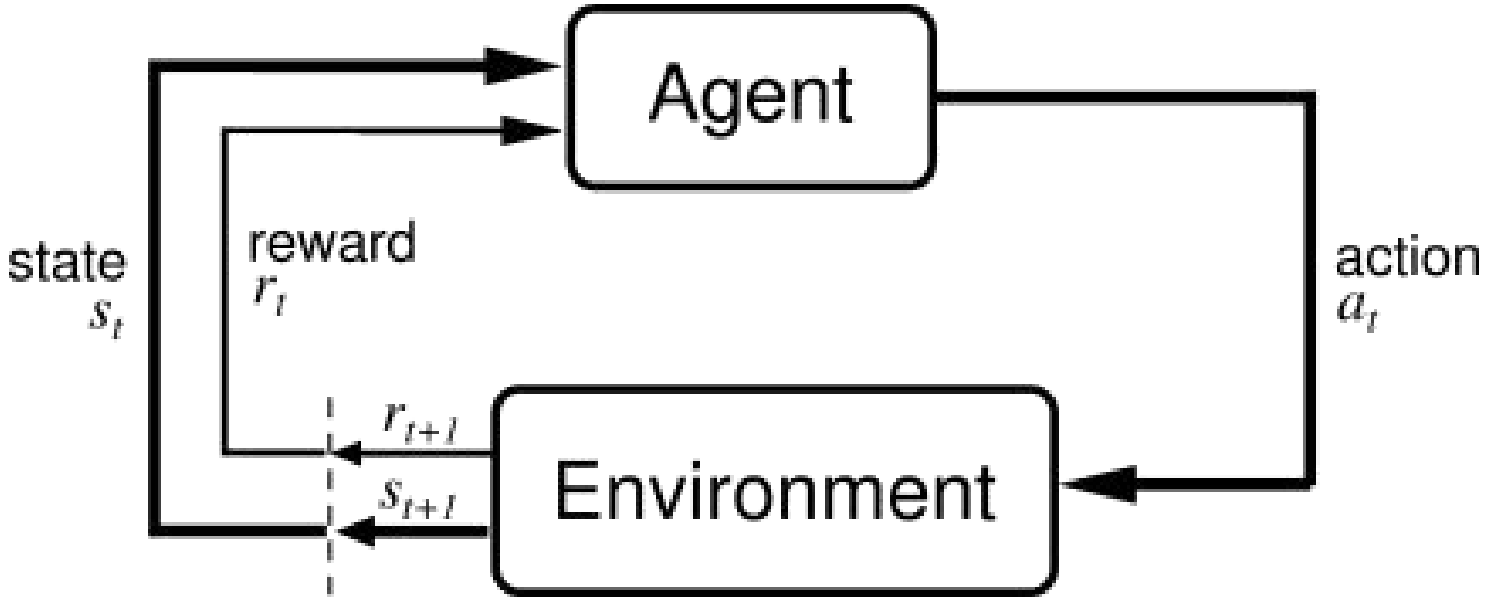
Reinforcement Learning in the Human Brain

Presented by Chris Foster

How can we model learning?

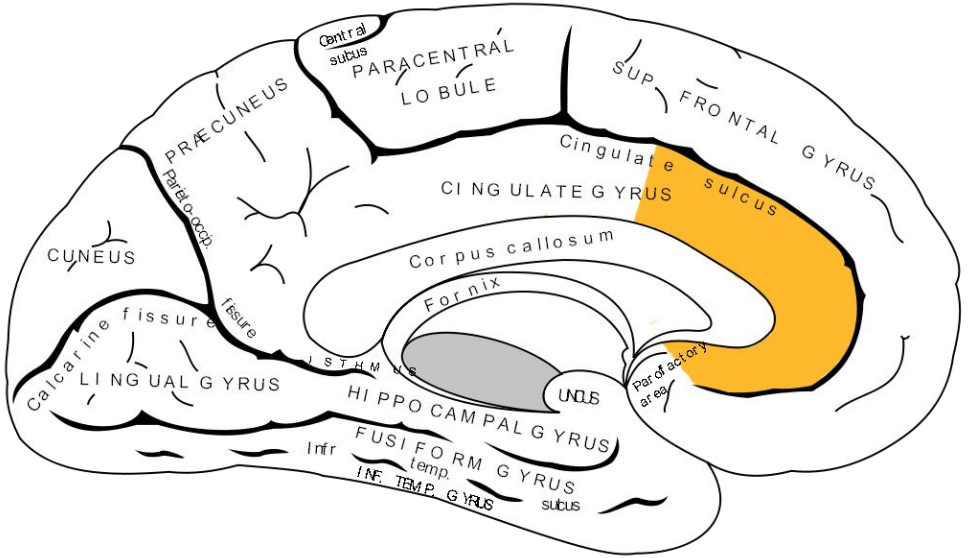
Agent-based Mathematical Models



What about the brain?

