Notes: Dopamine reward prediction-error signaling: a two-component response

Daniel Saunders

September 15, 2018

1 Introduction

Rewards induce behaviors that enable animals to obtain objects necessary for survival (e.g., food). "Reward" is commonly associated with happiness, but the scientific term has three functions:

1. They can act as positive reinforcers to induce learning
2. They elicit movement towards desired objects and constitute factors to be considered in economic decisions; value per individual decision maker is subjective and can be formalized as *economic utility*
3. It is associated with emotions such as pleasure and desire (difficult to measure in animals, whereas the first two can be quantitatively assessed using behavioral tasks)

Electrophysiological studies have identified individual neurons that signal reward-related information in certain brain areas. These reward neurons process specific aspects of rewards (amount, probability, subjective value, ...) in forms suitable for learning and decision-making. Most reward neurons show brief, phasic responses to rewards and reward-predicting stimuli. These responses code a *temporal reward prediction error*, reflecting a difference between received and predicted rewards at each moment in time. This *fast* dopamine signal differs strongly from slower dopamine activity increases (reflecting reward risk or behavioral reactivity) and from the tonic dopamine level needed to enable various neural processes.

Despite evidence for their involvement in reward coding, recent research has shown that some dopamine neurons show phasic activity increases in response to non-rewarding and aversive stimuli. This doesn't rule out a role for phasic dopamine release in reward processing; however, the extent to which phasic dopamine responses code rewards vs. non-reward information is difficult to resolve.

Recent studies have encouraged a revision of our views on the nature of the phasic dopamine response. These:

- Demonstrate distinct subcomponents of the phasic dopamine response
- Provide an alternate explanation for activations in response to aversive stimuli
- Document strong sensitivity to some unrewarded stimuli

The author suggests the reward prediction-error response be considered a *utility prediction-error signal*.

2 Processing of reward components

2.1 Reward components

Rewards consist of distinct *sensory* and *value* components. Rewards first impinge on the body through their physical sensory impact, and draw attention via their *physical salience*, which facilitates initial detection. Comparison with known objects determines reward novelty, drawing attention via *novelty salience* and *surprise salience*. Value is the feature that distinguishes rewards from other stimuli; it can be estimated from behavioral preferences elicited in choices; it provides *motivational salience*. The various forms of salience induce stimulus-driven attention, which selects information and modulates neural processing. This leads to internal decisions and overt choices to actions towards the chosen option and feedback that updates the neural representation of a reward's value.

2.2 Box 1: Fast, slow, and tonic dopamine functions

How can dopamine be involved in such diverse processes as movement, attention, cognition, and motivation? Some answers may lie in the different time scales across which it operates.

At the fastest timescale (subsecond), dopamine neurons show a two-part, phasic prediction error response that (apparently) transitions from salience and detection to reward value. This response constitutes a highly time-specific neuronal signal that is capable of influencing other fast neuronal systems involved in fast behavioral functions.

At a timescale in the second to minute range, a wide variety of brain functions are associated with slower changes in dopamine levels, which are unlikely to be driven by subsecond changes in dopamine impulses, and thus may be unrelated to reward prediction error. Their function may be to *homeostatically* adjust the sensitivity of the fast, phasic dopamine reward responses.

At the slowest timescale, dopamine exerts an almost tonic (movement, attention, cognition, motivation) influence on postsynaptic structures. Several diseases are associated with deficits in the tonic, finely regulated release of dopamine, which enables the function of postsynaptic neurons that mediate movement, cognition, attention, and motivation.

Altogether, dopamine neurotransmission exerts different influences on neuronal processes at different timescales.

