Dopamine and self-directed learning

Seth HERD [a,1], Brian MINGUS [a] and Randall O'REILLY [a]
[a] *Department of Psychology and Neuroscience,*
University of Colorado,
345 UCB,
Boulder, Colorado 80309

Abstract.
Humans are intrinsically motivated to learn. Such motivation is necessary to be a human-like learner, and helpful for any learning system designed to achieve general intelligence. We discuss the limited existing computational work in this area, and link them to known and theorized properties of the dopamine system. The relatively well-understood mechanisms by which dopamine release signals unpredicted reward can also serve to signal new learning. Dopamine release leads to maintenance of current representations, which serves to "lock" attention onto topics or tasks in which useful learning is occurring. We thus propose a novel but natural extension of known aspects of dopamine function to perform self-directed learning of arbitrary self-defined tasks. If this hypothesis is correct, detailed experimental evidence on dopamine function can help guide computational research into human-like learning systems.

Keywords. Motivation, Learning, Neural Network

Introduction

Children play not to learn but because play is fun. They play with things and in ways they find interesting, and cease once they become boring. Evolution, on the other hand, has designed children not to have fun, but to learn. Their pattern of play reflects an evolved tendency toward maximizing learning opportunities. Understanding our ability to efficiently self-direct learning is likely to be crucial for understanding and reproducing human intelligence.

Building a biologically inspired cognitive architecture (BICA) is intended to capitalize on the only example of a generally intelligent system we have available: the human brain. Similarly, there is only one example of a training set that produces general intelligence: that selected by the human learner from its natural environment. Designing a BICA as a human-like learner serves not only to make that agent more accessible to humans, but follows the only known working path to general intelligence.

While the dopamine system has been previously hypothesized to play a role in self-directed learning, we propose a more specific and sophisticated relation

[1]Corresponding Author: E-mail: seth.herd@colorado.edu

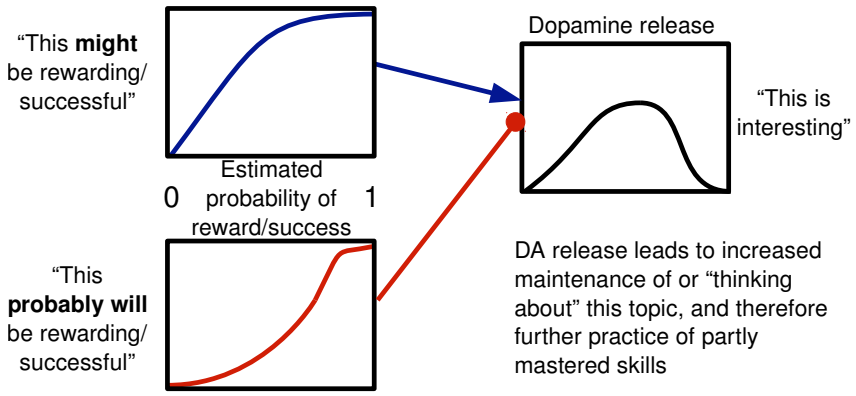

Figure 1. Self directed learning as an outcome of dopamine system function. The dopamine system's known property of signaling only unpredicted rewards could allow it to signal activities which are neither too easy nor too difficult for the learner's current abilities. An excitatory system learns "optimistically" to respond when reward is possible, while an inhibitory system learns more slowly or "pessimistically" when reward is likely. Dopamine release is approximately the difference in activation of the two systems. The phasic release of dopamine then serves to "lock" attention to the current topic, encouraging focused practice on tasks where success is possible but not certain.

between DA release and attentive learning. We work from the hypothesis that success at any physical or mental task acts as a reward, and show how known dopamine system mechanisms should then produce self-directed learning behavior much like, but importantly different from, existing computational approaches (Figure 1). We show how not only the antecedents but the consequences of DA release suit the dopeamine system for a crucial role in self-directed learning.