The ACC could be how the brain manages “options” through hierarchical reinforcement learning

Holroyd & Yeung developed a computational model explaining the role of the ACC



Anterior Cingulate Cortex (ACC)

Error Related Negativity (ERN)

This is also known as the Error Negativity (Ne).

A measurable and consistent negativity detected after the human had made an error while performing a task, and knew that they had performed an error.

The Error Related Negativity is *time-locked* to the response.

A New Event Related Potential

An Event Related Potential (ERP) is a stereotyped electrophysiological response to a stimulus.

Task: participants must press a button 1 second after an auditory cue. The allowed margin of error would shrink or grown such that their accuracy would be exactly 50%.

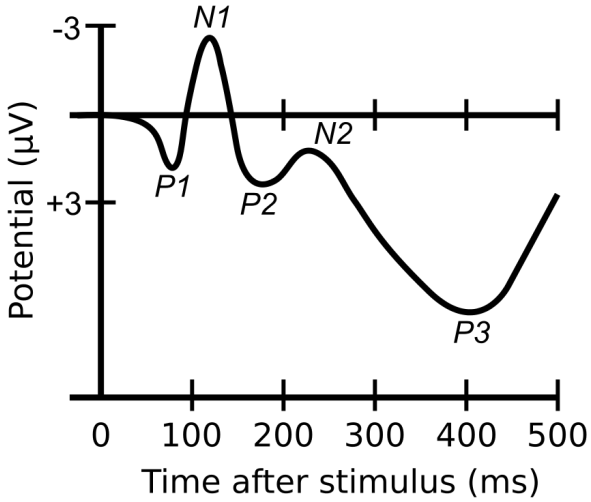
In response to the stimulus, a detectable negative potential with very similar scalp distribution and source localization to the ERN was detected.

We will call this the feedback ERN (fERN).

Two Negatives

We now have two negatives:

- Response-locked ERN
- Feedback-locked ERN (fERN)



What is their relationship?

Let’s look at a reinforcement learning paradigm!

When participants did not know stimulus-response mappings:

Negative feedback elicited the fERN while erroneous responses did not

When participants did know the stimulus-response mappings:

Response errors elicited an ERN while negative performance feedback did not

Is this the discovery of a generic error monitor system for the brain?

Meanwhile: a gambling experiment

Participants could win money and have made the “wrong” choice

Losses caused a negativity around 250ms after feedback -- regardless of whether the feedback indicated the choice was correct or an error.

Since it did not appear to track performance errors, they referred to it as the Medial Frontal Negativity (MFN).

Subsequent studies also referred to it as the Feedback Related Negativity (FRN) and also the Feedback Negativity (FN).

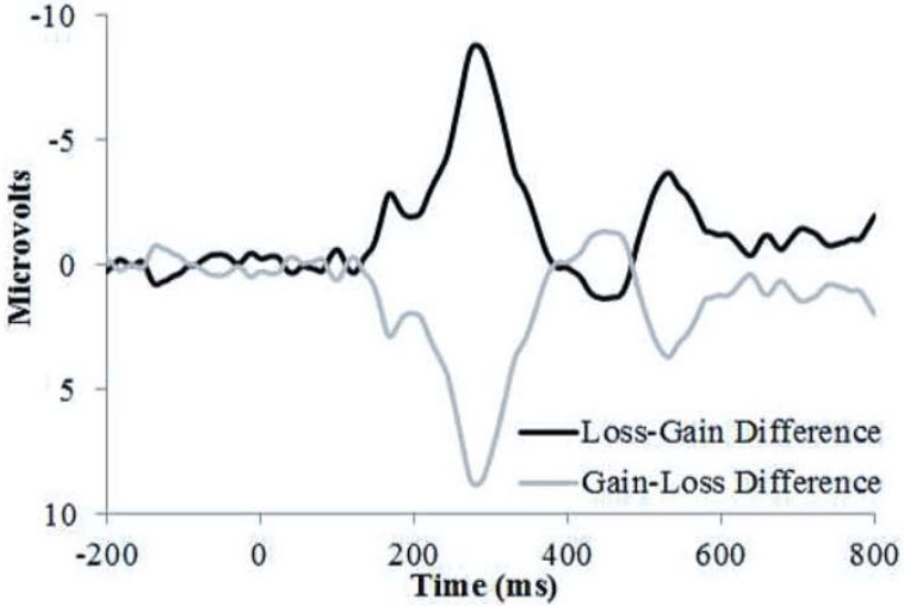
What’s in a name?

