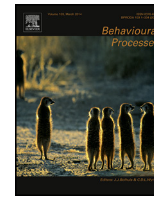
Behavioural Processes 120 (2015) 54–63

Contents lists available at ScienceDirect

Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc



Unified-theory-of-reinforcement neural networks do not simulate the blocking effect

Nicholas T. Calvin [*,1], J. J McDowell

Emory University, Atlanta, Georgia

ARTICLE INFO

Article history:
Received 2 April 2015
Received in revised form 18 August 2015
Accepted 21 August 2015
Available online 28 August 2015

Keywords:
Agent-based modeling
Complexity theory
Neural networks
Blocking effect

ABSTRACT

For the last 20 years the unified theory of reinforcement (Donahoe et al., 1993) has been used to develop computer simulations to evaluate its plausibility as an account for behavior. The unified theory of reinforcement states that operant and respondent learning occurs via the same neural mechanisms. As part of a larger project to evaluate the operant behavior predicted by the theory, this project was the first replication of neural network models based on the unified theory of reinforcement. In the process of replicating these neural network models it became apparent that a previously published finding, namely, that the networks simulate the blocking phenomenon (Donahoe et al., 1993), was a misinterpretation of the data. We show that the apparent blocking produced by these networks is an artifact of the inability of these networks to generate the same conditioned response to multiple stimuli. The piecemeal approach to evaluate the unified theory of reinforcement via simulation is critiqued and alternatives are discussed.

1. Introduction

The unified theory of reinforcement (UTR) has been used to develop neural network models to evaluate its neural-mechanistic account of behavior (Burgos, 1996, 1997, 2003, 2005, 2007; Burgos and Murillo-Rodríguez, 2007; Burgos et al., 2008; Burns et al., 2011; Donahoe, 2002; Donahoe and Burgos, 1999, 2000; Donahoe et al., 1993, 1997a,b; Sánchez et al., 2010). At its simplest, the unified theory of reinforcement proposes that the same neurological systems control behavior under both operant and respondent contingencies (Donahoe et al., 1993). The ventral tegmental and CA1 areas of the brain are central to this theory because they control changes in neural connectivity via diffuse reinforcement systems. The internal changes caused by these diffuse reinforcement systems result in an animal's behavior adapting to its environment. The theory's account is selectionistic, as behaviors that precede positive consequences become more likely to occur in the future. While this theory is a biological argument, its plausibility has been evaluated by simulating neural networks that interact with environments.

UTR neural networks have successfully mimicked a number of important phenomena. Early in their development, these networks were shown to be capable of exhibiting conditioned responding, extinction, and facilitated reacquisition under both respondent and operant contingencies (Donahoe et al., 1993). The model has also been argued to demonstrate behavior that is consistent with the blocking phenomenon (Donahoe et al., 1993). Other phenomena that UTR neural network models have mimicked are revaluation (Donahoe and Burgos, 2000), latent inhibition (Burgos, 2003), and autoshaping (Burgos, 2007). The behavior of UTR neural networks have also been compared to those of animals to suggest that this type of neural network generates qualitatively accurate predictions (Burgos et al., 2008; Burns et al., 2011). Unfortunately, despite these successes, all published work with these neural networks has been led by the original group of researchers, which has limited the degree of discussion regarding the model's design and behavior.

The ability of UTR neural networks to simulate both blocking and facilitated reacquisition phenomena is very important. The blocking phenomenon (Kamin, 1969) is when a novel stimulus fails to become effective during conditioning due to a strong, previously conditioned stimulus/stimuli being presented with the new stimulus. This phenomenon is particularly important in the history of conditioning because of the Rescorla-Wagner model (Rescorla and Wagner, 1972; Wagner and Rescorla, 1972), which predicts it. This model is still referenced and discussed to the current day (*e.g.*, Ruprecht et al., 2014; Culver et al., 2015) because many of its most important predications are accurate (Miller et al., 1995). There are, however, phenomena that it fails to predict, and facilitated reacquisition is one of them (Miller et al., 1995). By being able

* Corresponding author.
E-mail addresses: ncalvin@emory.edu, elitegreenpaladin@gmail.com (N.T. Calvin).
[1] Nicholas T. Calvin and Jack J. McDowell, Department of Psychology, Emory University.

http://dx.doi.org/10.1016/j.beproc.2015.08.008
0376-6357/

to predict both phenomena UTR neural networks were established as a potentially important model of conditioning.

Unified-theory-of-reinforcement neural network models are relatively complex compared to other models because the networks attempt to mimic neural systems. Despite their complexity, these models are built from only two component types: neural processing units (NPUs) and connections. The primary function of NPUs is to calculate an activation level, which is a rough determination of how the network should behave. The current activation level of each NPU is a function of its previous level and the activation levels of the NPUs that connect to it. There are also specific circumstances where a NPU's activation level is set directly by the simulation rather than being calculated. Connections, the other component type, link pairs of NPUs and are unidirectional. Each connection has a weight value which modulates the extent to which the preceding NPU's activation level influences the terminal NPU's activation level. Connections enable networks to perform complex calculations by linking multiple NPUs together, thus combining their computational power. The exact equations used to manipulate these processes are provided in Appendix A but are not necessary to understand how UTR neural networks generally function. In fact, the mathematical details of these neural networks are not considered a central component to how the UTR is translated into simulations (Donahoe, 2002).

