Reversal of Fear Learning in the Human Brain

Daniela Schiller[1,2], Joseph E. LeDoux[1] and Elizabeth A. Phelps[1,2]

[1]Center for Neural Science, [2]Psychology Department, New York University

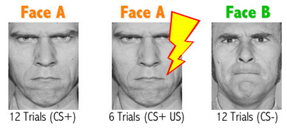
Introduction

- Fear learning is typically rapid and resistant to modification. This tendency to persist prevents the need for relearning about danger and can be adaptive in promoting avoidance in the face of danger.
- However, the ability to flexibly readjust behavior is also advantageous, and this ability may be impaired in fear disorders. One way of studying fear modification learning involves reversal of reinforcement contingencies.
- Research of the neural mechanisms of reversal learning has strongly implicated the orbitofrontal cortex across species (rats, monkeys and humans). However, these studies typically used reversal of appetitive reinforcement learning while the mechanisms underlying reversal of fear learning in humans are largely unknown.
- To investigate these mechanisms we used whole brain fMRI during a fear learning and reversal paradigm.

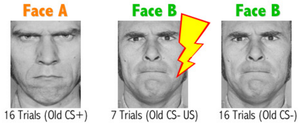
Experiment Methods

Simple visual discrimination and reversal, partial reinforcement paradigm

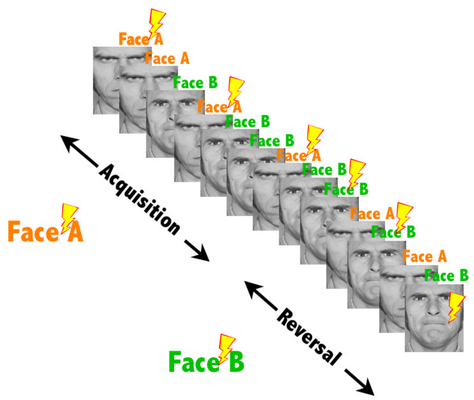
- **Acquisition**



- **Reversal**



Galvanic skin response (GSR) served as an index of FEAR



FMRI data acquisition

- 3 T Siemens head-only scanner
- 39 slices were obtained parallel to the AC-PC plane
- 3 mm isotropic voxels
- TR = 2000 ms; TE = 25 ms
- Event-related design

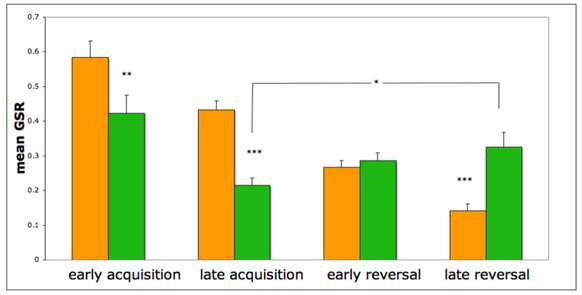
FMRI data analysis

- 17 participants
- The data were initially corrected for motion, spatially smoothed using a 3-D Gaussian filter (6-mm FWHM), temporally processed to remove scanner drift, slice scan time was corrected, and functional data was transformed into Talairach space to allow for group analyses.
- A random-effects General Linear Model was used to construct group maps.

Acknowledgements

This research was supported by NIH grant #P50 MH058911 to JEL, #MH62104 to EAP and a Fulbright award to DS

Physiological Results



Face A: Shocks → No shocks **Face B:** No shocks → shocks

- **GSR scores from each subject served as a predictor for brain activation during reversal learning**
- **A predictor for brain activation was also constructed based on the Temporal Difference (TD) Learning model:**

Prediction Error (PE):

$$\delta(t) \equiv r(t) + V(t+1) - V(t)$$

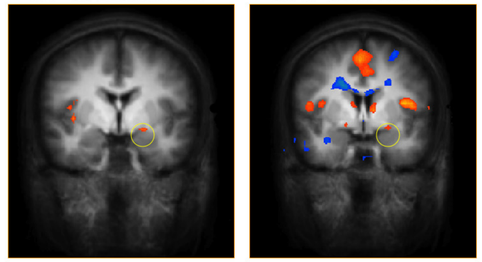
shock outcome (1 or 0); future expectations; current expectation (Vold)

Vnew = Vold + learning rate * PE

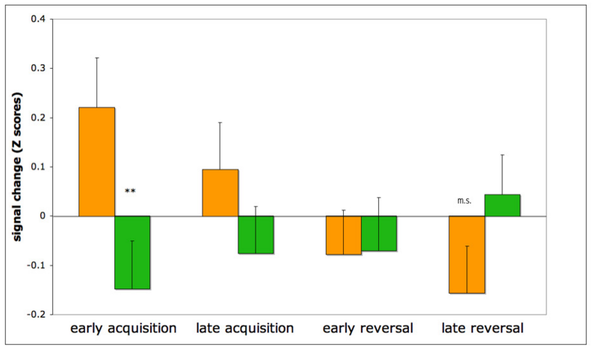
FMRI Results

Brain areas positively correlated with the physiological data were also correlated with the pattern predicted by the TD model :