We now have a bunch of negativities including:

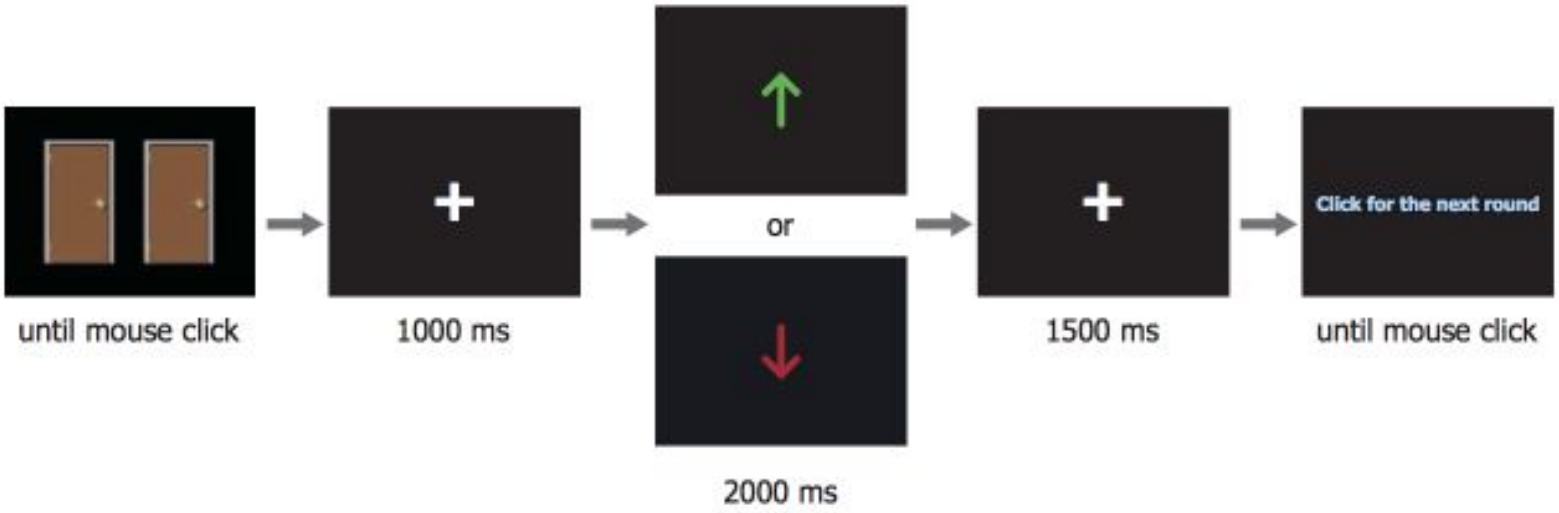
1. Feedback Event Related Negativity (fERN)
2. Medial Frontal Negativity (MFN)
3. Feedback Related Negativity (FRN)
4. Feedback Negativity (FN)

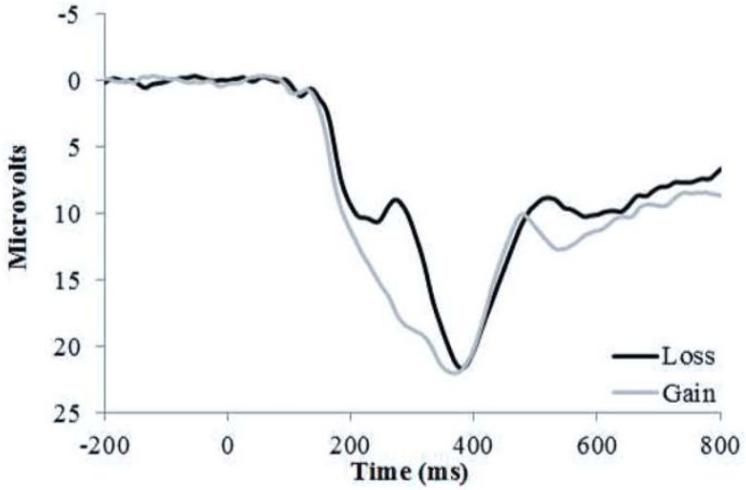
But what if these were all the same thing?