1. Dopamine function and self-directed learning

The best developed current theory of self-directed learning (or "intrinsic motivation") proposes a heuristic of increasing predictability [1]. This principle, dubbed Intelligent Adaptive Curiosity (IAC), directs the learner toward activities in which it is currently on a steep part of a learning curve. In essence, the system predicts the outcomes of its actions, and keeps a record of the quality of those predictions. Domains in which predictions have recently improved are probably those in which significant learning has occurred. The system therefore chooses action domains stochastically but based on that metric.

Success at arbitrary laboratory tasks seems to act as a reward and trigger dopamine release in humans [2]. This could result from an innate reward signal from successful prediction, from success being strongly predictive of primary rewards (one intriguing hypothesis is that humans find positive regard from other humans intrinsically rewarding [3]), or for other reasons. Here we remain agnostic

on the particular underlying cause, and focus on the consequences of the general hypothesis.

If success at a self-defined task results in a reward signal, the dopamine system appears to have the right properties for producing something similar to the IAC algorithm described above. The relatively well understood mechanisms by which dopamine directs learning and memory toward reward and reward-predictive events may also serve to direct attention toward tasks and topics for which learning is currently happening rapidly. We first discuss how DA acts to direct learning toward unpredicted reward predictors, then show how that function can generalize to self-direct learning on arbitrary sensorimotor or cognitive tasks.

2. Dopamine release directs learning to unpredicted reward predictors

Expected rewards are "discounted" by the dopamine signal: after sufficient learning, a predictable reward causes little to no DA release. The Temporal Differences (TD) algorithm [4] has been widely used to model this effect. Physiologically, the absence of DA release for a well-predicted reward probably results from an active cancellation (evidence reviewed in [5,6]). An excitatory system pushes for DA release and learns quickly, while an inhibitory system learns more slowly (or conservatively - the role of uncertainty in guiding behavior [8] is an important one that we do not address here). The overall result is that, early in learning, a predictable reward causes phasic DA release; later in learning, that release is canceled by the slower-learning inhibitory system (as illustrated in Figure 1).

After sufficient learning, phasic DA release is instead triggered by the stimulus that predicts the upcoming reward. In classical experiments, a light or tone that reliably signals an upcoming food reward will trigger dopamine release after sufficient learning. Dopamine thus signals newly predicted rewards, while remaining silent for rewards that have been previously predicted.

Phasic DA release enhances learning. For instance, playing a tone just before inducing a phasic DA release dramatically increases the area of cortex in which neurons respond to that tone in the future [9]. The combination of signaling new reward predictions and directing learning allows the dopamine signal as currently understood to perform a relatively weak type of self-directed learning, as outlined below.

This attentional focus causes the organism to learn about events immediately preceding reward; e.g., an infant learns which motor commands scoop applesauce into its mouth, producing a primary reward of nourishing sugar. But this focus on the moment of reward becomes counterproductive. Once those motor commands are mastered, it becomes useful to learn about the conditions surrounding the predictor of reward (in this case, the jar of applesauce). As the success of those motor movements becomes predictable, the dopamine spike at the moment of reward is cancelled by the inhibitory component of the system. At the same time, any stimulus that reliably predicts reward (in the example, the applesauce jar) starts to produce phasic DA release. This moves the focus of learning toward the next step in the causal chain. The infant now learns to reach or ask for the applesauce jar.

While the TD algorithm has worked well to explain how similar chains of learning are performed in laboratory tasks, its reliance on temporal rather than semantically associative chaining limits its applicability to rich and varying environments. Brain mechanisms for a similar, but more general algorithm (one that chains semantically rather than only temporally) are reviewed in [6].

In sum, the DA signal is thought to provide a simple form of self-directed learning that enhances learning to important (reward-predicting) events. The DA system may also have been co-opted by evolution to produce a more flexible and powerful form of self directed learning, as discussed below.

