Reward processing: a global brain phenomenon?

Andrew M. Clark

Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland

Submitted 23 January 2012; accepted in final form 15 July 2012

Clark AM. Reward processing: a global brain phenomenon? J Neurophysiol 109: 1–4, 2013. First published July 18, 2012; doi:10.1152/jn.00070.2012.—Rewards and punishments (reinforcement) powerfully shape behavior. Accordingly, their neuronal representation is of significant interest, both for understanding normal brain-behavior relationships and the pathophysiology of disorders such as depression and addiction. A recent article by Vickery and colleagues (Neuron 72: 166–177, 2011) provides evidence that the neural response to rewards and punishments is surprisingly widespread, suggesting the need for examination of the specific roles of areas not commonly included in the canonical reward circuitry in processing reinforcement.

Whether or not to engage in a particular behavior, opt for one course of action over another, or approach or withdraw from a particular situation is all strongly influenced by the outcomes previously experienced under similar conditions (Mazur 2012). That is, motivation, decision making, and social interaction are all dependent on expectations that a particular action, choice, or environment will yield a given result. In general, these returns, be they pleasant (rewards) or aversive (punishments), powerfully shape behaviors from the simple to complex, across timescales that range from milliseconds to years (Mazur 2012). This influence of rewards and punishments is disrupted in clinical brain disorders such as depression, addiction, and obsessive-compulsive disorder (Charney and Nestler 2009). Individuals with these conditions no longer seek to gain rewards and avoid punishments, compulsively seek particular rewards to the detriment of other needs, or attribute rewards to maladaptive behaviors, respectively (Charney and Nestler 2009). Thus, an understanding of the neurobiological representation of rewards and punishments is essential for understanding both the healthy and diseased brain.

Evidence for a dedicated neurobiological pathway representing reinforcement first came from experiments that found rats would lever-press to earn electrical microstimulation of brain sites previously identified as critical for seeking natural rewards (Olds and Milner 1954). Subsequently, it was demonstrated that a key feature of these sites, including the ventral striatum, amygdala, medial prefrontal cortex, and orbital prefrontal cortex was innervation from midbrain dopaminergic neurons (for a review, see Haber and Knutson 2010). Studies demonstrating significant responses to primary (e.g., food) and secondary (e.g., money) reinforcements and reinforcement-predictive stimuli in these areas, modulation of these neuronal responses by dopamine agonists and antagonists, and behavioral impairments following dopamine depletions served to cement the primary role of dopamine and its cortical and subcortical efferents in encoding and valuing rewards (for a review, see Schultz 2007).

Given the above evidence for a well-conserved and conserved pathway for encoding rewards, it is surprising that a recent study reported widespread encoding of reinforcement signals across multiple cortical and subcortical domains (Vickery et al. 2011).

In this study, Vickery and coworkers (2011) examined blood oxygen level-dependent (BOLD) responses using functional magnetic resonance imaging (fMRI) while human subjects played a simple competitive game, matching pennies or rock-paper-scissors, against a computer opponent. During these experiments, the subject’s task was to guess the computer’s choice, either heads or tails in matching pennies, or rock, paper, or scissors in rock-paper-scissors. In the matching pennies game, guesses that matched the computer’s selection were counted as wins and mismatches were counted as losses. Subject-computer choice relationships in rock-paper-scissors determined wins, ties, and losses (paper > rock > scissors > paper). In matching pennies, any block of trials that ended with the subjects having answered correctly on greater than 50% of the trials resulted in a monetary reward. Payouts in rock-paper-scissors were earned whenever subjects won more than they lost over a block of trials. In both games, subjects used previous outcomes to guide future choices, exhibiting a greater tendency to switch their response following a loss on the previous trial and stay with the same response following a win on the previous trial (“win-stay lose-shift” strategy).

To examine whether fMRI responses in a region of interest (ROI) encoded the outcome of the previous trial, Vickery et al. (2011) used two separate methods, a standard generalized linear model (GLM) and a multivoxel pattern analysis (MVPA), to decode previous trial outcomes during an intertrial period. Unlike GLM, in which classification of the outcome is based on a comparison of the average activity following wins and losses, MVPA uses the pattern of activity across the voxels in an ROI to determine whether the previous trial was a win or loss (Fig. 1A). This technique has previously been used, for example, to decode subjective perception of visual stimuli from activity in primary visual cortex (Kamitani and Tong 2005).