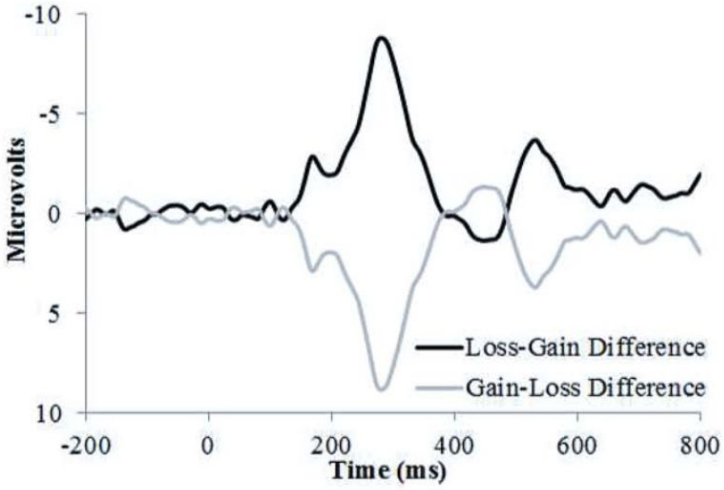
Reward Positivity

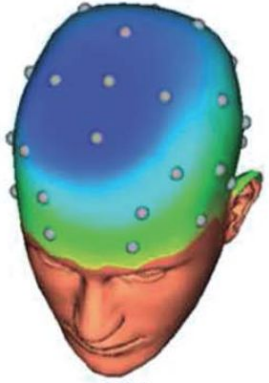


Doors Task


until mouse click
1000 ms
or
2000 ms
1500 ms
Click for the next round
until mouse click

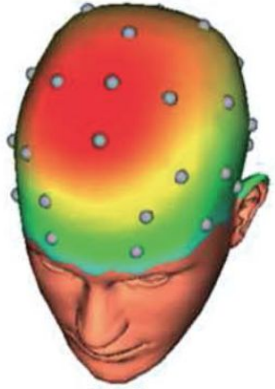






ΔFN
(Loss-Gain)
250-350 ms

-6.2μV 0μV



ΔFN
(Gain-Loss)
250-350 ms

0μV 6.2μV

So is it a loss-related negativity or a gain-related positivity?

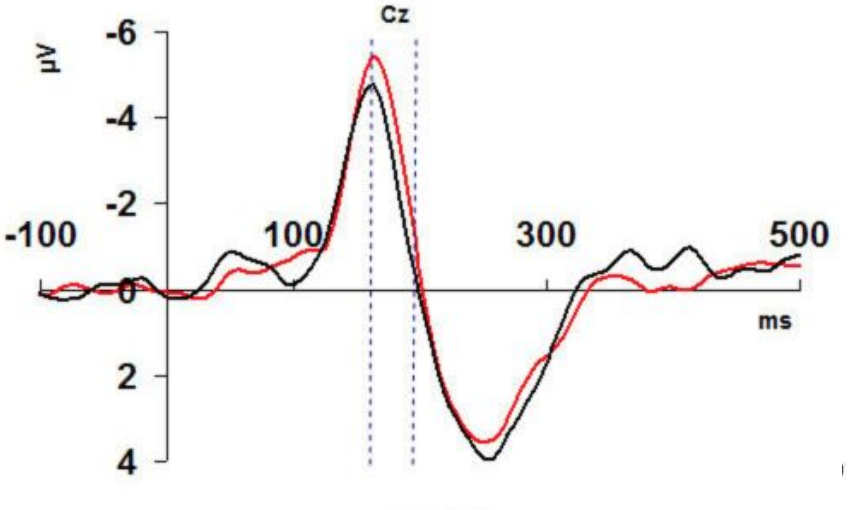
Depends on how you look at it based on the graphs above.

What can we call the “baseline response”?

In further experiments, it was found that breaking even had the same response as losing. Therefore it can be considered the baseline response to not have a positive feedback.