3. How the DA system can signal opportunities for learning

Our main novel proposal is that the same set of mechanisms described above could act to direct attention toward activities for which average success has recently become greater – those which are being learned rapidly. In the better-researched case of laboratory tasks and rewards, attention is directed toward predictors of reward that are themselves poorly predicted (e.g., the unexpected appearance of an applesauce jar). Supposing that self-defined success at any task acts as a reward has interesting consequences.

We deliberately use a very general supposition: some close correlate of successful performance (e.g., accurate prediction) acts as a reward from the perspective of the DA system. For instance, when an infant successfully places one block atop another, its reward circuits fire. Treating an arbitrary happening or concept as a reward could be crucial for human cognition [7], allowing us to work toward abstract concepts like "money" and "trustworthiness" as well as concrete rewards like food and shelter. Successful performance of an arbitrary task is one such reward-substitute with particularly important consequences.

If success counts as a reward, the reward discounting properties of the DA system become useful in directing learning toward new successes. When successful skill execution is fully predictable, as in a task that's been mastered, the same circuits that cancel DA release for physical rewards, cancel that for successful prediction. Thus frequent phasic DA release indicates a task that's sometimes successful but not yet mastered. Thus the DA system directs learning toward new tasks, exactly as it directs learning toward new steps in a causal chain leading to physical reward.

4. How dopamine release can direct learning

There is a second important reason to favor dopamine release as a mechanism for self-directed learning. Not only does the DA system have the right properties to signal a useful learning opportunity, but DA will direct attention and therefore learning to whatever is happening when it's released. Dopamine has a relatively well understood role in working memory (reviewed in [10]). Working memory function is also thought to be central to cognitive control [11,12].

In essence, sustained firing of neurons in prefrontal cortex and elsewhere is thought to be the basis of working memory. In the terms of the biased competi-

tion model [13], working memory representations are strategically maintained biases that direct attention toward appropriate items. These maintained representations can also act to direct attention toward topics, tasks, or stimulus-response mappings by acting as one constraint in a brain-wide constraint satisfaction [14].

Phasic dopamine release and tonic dopamine levels (resulting from frequent recent phasic releases) both play a role in working memory maintenance. At normal levels, increased tonic dopamine levels in cortex increase the maintenance of information [15]. Phasic dopamine release also biases the striatum toward "Go" over "NoGo" decisions, one likely consequence of which is the maintenance of information currently represented in associated regions of prefrontal cortex [16].

Both of these factors make dopamine release tend to "lock" the current topic of thought in mind. If a hungry animal sees food that's not immediately obtainable, it will tend to keep maintaining that representation, and so guide its behavior toward obtaining it. Similarly, if a child performs above its predictions at a particular block-stacking task, it will tend to keep thinking about it and so performing similar tasks until they become too predictably successful (or unsuccessful).

5. Summary and Conclusions

The logic above describes how humans may have adapted the evolutionarily old dopamine system to enhance general learning. The system evolved to direct learning toward progressive events in a causal chain leading to reward. The same basic mechanisms can direct learning toward learning itself with the simple adaptation of making successful prediction (or any other correlate of successful learning) rewarding in itself. The system is also ideally suited for the task of self-directed learning because dopamine release in itself serves to "lock" attention on the item, task, or concept currently being attended to or represented. Because both the causes and effects of dopamine release are so well suited to usefully directing learning toward optimal areas, it is likely that the dopamine system plays a crucial role in this important human adaptation.

This hypothesis is compatible with the approach of Huang and Weng [17]. The dopamine system is not only well known to signal reward, but seems to also signal punishment and novelty, the three components they suggest are the minimum for an effective self-directed learning system.

Other aspects of human self directed learning deserve attention as well. For instance, there are likely biases in the self-direction of learning that prevent people from searching behavior-space at random. Infants may be biased toward making motor actions and sounds ("babbling") that may pre-train systems for the more deliberate tasks discussed here. Imitation likely provides an important constraint guiding learners to useful parts of behavior-space.