2.3 Sequential processing in other systems

Research in sensory, cognitive, and reward systems recognize the component nature of stimuli and objects. Although simple stimuli are processed too rapidly to reveal their dissociable components, more sophisticated events take longer to identify, discriminate, and valuate. Examples: *visual search tasks* (50-120ms to discriminate targets from distractors), perceptual discriminations between partly coherently moving dots (activity in certain brain regions distinctive after 120-200ms after initial detection); temporal evolution from grossly tuning for stimulus properties to more finely tuned discrimination responses. These results demonstrate a sequential processing flow, but typical feature-selective neurons in the IT cortex process specific stimulus properties of visual objects at the same time they detect / identify them. However, in general, many sensory / cognitive neurons process different components of complex stimuli in consecutive steps.

Processing a reward requires an additional valuation step. Initial sensory response, follow 60-300ms later by a separate value component; thus, neural processing of rewards may require sequential steps.

2.4 Sequential processing in dopamine neurons

Phasic dopamine reward prediction error responses show temporal evolution with sequential components. An initial, short latency and duration activation of dopamine neurons detects objects before IDing and valuing it. The subsequent evolving response properly IDs and values the object in fine gradations.

3 The initial component: detection

3.1 Effective stimuli

The initial component of the reward prediction-error response is a brief activation that occurs unselectively in response to a large variety of unpredicted events; e.g., rewards, reward-predicting stimuli, unrewarded stimuli, aversive stimuli, ... This initial activation is sensitive to the time of stimulus occurrence and thus codes a temporal-event prediction error. The *unselective* and *multisensory* nature of the initial activation corresponds to the large range of possibly rewarding stimuli in the environment.

3.2 Sensitivity to stimulus characteristics

Several factors can enhance initial dopamine activation. Stimulus of sufficient intensity elicit the initial dopamine activation in a graded manner, regardless of positive / negative value associations. Weaker stimuli elicit little or no initial dopamine activations, and will only induce dopamine activation if they're rewards or associate with rewards.

The context in which a reward is presented can enhance initial dopamine activation. Unrewarded stimuli elicit little activation when well separated from reward; however, the same stimuli effectively elicit dopamine activation when presented in the same context as a reward. Neurons might be primed by a rewarded context and process every un-IDed stimulus as potential reward until the opposite is proven.

The resemblance of a stimulus to other stimuli associated with rewards can enhance initial dopamine activation through generalization. This is analogous to behavior generalization, in which "neutral" stimuli elicit similar reactions to similar target stimuli.

Novel stimuli can enhance dopamine activation; the response of dopamine neurons to an initally novel stimuli decreases with stimulus repetition. However, physically weak novel stimuli fail to induce a dopamine response. Novelty detection requires ID to compare with known stimuli, as does generalization.

Stimuli of high intensity are potential rewards and should be prioritized. Even the early stages of dopamine detection response are geared towards rewards.

3.3 Salience

The factors that enhance initial dopamine activation are closely related to different forms of stimulus-driven salience. The mechanism by which salience induces intitial dopamine response may apply mostly to rewarding stimuli, because the negative value of stimuli is unlikely to induce dopamine activation. Distinctions between different forms of salience may be important because they're thought to affect difference aspects of behavior.

3.4 Benefits of initial unselective processing

It might be assumed that unselective responses constitute inaccurate neural signals prone to erroneous behavioral reactions. It is in fact sensitive to several factors related to potential reward availability. The wide and multisensory sensitivity of the response facilitates the detection of a maximal number of potential rewards. The early dopamine activation component may serve to transiently enhance the ability of rewards to induce learning and action. This is formalized in the attentional Pearce-Hall learning rule, in which suprise salience derived from reward prediction errors enhances the learning rate, as do physical and motivational salience. By conveying different forms of salience, the initial dopamine response might boost / sharpen subsequent reward value processing and increase action accuracy.

The initial dopamine activation might provide a temporal advantage by inducing preparatory processes that lead to faster behavioral reactions to important stimuli. Since the response is faster than most behavioral reactions, there is still time to cancel behavioral initiation processes if subsequent valuation of a stimulus labels it worthless or damaging.