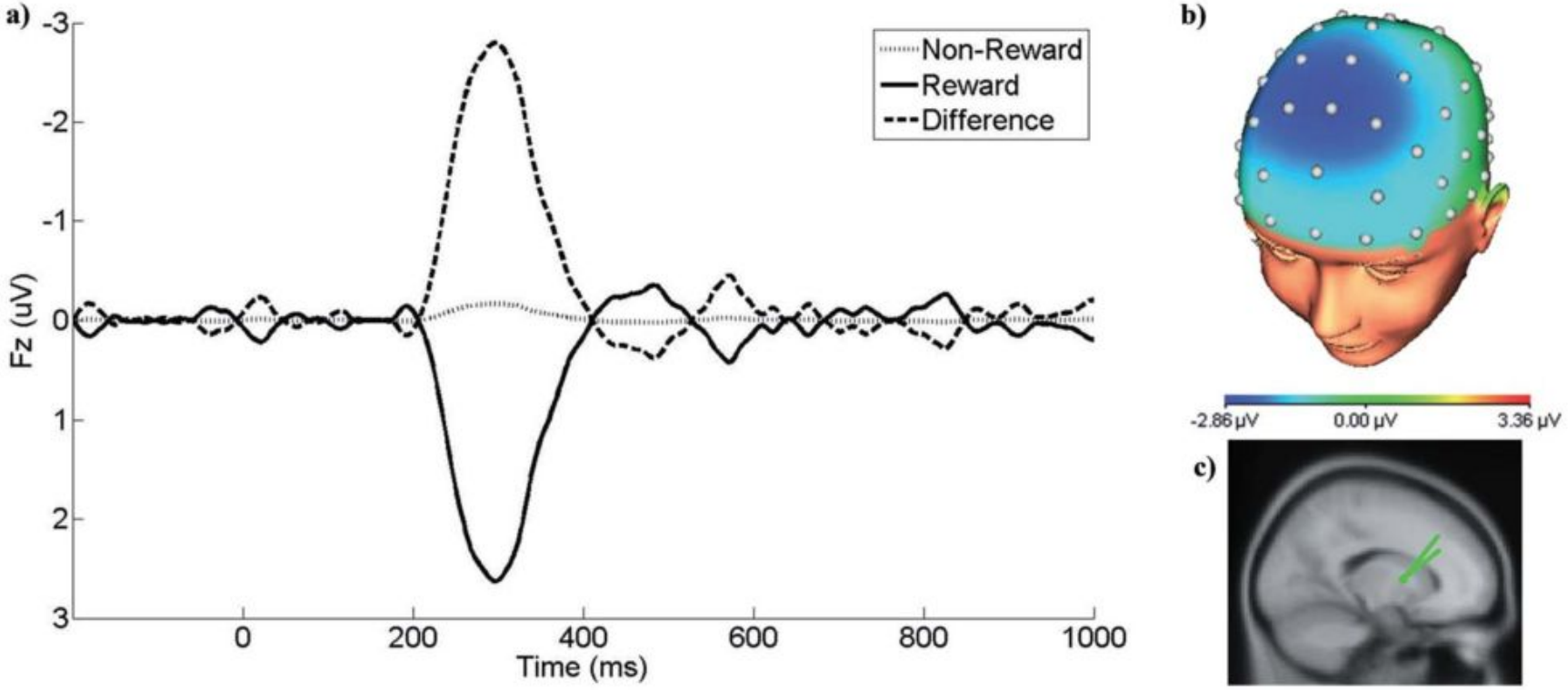
The apparent loss-related negativity is extremely close to a N200, a baseline response when a stimulus contains no information. It’s possible that the N200 is the baseline response and it is suppressed by the RewP which occurs in the same time range.

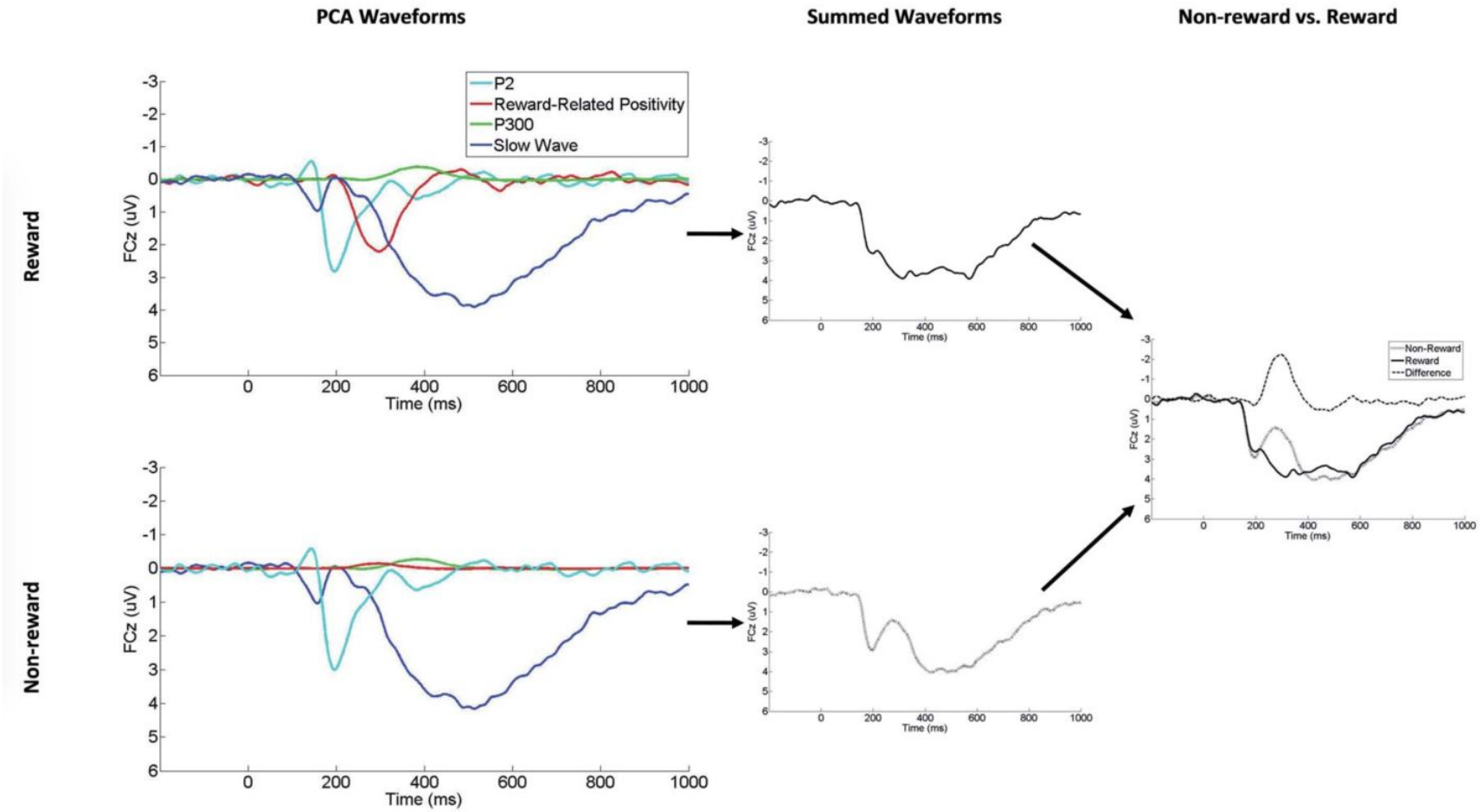
So in this case, it appears that instead of a feedback-related negativity this is in fact a feedback-related positivity when there is a reward!



