Notes on Lefebvre et al., 2017 and Katahira (2018)

Lefebvre find that in a binary outcome task with gain/no-gain given as payoffs (.5, 0) people weight positive prediction errors more than negative prediction errors.

They also found this in Experiment 2 which where participants received either a gain or loss (.5, -.5).

This finding seems very believable in an all-gains task, but I don't know if it will hold up in studies with gains and losses – for example, I think the overall poor performance in the Soochow Gambling task is driven by attention to frequent losses. The finding also seems to run counter to evidence that "losses loom larger" than gains supported by Prospect theory and decades of research. Perhaps prediction errors are different somehow, though.

Katahira, 2018 suggests that the delta model with separate learning rates for positive and negative prediction errors allows the model to account for autocorrelation, and that the higher learning rates for positive PEs is indicative of perseveration rather than a positivity or confirmation bias. Katahira calls it a "pseudo-confirmation bias."

Right now I think Katahira might be right, but more work needs to be done.