It is suggested the the initial dopamine response component affords a gain in speed and processing without substantially compromising action accuracy, which supports the function of the phasic dopamine reward signal.

4 The main component: valuation

The dopamine reward prediction-error response evolves from unselective stimulus detection into increasingly specific ID / valuation of stimulus. The latter component defines the function of the phasic dopamine response / reflect evolving neural processing needed to valuate the stimulus. Higher-than-predicted rewards induce brief dopamine increase, lower-than-predicted rewards induce decrease, and accurately predicted rewards don't change the activity. These responses constitute a bio implementation of the error term for RL according to the *Rescorla-Wagner model* and *temporal difference learning models*.

4.1 Subjective reward value

Value is a theoretical construct used to explain learning + decision making; construction of reward value involves brain mechanisms mediated by dopamine neurons.

Though dopamine responses increase w/ expected reward value, it's unclear whether dopamine neurons code for objective / subjective reward value. Comparison between rewards that are objectively equal show subjective value; preferences for risky vs. safe rewards w/ identical volume suggest increased subjective value. Another way to estimate: choose b/w reward in question and a reference reward. Dopamine neurons show higher activation due to preferred reward; activity correlates w/ indiff. amounts in choices b/w risky / safe rewards, indicating neurons code subjective rather than objective reward value.

Rewards lose subjective value after delays, though physically unchaged. Although initial comp. of dopamine response to a stim. that predicts a delayed reward stays constant, second comp. decreases as delay increases.

4.2 Utility

Econ. utility provides a constrained + principled def. of subjective value of rewards. Utility has the important potential to provide an internal / private metric of subjective reward value to individual decision makers. Subjective value derived from choice prefs. / indiff. provides a measure in physical units, but doesn't tell us how much a physical unit of ref. reward is privately worth to decision makers.

A private / internal metric of reward value would help to establish a neural reward function. This would relate the freq. of action potentials to internal reward value that matters to the decision maker. The number of APs after a stim. would quantify how much the reward is valued by neurons -¿ how much it's worth privately.

Econ. suggests numeric estims. of utility can be obtained experimentally in choices involving risky rewards. To obtain util. functions, we can use structured choices b/w gambles and variable safe rewards, and estim. "certainty equivalents", the amount of safe reward needed for the agent to select the reward as often as the gamble. The CEs are used to construct util. fns. A monkey's / human's choices reveal nonlinear utility / neural value fns. It's thus possible to estim. numeric econ. util. fns. that are suitable for corrs. w/ numeric neural reward responses.

Dopamine neurons show initial / uniform detection response to reward onset, unaffected by reward amount, and the 2nd response comp. increases monotonically w/ final amount -¿ signals value. The 2nd comp. increases nonlinearly with reward amount; gradually -¿ steeply -¿ gradually. So, the fully-evolved dopamine response corrs. w/ util.

To determine whether dopamine response codes for util., tests should use well-def.'d gambles that sat. conds. for util., rather than unpredicted rewards where risk is poorly def.'d.

Since the full dopamine reward prediction-error response codes for util., the phasic dopamine reward pred.-error response can be spec.'d as a util. pred.-error signal. Dopamine responses seem to repr. a phys. corr. for util.

5 Downstream influences

5.1 Correct behavior based on late component

The 2nd response comp. persists through behavior until reward is received, as revealed by graded positive pred.-error response to reward deliv. This response is large w/ intermediate reward prob. -¿ gens. intermed. value pred., decreases progressively when reward-prob. predictions lead to less-surprising rewards. Despite the transient / inaccurate 1st dopamine comp., the quickly following 2nd comp. allows neurons to distinguish rewards from non-rewards early enough to change behav. reactions.

The two-comp. mech. operates on a 10ms timescale that reqs. precise processing. Any changes in temp. structure of phasic dopamine response might distub the valuation comp. -¿ lead to impaired post-synpatic processing of reward info.

5.2 Updating predictions and decision variables