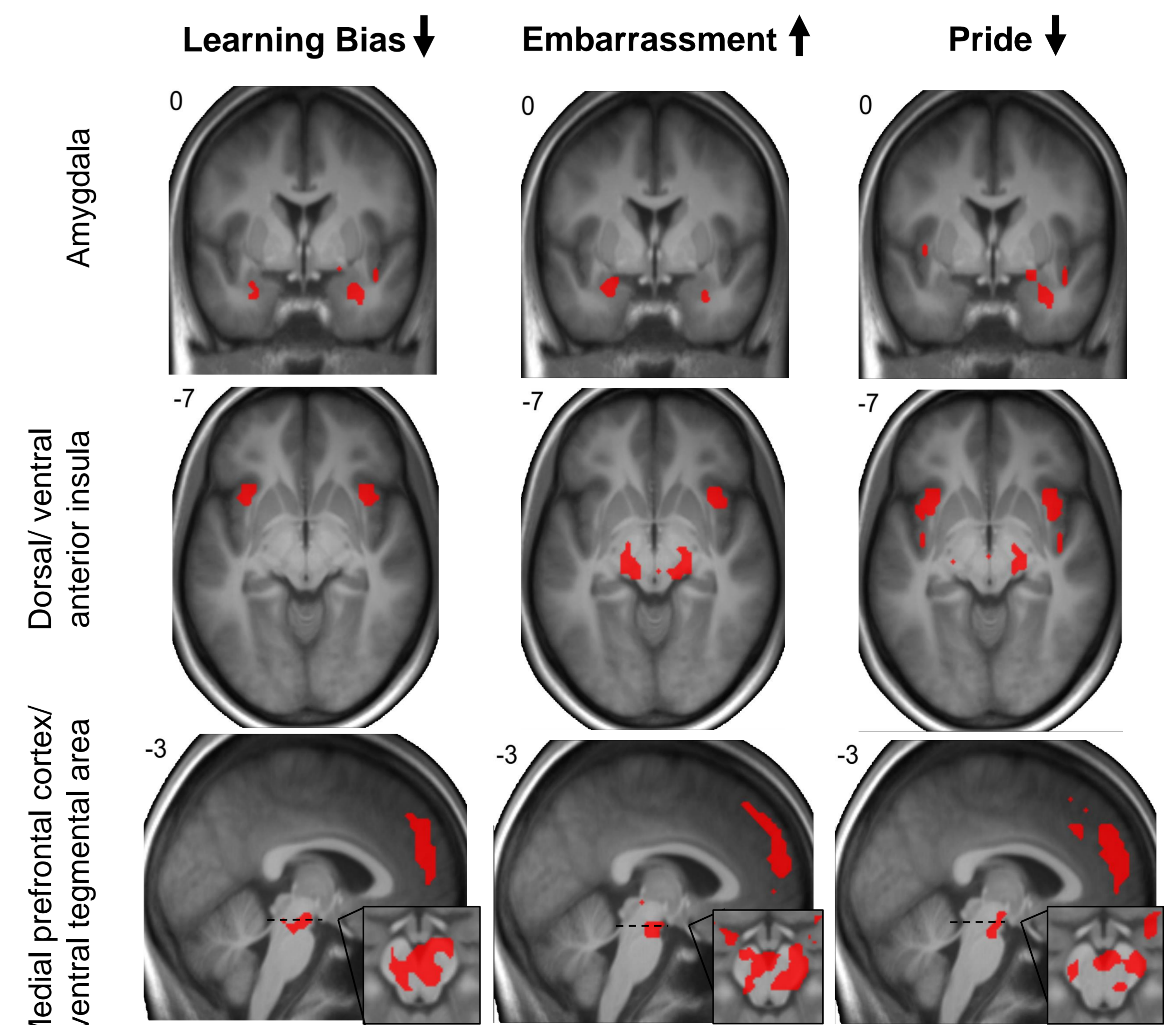
RESULTS

Embarrassment and pride ratings are associated with biased updating of ability beliefs



Participants who updated their ability beliefs more negatively:

- experienced stronger embarrassment and less pride
- had stronger responses to more negative prediction errors in brain regions associated with value representation, prediction error processing in general, and attentional processes



Increased activity in response to negative PEs vs. positive PEs. $p < .005$, uncorrected for displaying purposes, statistical results FWE-corrected within ROIs

References

1. Rutledge, R. B., Skandali, N., Dayan, P., & Dolan, R. J. (2014). A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences*, 111(33), 12252-12257.
2. Stolz, D. S., Müller-Pinzler, L., Krach, S., & Paulus, F. M. (2020). Internal control beliefs shape positive affect and associated neural dynamics during outcome valuation. *Nature communications*, 11(1), 1230.
3. Bromberg-Martin, E. S., & Sharot, T. (2020). The value of beliefs. *Neuron*, 106(4), 561-565.
4. Müller-Pinzler, L., Czekalla, N., Mayer, A. V., Stolz, D. S., Gazzola, V., Keysers, C., ... & Krach, S. (2019). Negativity-bias in forming beliefs about own abilities. *Scientific reports*, 9(1), 14416.

5. Czekalla, N., Stierand, J., Stolz, D. S., Mayer, A. V., Voges, J. F., Rademacher, L., ... & Müller-Pinzler, L. (2021). Self-beneficial belief updating as a coping mechanism for stress-induced negative affect. *Scientific Reports*, 11(1), 17096.
6. Koban, L., & Pourtois, G. (2014). Brain systems underlying the affective and social monitoring of actions: an integrative review. *Neuroscience & Biobehavioral Reviews*, 46, 71-84.
7. Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., & Gabrieli, J. D. (2006). Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron*, 50(3), 507-517.
8. Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., & Rangel, A. (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *Journal of neuroscience*, 28(22), 5623-5630.