Further understanding the function of dopamine and other neuromodulatory systems (e.g., norepinephrin [8]) will help us understand how the human brain usefully directs its own learning. This will in turn allow us to design more human-like and more effective learners. Humans' poorly understood ability to select our own tasks and learning examples may well be a crucial ingredient in our still-unique ability to arrive at a rich understanding of our world.

References

[1] P. Oudeyer, F. Kaplan, V. Hafner, Intrinsic motivation systems for autonomous mental development, Evolutionary Computation, IEEE Transactions on evolutionary computation 11 (2007) 265–286.

[2] A. R. Aron, D. Shohamy, J. Clark, C. Myers, M. A. Gluck, R. A. Poldrack, Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning., Journal of neurophysiology 92 (2) (2004) 1144–1152.
URL http://www.ncbi.nlm.nih.gov/pubmed/15014103

[3] M. Tomasello, The Cultural Origins of Human Cognition, Harvard University Press, Cambridge, MA, 2001.

[4] R. S. Sutton, Learning to predict by the method of temporal differences, Machine Learning 3 (1988) 9–44.

[5] R. C. O'Reilly, M. J. Frank, T. E. Hazy, B. Watz, Pvlv: The primary value and learned value pavlovian learning algorithm., Behavioral Neuroscience 121 (2007) 31–49.
URL http://www.ncbi.nlm.nih.gov/pubmed/17324049

[6] T. E. Hazy, M. J. Frank, R. C. O'Reilly, Neural mechanisms of acquired phasic dopamine responses in learning., Neuroscience and biobehavioral reviews 34 (5) (2010) 701–720.
URL http://www.ncbi.nlm.nih.gov/pubmed/19944716

[7] R. Montague, *Why Choose This Book?*, Dutton, New York, New York, 2006.

[8] G. Aston-Jones, J. D. Cohen, An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance., Annual review of neuroscience 28 (2005) 403–450.
URL http://www.ncbi.nlm.nih.gov/pubmed/16022602

[9] S. Bao, V. T. Chan, M. M. Merzenich, Cortical remodelling induced by activity of ventral tegmental dopamine neurons., Nature 412 (2001) 79–82.
URL http://www.ncbi.nlm.nih.gov/pubmed/11452310

[10] J. K. Seamans, C. R. Yang, The principal features and mechanisms of dopamine modulation in the prefrontal cortex., Progress in neurobiology 74 (2004) 1–57.
URL http://www.ncbi.nlm.nih.gov/pubmed/15381316

[11] J. B. Morton, Y. Munakata, Active versus latent representations: A neural network model of perseveration and dissociation in early childhood, Developmental Psychobiology 40 (2002) 255–265.

[12] G. Deco, E. T. Rolls, Attention, short-term memory, and action selection: a unifying theory., Progress in neurobiology 76 (4).
URL http://www.ncbi.nlm.nih.gov/pubmed/16257103

[13] R. Desimone, J. Duncan, Neural mechanisms of selective visual attention., Annual Review of Neuroscience 18 (1995) 193.

[14] S. A. Herd, M. T. Banich, R. C. O'Reilly, Neural mechanisms of cognitive control: an integrative model of stroop task performance and fmri data., Journal of cognitive neuroscience 18 (2006) 22–32.
URL http://www.ncbi.nlm.nih.gov/pubmed/16417680

[15] J. K. Seamans, T. W. Robbins, Dopamine modulation of the prefrontal cortex and cognitive function, 2010, pp. 373–398.

[16] T. E. Hazy, M. J. Frank, R. C. O'Reilly, Banishing the homunculus: Making working memory work., Neuroscience 139 (2006) 105–118.
URL http://www.ncbi.nlm.nih.gov/pubmed/16343792

[17] X. Huang, J. Weng, Inherent Value Systems for Autonomous Mental Development, International Journal of Humanoid Robotics 4 (2007) 407–433.