Even under the strictest statistical criterion, the MVPA model performed significantly better than chance in decoding win-loss information in an approximately fourfold greater number of areas than the GLM model (greater than 80% of the areas studied). This suggests the more limited representation of reinforcement described in previous imaging studies (for a review, see Haber and Knutson 2010) might have been a result of only looking for differences in the mean signal. To examine
the overlap between reward and punishment signals, separate MVPA models were used to classify wins vs. ties and ties vs. losses in the rock-paper-scissors game. Most areas that encoded wins vs. ties also encoded ties vs. losses, leading the authors to suggest that reinforcement and punishment are encoded within the same circuit. A notable exception to this pattern was the nucleus accumbens, an area previously identified as a key node in the reward circuit and hypothesized to implement an important step in leading computational models of reinforcement learning (for a review, see Houk et al. 1995), which did not encode losses. More generally, this finding also conflicts with studies reporting a segregation of responses to reward and punishment (Matsumoto and Hikosaka 2007).

Finally, two important control analyses were run to ensure that these results were not an artifact of a particular statistical model and could not be attributed to emotional or cognitive factors that co-vari with outcome. First, to ensure that the widespread encoding of reinforcement signals was not simply an artifact of an overly sensitive method, MVPA was also used to decode the visual, feedback, motor response, and decision strategy signals from activity during the feedback, response and inter-trial periods, respectively. In all cases, significant encoding was restricted to a small subset of ROIs (fewer than 15% of those examined) in which, based on a large corpus of evidence, each type of signal would be expected to be observed (e.g.,...
visual feedback could only be decoded from activity in visual cortex).

Second, to determine the extent to which the signals observed during the inter-trial period reflect reinforcement and not other variables such as arousal, separate MVPA classifiers were constructed using only win and tie outcomes or tie and loss outcomes in the rock-paper-scissors game. Assuming that wins and losses are more emotionally or cognitively “salient,” i.e., elicit greater happiness or frustration, respectively, or grab more attention, than ties, the tendency of each classifier to assign untrained losses (for the win-tie classifier) or wins (for the tie-loss classifier) as a win or a loss can be taken as an indication of the extent to which these signals reflect outcome salience rather than reinforcement. Because neither of these models showed a significant tendency to misclassify signals from any ROI in this manner, Vickery et al. (2011) concluded that these signals do not reflect salience.

Current theories of reward-guided behavior posit that sensory, motor, and mnemonic information is integrated with information about reinforcement in a specialized pathway for valuing and selecting rewards (Fig. 1B) (for a review, see Haber and Knutson 2010). The work of Vickery and colleagues (2011) suggests at the very least two alternative interpretations, namely: 1) reward is encoded far more broadly than previously hypothesized, or 2) processing within this specialized reward pathway dramatically influences subsequent processing throughout nearly the entire brain. Given the aforementioned importance of reinforcement to understanding normal and pathological behavior, it is crucial to critically evaluate the results of Vickery et al. to determine the most promising avenue for further research.

While the results of the control analyses examining visual feedback, response, and decision signals are strong evidence that the ubiquitous encoding of trial outcomes observed during the inter-trial period was not simply due to the sensitivity of the MVPA method, there remain several open questions that must be answered to better interpret the central findings of Vickery et al. (2011).

The results of the separate win-tie/tie-loss classifier analysis are strong evidence that these signals do not reflect the degree of emotional arousal or attention elicited by a particular outcome. However, when planning future choices, ties are potentially as strategically salient as wins and losses (i.e., just as informative about what to do next). The above analysis does not control for salience in this latter, informational sense. Furthermore, the manner in which subjects used this information might have been dependent on their knowledge that they were playing against a computer rather than human opponent. A recent study reported encoding of socially dependent strategies in many of the core areas of the canonical reward pathway (Bault et al. 2011). Examining whether a subject’s task strategies and the pattern of neural encoding of outcomes change when games are played against human vs. computer opponents could shed further light on the degree to which the signals described in Vickery et al. (2011) are dependent on these factors.