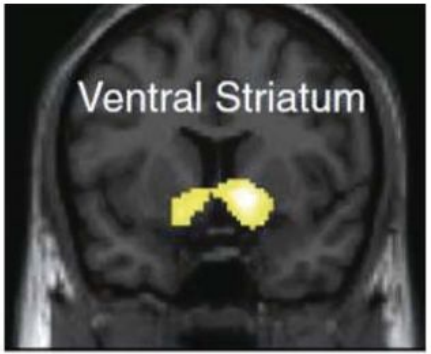
N200

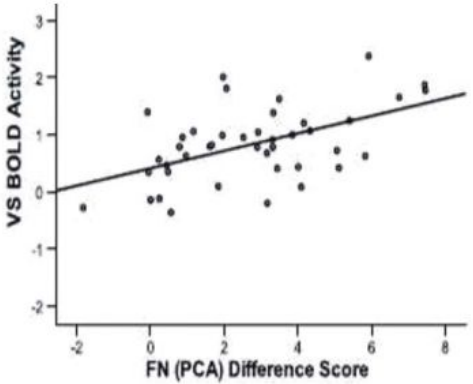
PCA

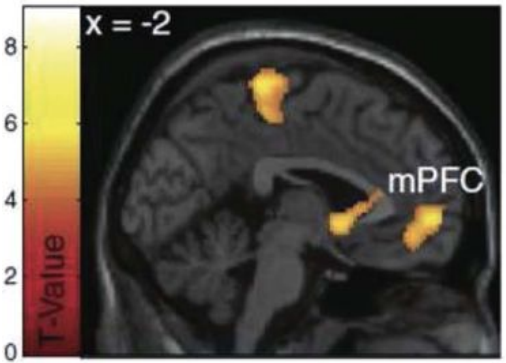

a)
b)
c)
Non-Reward
Reward
Difference
Fz (uV)
Time (ms)
-3
-2
-1
0
1
2
3
0
200
400
600
800
1000
-2.86 µV
0.00 µV
3.36 µV

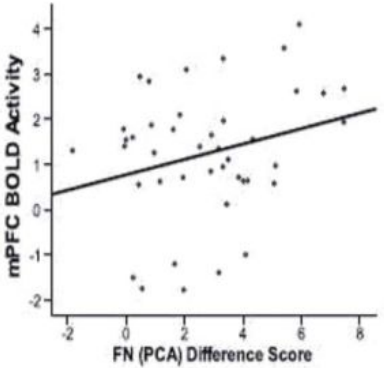

PCA Waveforms
Summed Waveforms
Non-reward vs. Reward
Reward
Non-reward
P2
Reward-Related Positivity
P300
Slow Wave
FCz (uV)
Time (ms)
Non-Reward
Reward
Difference

Analysis of RewP in fMRI









Reward Positivity between groups

Group 1: Receives their task payout in real cash each block

Group 2: Does not receive their task payout

When the payout of the doors task was real, the average RewP response within a group was larger than when the payout was not real.

