Neuroscience 345 (2017) 27–37

ORBITOFRONTAL CORTEX REFLECTS CHANGES IN RESPONSE–OUTCOME CONTINGENCIES DURING PROBABILISTIC REVERSAL LEARNING

L. R. AMODEO,[a] M. S. MCMURRAY[b] AND J. D. ROITMAN[a,c]*

[a] *Department of Psychology, University of Illinois at Chicago, 1007 West Harrison Street, Chicago, IL 60607, USA*

[b] *Department of Psychology, Miami University, 90 N Patterson Avenue, Oxford, OH 45056, USA*

[c] *Laboratory of Integrative Neuroscience, University of Illinois at Chicago, 1007 West Harrison Street, Chicago, IL 60607, USA*

Abstract—In a continuously changing environment, in which behavioral outcomes are rarely certain, animals must be able to learn to integrate feedback from their choices over time and adapt to changing reward contingencies to maintain flexible behavior. The orbitofrontal region of prefrontal cortex (OFC) has been widely implicated as playing a role in the ability to flexibly control behavior. We used a probabilistic reversal learning task to measure rats' behavioral flexibility and its neural basis in the activity of single neurons in OFC. In this task, one lever, designated as 'correct', was rewarded at a high probability (80%) and a second, spatially distinct lever, designated as 'incorrect', was rewarded at a low probability (20%). Once rats reached a learning criterion for reliably selecting the correct lever, reward contingencies of the two levers were switched, and daily sessions were conducted until rats reliably selected the new correct lever. All rats performed the initial Acquisition and subsequent Reversal successfully, with more sessions needed to learn the Reversal. OFC neurons were recorded during five behavioral sessions spanning Acquisition and Reversal learning. The dominant pattern of neural responding in OFC, identified by principal component analysis of the population of neurons recorded, was modulated by reward outcome across behavioral sessions. Generally, activity was higher following rewarded choices than unrewarded. However, there was a correlation between reduced responses to reward following incorrect choices and the establishment of the preference for the correct lever. These results show how signaling by individual OFC neurons may participate in the flexible adaptation of behavior under changing reward contingencies.

This article is part of a Special Issue entitled: Cognitive Flexibility.

Key words: cognitive flexibility, probabilistic reversal learning, orbitofrontal cortex, reward, electrophysiology.

INTRODUCTION

A critical cognitive faculty is the ability to flexibly adapt behavior as the consequences of our actions change in a dynamic environment. It has been demonstrated in a wide variety of species and settings that the orbitofrontal cortex (OFC) is critical for successful behavioral flexibility, when learned reinforced contingencies change without warning (Teitelbaum, 1964; Jones and Mishkin, 1972; Dias et al., 1996; Schoenbaum et al., 2002; Chudasama and Robbins, 2003; Izquierdo et al., 2004). While deficits in OFC function impair reversal learning, they do not typically impair acquisition of the initial discrimination (Ragozzino, 2007). Although there is debate over the basis of this impairment (Schoenbaum et al., 2011), OFC appears necessary to make shifts in associative learning. OFC lesions in both monkeys and rodents reveal deficits in discriminative performance when previously learned behavior is devalued, by treating the valued and devalued reward-cues similarly (Gallagher et al., 1999; Izquierdo et al., 2004; Machado and Bachevalier, 2007). Further, electrophysiological studies have shown that activity in the OFC flexibly encodes conditioned stimuli that predict outcomes, irrespective of specific cue features (Schoenbaum and Eichenbaum, 1995; Morrison and Salzman, 2009). This work has increasingly demonstrated that the OFC represents associative information, particularly information important for value-based decision-making (Murray et al., 2007; Wallis, 2007; Rangel et al., 2008; Padoa-Schioppa, 2011).

Typically, modeling the neural basis of behavioral flexibility with reversal learning uses reinforcement contingencies in which 'correct' choices are always rewarded, and 'incorrect' choices are not. However, in everyday situations, choice outcomes often carry a degree of uncertainty about whether or not the potential reward will be delivered. Probabilistic reinforcement increases the complexity of reversal learning, as both 'correct' and 'incorrect' choices can lead to reward or its omission, but with different probabilities of the two outcomes. Thus, the information from any given choice is insufficient to guide future behavior, and subjects must engage cognitive functions to track the reward history for both options to determine the overall more beneficial option.

*Correspondence to: J. D. Roitman, Department of Psychology, 1007 W Harrison Street MC 285, Chicago, IL 60607, USA. Tel: +1-312-355-1458.
E-mail address: jroitman@uic.edu (J. D. Roitman).