In a similar vein, while competitive games such as matching pennies or rock-paper-scissors are attractive, in that reinforcement is not dependent on a particular stimulus or response, and previous outcomes guide future choices, the relationship between any individual outcome and the ultimate payout (determined at the end of a block of trials) is complex and temporally variable. Thus, the widespread encoding of reinforcement seen in these tasks could be related to either task complexity or temporal distance from actual reward. Either 1) providing real-time feedback on the overall probability of reward as subjects advance through a block of trials, 2) examining whether distance from payout significantly affected outcome encoding, or 3) employing a simpler task, for example, one in which each choice is directly rewarded or punished, would be interesting manipulations to test in further studies.

Studies examining the relationship between BOLD and underlying neural activity have generally reported that BOLD reflects synaptic input and/or local processing within an ROI (Logothetis et al. 2001). However, a recent paper combining intrinsic signal optical imaging (measuring a similar hemodynamic response to the BOLD signal) with multi-unit and local field potential recordings in non-human primate primary visual cortex reported a striking dissociation between electrophysiological and hemodynamic responses to expected reward (Sirotin and Das 2009). Specifically, Sirotin and Das (2009) found a significant increase in the hemodynamic signal, but no corresponding increase in any of their electrophysiological measures, when monkeys anticipated upcoming trials (and thus rewards). These authors suggested that this hemodynamic modulation could be the result of dopamine-driven changes in vascular tone in anticipation of upcoming neural processing. Because subjects might have been anticipating future outcomes during the time period that Vickery et al. (2011) reported widespread reinforcement encoding, these signals could reflect anticipatory changes in local cerebral blood flow that are uncoupled from the neural activity in their ROIs. Largely confirming these effects, with obvious differences in spatial resolution, with a more direct measure of neural activity (e.g., MEG or EEG), would at least provide assurance that the BOLD signals measured here reflect underlying neural activity in these conditions (although this still might not reflect the spiking output of these regions). Conversely, differences in BOLD and electrophysiological responses in such an experiment would strongly suggest that the signals reported by Vickery et al. (2011) reflect changes in local processing, perhaps an increase in the sensitivity of these circuits, driven by activity in distant structures.

Finally, in future studies, it would be helpful to correlate the performance of MVPA models like those used here with estimates of functional connectivity (Fox et al. 2006). This would allow for the determination of whether reinforcement signals are stronger in particular networks. Such clustering data would allow for the targeted investigation of signal flow within these circuits by, for example, using reversible inactivations, focal lesions, and disconnections in primate and rodent models to determine whether, and under what conditions, the reinforcement signals found in specific brain regions are necessary for normal reward-guided behavior.

The greater variety of techniques available for recording and manipulating neural activity, and concomitant behavioral output, in animal models can be deployed to great use in extending the work of Vickery et al. (2011). Indeed, these authors’ work is in agreement with accumulating evidence in both rodents and monkeys that reinforcement signals are processed more widely than previously thought. For example, it has been reported that a large fraction of units in rodent primary visual...
cortex (V1) signal the magnitude or timing of an expected reward in the absence of any visual stimulation (Shuler and Bear 2006). Similarly, a recent study found that removal of macaque rhinal cortex (Rh), the final stage of the ventral stream of visual processing in the primate, induced a severe impairment in reward valuation in conditions in which no visual stimuli predicted reward value (Clark et al. 2012). Importantly, both V1 and Rh showed evidence of reinforcement processing in Vickery et al (2011).

In conclusion, maximizing rewards while minimizing losses is essential for survival. Thus, widespread abstract reinforcement signals, unrelated to any particular stimulus, response, or choice, could potentially be extremely useful not only in finding food and avoiding predators, but also, for example, in deciding whether specific information or a particular episode is valuable enough to encode in memory. It will be necessary to determine whether, as has been thought for more than fifty years, such signals are exclusive to dedicated pathways for processing reward and punishment, or whether wider networks of sensorimotor and association areas participate directly in encoding rewards and punishments.

REFERENCES