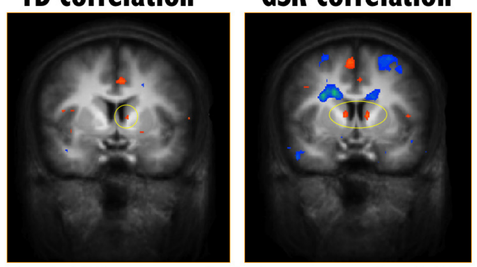
AMYGDALA



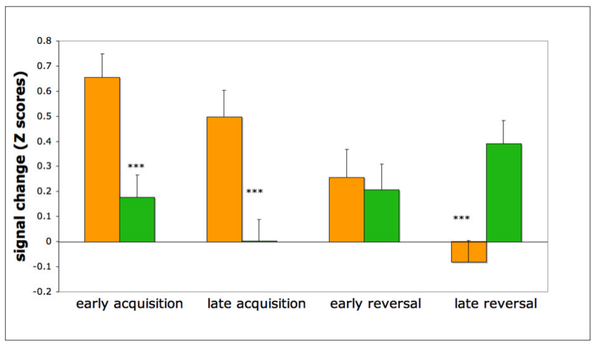
$P < 0.001$ -19, -1, -6 $P < 0.005$



STRIATUM

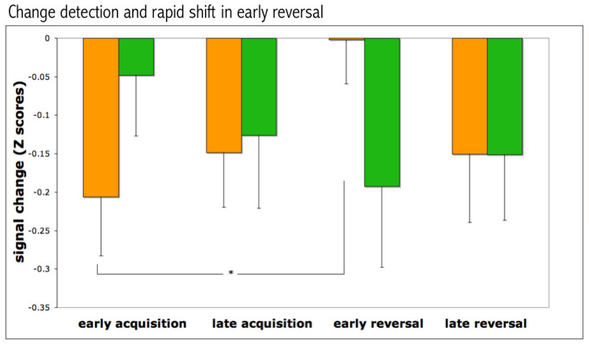


$P < 0.001$ -8, 2, 8 $P < 0.005$

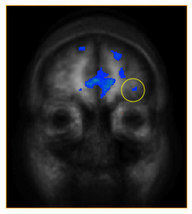


Brain areas negatively correlated with the physiological data:

Change detection and rapid shift in early reversal

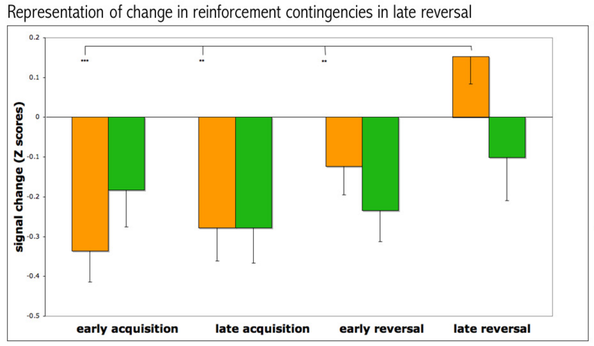


Lateral OFC

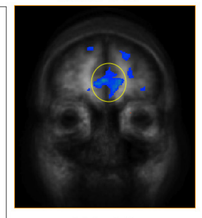


BA 11
-33, 50, 2
$P < 0.001$

Representation of change in reinforcement contingencies in late reversal



vmPFC



BA 10
5, 50, 6
$P < 0.001$

Discussion

- BOLD responses in the amygdala and the striatum represented the predictive aversive values of the conditioned stimuli. These responses were flexibly adjusted when reinforcement contingencies were reversed.
- The BOLD responses of these areas were positively correlated with GSR responding and more selectively and strongly with the pattern predicted by the TD model.
- Two areas in the prefrontal cortex (PFC) showed dissociable effects during reversal: the lateral orbitofrontal was the first to detect a change in reinforcement contingencies by quickly reversing its differential responding to the stimuli during early reversal. The ventromedial PFC represented this change in reinforcement contingencies in the late phase of reversal, and exhibited increased responding to the stimulus that was no-longer reinforced.

Conclusions

- **Our results suggest that the amygdala and the striatum promote fear learning and its reversal by updating the current aversive value of the conditioned stimuli based on prediction errors.**
- **Our findings also point to a specific role of the PFC areas in change detection and rapid reversal (lateral OFC), and possibly in providing a "reward" signal when cues cease to predict aversive outcomes (vmPFC).**

schiller@cns.nyu.edu