Abbreviations: ANOVA, analysis of variance; DMS, dorsomedial striatum; OFC, orbitofrontal cortex; PC, principal component; PCA, principal component analysis; PRL, probabilistic reversal learning.

http://dx.doi.org/10.1016/j.neuroscience.2016.03.034
0306-4522/

Deficits in this more complex form of probabilistic reversal learning (PRL) have been demonstrated in a variety of psychiatric conditions in both humans (Waltz and Gold, 2007; D'Cruz et al., 2013) and non-human animals (Amodeo et al., 2012), as well in patients with OFC damage (Tsuchida et al., 2010). The receipt of probabilistic rewards modulates individual OFC neurons in monkeys (O'Neill and Schultz, 2010) and rodents (Roitman and Roitman, 2010; McMurray et al., 2015) during decision-making tasks. These data suggest that the OFC may play a broader role in processing probabilistic reward signals to determine the values of available options. The goal of the present study was therefore to investigate how OFC represents probabilistic response–outcome associations as they change across task conditions. To that end, we recorded the activity of individual OFC neurons in rats across acquisition and reversal learning in a PRL task. We hypothesized that encoding of reward contingencies by OFC neurons would correspond with acquisition and be modulated across PRL.

EXPERIMENTAL PROCEDURES

Animals

Eight adult male Sprague–Dawley rats (Charles River, Chicago, IL, USA) weighing 300–350 g at the start of the experiment were individually housed in standard polycarbonate cages in a colony room maintained on a 12-h light/dark cycle (lights on at 7:00 h). Rats had at least one week of free access to standard rodent chow (LabDiet 5012, Richmond, IN, USA) prior to the start of the experiment. Immediately prior to experimental testing, food was restricted to maintain at least 90% of free feeding weight with *ad libitum* access to water. Behavioral testing was conducted in a separate experimental room during the light phase between 10:00 and 15:00 h. Subjects were cared for in accordance with guidelines set by the National Institutes of Health, with approval of the University of Illinois Chicago Animal Care Committee.

Surgical procedures

Rats were anesthetized with ketamine (100 mg/kg, i.p.) and xylazine (20 mg/kg, i.p.) and placed in a stereotaxic instrument. Stainless steel electrode arrays were custom-designed (Micro Probe, Gaithersburg, MD, USA) into two columns of four microwires (50-μm diameter; tip separation 0.25 mm spanning 1 mm). Arrays were implanted bilaterally in OFC (AP + 3.0, ML ± 3.2, DV −4.0 relative to the brain surface [Paxinos and Watson, 2007]). Silver ground wires for each array were inserted into the brain ipsilateral to electrode arrays, located roughly midway between bregma and lambda. Connectors for the microwire arrays were anchored to the skull via four stainless steel screws and dental acrylic. Electrode arrays were implanted at least 1 week prior to behavioral training and 2 weeks before the first recording.

Behavioral test chamber

Recording sessions were conducted in an operant chamber (30.5 × 24.1 × 21.0 cm; Med Associates, St Albans, VT, USA) housed in a sound-attenuating cabinet. Each operant chamber was equipped with a food receptacle mounted on one wall. Head entry into the receptacle to retrieve sugar pellets (45 mg sugar; BioServ, Frenchtown, NJ, USA) was measured via photobeam. On either side of the receptacle was a retractable lever with a cue light positioned directly above it. A house light was positioned at the top center of the opposite wall. Events were controlled (cue light illumination, lever extension, etc.) and monitored (lever press, head entry, etc.) by a PC computer using a commercially available software program (MED-PC; Med Associates). Such events were marked with TTL outputs that were time-stamped to enable temporal alignment of electrophysiological activity.

Probabilistic reversal learning

Training. Prior to surgery, rats underwent initial lever press training of daily sessions in which one lever was continuously available and each press resulted in the delivery of one sugar pellet (45 mg, Bio-Serve Precision Pellets; Frenchtown, NJ, USA). Each lever was presented on alternative days. Training continued until each rat met a criterion of at least 50 presses in 1 h for each lever, and typically lasted 5–7 days. After recovery from surgery, rats were habituated to the electrophysiology recording cable for three additional training sessions.

Acquisition. For each subject, the lever on one side of the food receptacle was designated as 'correct' and the lever on the other side was 'incorrect'. Throughout the experiment, the lever designated as 'correct' resulted in a higher rate of reward when pressed. The side of designation was counterbalanced across rats. During the Acquisition phase, rats performed one session per day, with 30 trials per session. Each trial began with the illumination of the house light. A nose poke into the center receptacle port caused the house light to extinguish and both cue lights to illuminate. Two seconds after cue light illumination, both levers extended. Eighty percent of 'correct' lever presses were rewarded with the delivery of one sugar pellet, and the remaining 20% of correct presses were not rewarded. Presses of the 'incorrect' lever resulted in omission of reward on 80% of presses and delivery of one sugar pellet on the remaining 20% (Fig. 1A). Following reward delivery or omission, a variable 10–15-s interval preceded the illumination of the house light to signal the beginning of the next trial. The criterion for successful learning in the Acquisition phase was pressing the correct lever on at least 80% of trials on three consecutive days.

Reversal learning. Reversal learning began the session after the criterion for Acquisition was met. In the Reversal phase, sessions were the same as the Acquisition phase,