This indicates that the magnitude of the RewP is tied to the level of reward itself!

Reward Positivity between individuals

The Reward Responsiveness Scale is a numerical measure test of a person's response level to a general reward stimulus. If RewP works as we would expect it to, it should model other reward measurement mechanisms.

In fact, studies did find that an individual’s RewP magnitude was correlated to a person’s location on the Reward Responsiveness Scale!

Aside: RewP correlation to depression

A further study revealed that increased depressive symptoms of 85 college undergraduates could be predicted based on a smaller RewP magnitude.

Subsequently, smaller RewP was found among adults with diagnosed MDD.

Even among ages 8-13, self reported depression symptoms correlated with RewP!

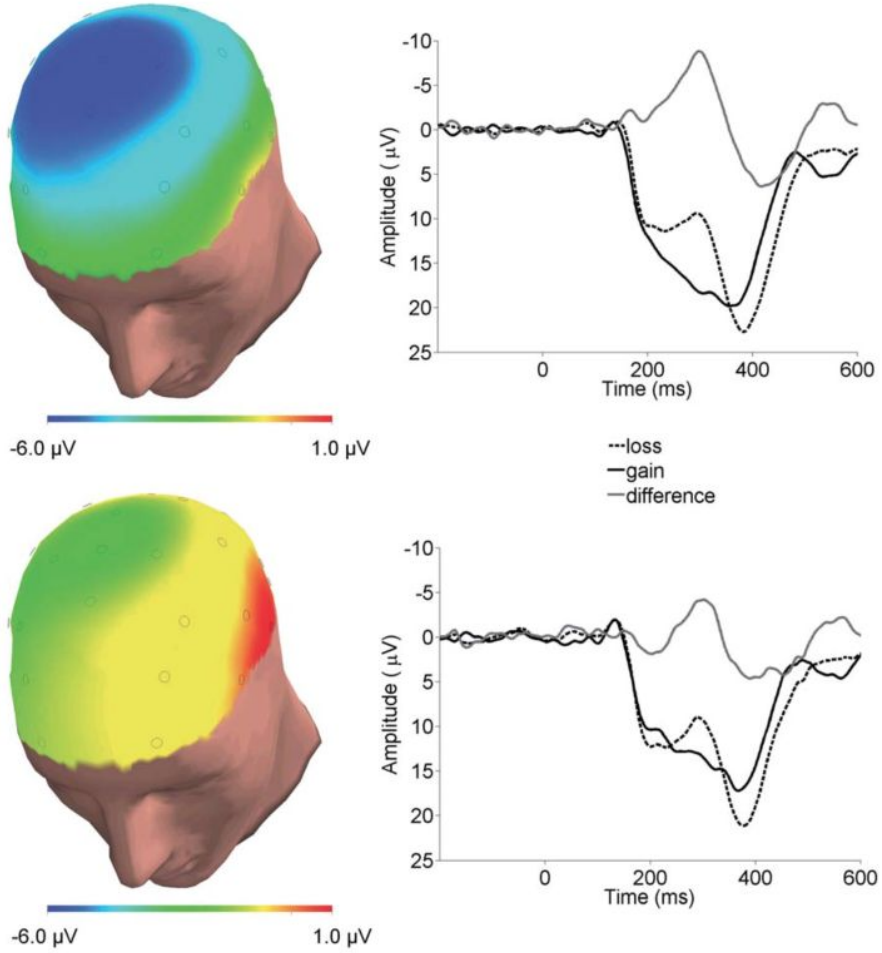
Possible depression predictor

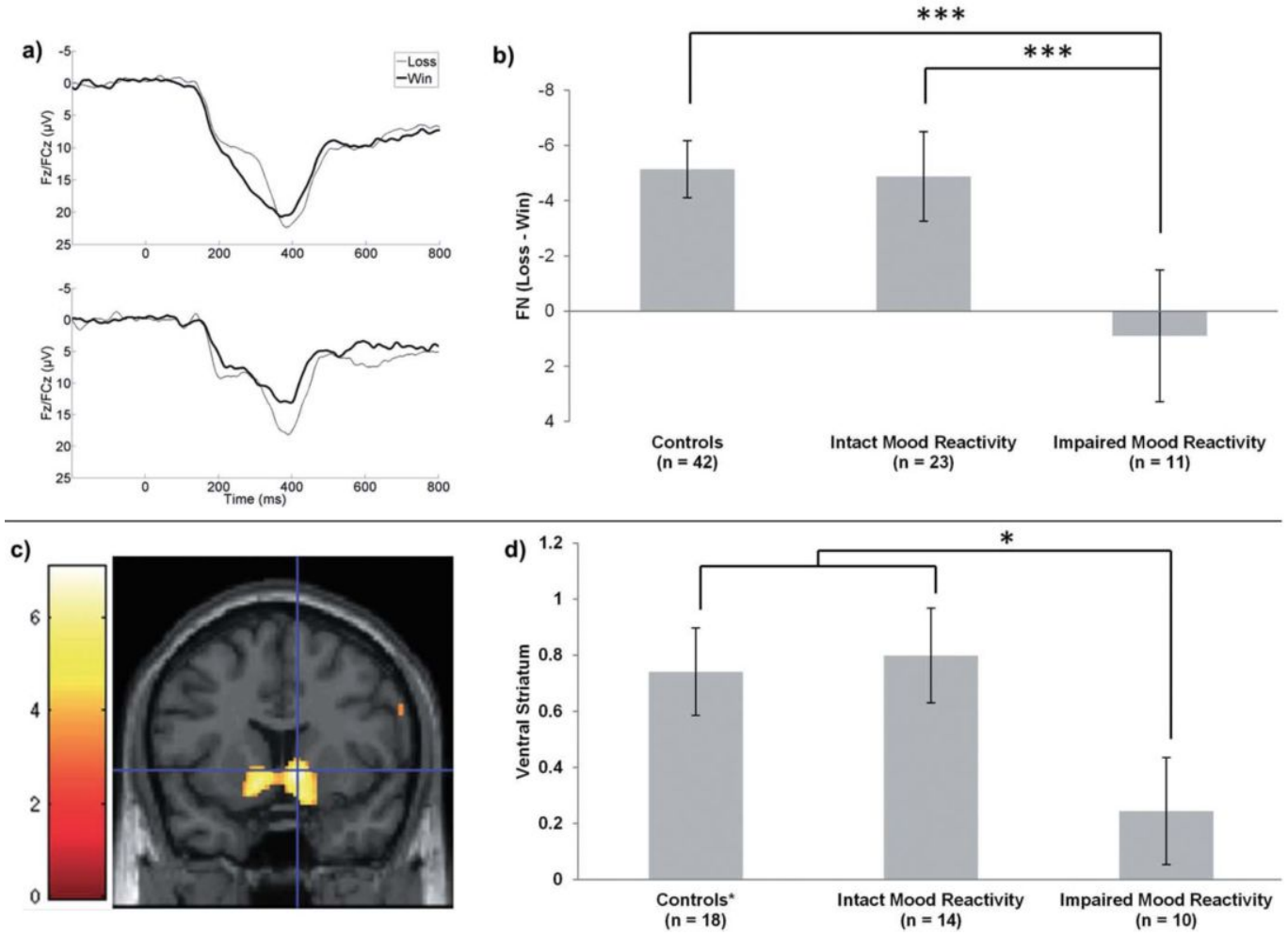
Longitudinal study of 407 never-depressed 9 year olds.

Maternal history of MDD is the strongest known predictor of MDD, and maternal history of MDD was found to be associated with a smaller RewP.

Further, RewP was most reduced among children who had the most severe maternal history of MDD.

Preliminary results suggest RewP predicts increases in depressive symptoms and the onset of depression - a reduced RewP was associated with increased depressive symptoms two years later in the same subjects


-6.0 μV
1.0 μV
-10
-5
0
5
10
15
20
25
Amplitude (μV)
0
200
400
600
Time (ms)
loss
gain
difference
-6.0 μV
1.0 μV
-10
-5
0
5
10
15
20
25
Amplitude (μV)
0
200
400
600
Time (ms)


a)
Loss
Win
Fz/FCz (μV)
Time (ms)
b)
***
***
FN (Loss - Win)
Controls
(n = 42)
Intact Mood Reactivity
(n = 23)
Impaired Mood Reactivity
(n = 11)
c)
d)
*
Ventral Striatum
Controls*
(n = 18)
Intact Mood Reactivity
(n = 14)
Impaired Mood Reactivity
(n = 10)

Bring it back around:

How do we measure learning?

(Recall)

When participants did not know stimulus-response mappings:

Negative feedback elicited the fERN while erroneous responses did not

When participants did know the stimulus-response mappings:

Response errors elicited an ERN while negative performance feedback did not

Bring it back around: how to measure learning

We’ve seen that as the mapping develops in the brain, RewP diminishes. RewP can be theorized to represent the “surprise” from an unexpected positive feedback. This is more like a “reward prediction error signal”.

In reinforcement learning, there is significant amounts of uncertainty involved. RewP can be used to measure the confidence in a person’s predictions and how their learning develops over time.

What is next for RewP?

Thanks!

Sources

Proudfit, 2015
Sutton and Barto, 1998
Holroyd & Yeung, 2012

http://www.enspire.com
http://springernature.com
https://wikimedia.org