The arrangement of connections and NPUs, a network's architecture, is the most important part of the theory's implementation (Donahoe et al., 1993). The architecture of UTR neural networks is typically organized into four distinct layers, as shown in Fig. 1. From left to right these layers are the input (IN), hippocampal (HIP), dopaminergic (DOP), and output (OUT). The layers are listed in the order of processing, with the input layer being affected by stimuli before the hippocampal layer, the hippocampal before the dopaminergic layer, and so on. The architecture of UTR neural networks can be further subdivided into two parallel pathways that run through all four layers. These are the response and learning pathways, which are differentiated from each other in Fig. 1 by shading; the learning pathway is shaded gray. The response pathway determines the behaviors that are evoked or elicited by stimuli. The learning pathway adapts the response pathway to the environment such that the network will more frequently express behaviors that have previously resulted in beneficial consequences. This unique interaction of the learning and the response pathways through the processing layers is the primary way in which the theory is manifested within UTR neural networks.

The function of the response pathway is to observe and engage with the environment. Stimuli are observed at the input layer

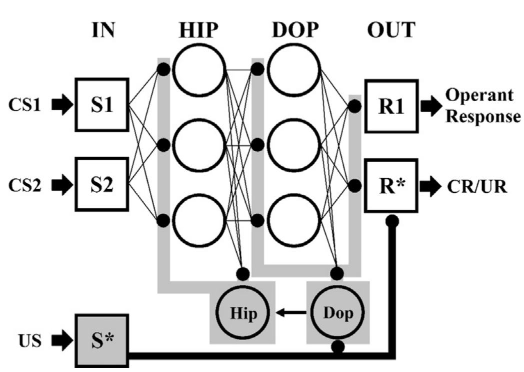


Fig. 1. The architecture of a standard 2_1-3_1-3_1-2 unified-theory-of-reinforcement neural network. Standard UTR neural networks have 4 layers, which are input (IN), hippocampal (HIP), dopaminergic (DOP), and output (OUT). The letters within the NPUs indicate special functions (S—stimulus detecting, R—response emitting, Hip—hippocampal, Dop—dopaminergic, *—unconditioned). The design of this figure is based on Figs. 1 and 4 from the Sánchez et al. (2010) article.

by unique stimuli-detecting NPUs. These NPUs are unique in that their activation level is set to a specific value if in the presence of a specific stimulus. For example, if a conditioned stimulus is observed then the activation level of the appropriate conditioned stimulus detecting NPU, for example S1 in Fig. 1, would be set to a certain value, with larger values indicating greater stimulus salience. This environmental information is passed from the input layer to the hippocampal layer via connections, which are shown in Fig. 1 as thin black lines that end in black circles. The model's hippocampal interneuron layer has been argued to function as a sensory-association area because it can combine disparate stimuli, such as legs, a seat, and a back, into a more complex stimulus, like a chair (Burgos, 2003; Burgos et al., 2008; Donahoe, 2002; Donahoe et al., 1997a,b). This complex-stimulus information is passed to and processed by the dopaminergic interneuron layer. The dopaminergic layer has been argued to behave as a motor-association area because it determines how the network will engage with the environment (Burgos, 2003; Burgos et al., 2008; Donahoe, 2002; Donahoe et al., 1997a,b). This response planning information is then passed to the response layer where the actual behavior is generated. The activation levels of the NPUs in the response layer determine which behaviors the network will engage in. Through these steps the neural network model observes and interacts with the environment.

The learning pathway adapts the neural network's response pathway to the environment and controls unconditioned responses. These changes occur due to unconditioned stimuli which are detected by the NPUs of the learning pathway's input layer. If an unconditioned stimulus is available, then the activation level of the unconditioned stimulus NPU, which is S* in Fig. 1, is set to a specific value (Donahoe, 1993). Larger values indicate greater reinforcer magnitudes. The activation level of the unconditioned stimulus NPU is then directly passed to the unconditioned/conditioned response NPU and dopaminergic NPU, which are respectively labeled R* and Dop in Fig. 1, through unique connections. The connections from the unconditioned stimulus NPU are unique in that they set the activation level of the target NPUs equal to the unconditioned stimulus NPU's activation level. Despite functioning identically, these two connections serve very different purposes. The connection between the unconditioned stimulus NPU and unconditioned/conditioned response NPU creates a reflex; the unconditioned stimulus always results in the unconditioned response. The unique connection to the dopaminergic NPU is the most important connection in the entire neural network, because it provides the reinforcement signals that enable the network to adapt to the environment. These two unique connections give the network its ability to adapt and react.

The reinforcement signal in the learning pathway is used to adapt the neural network to the environment by changing the weights of connection. Connection weights are changed based on the difference between the network's prediction and the actual consequence, with greater differences leading to greater changes. Connection weights are updated in two different ways. The updating method that is used depends on the layer at which the connection ends. Connections that end at the dopaminergic or response layer are adapted to the environment via a diffuse reinforcement signal that is produced by the dopaminergic NPU. The magnitude of the reinforcement signal is determined by the dopaminergic NPU's activation level, which changes most dramatically when it receives a reinforcement signal from the unconditioned stimulus NPU. The presence of an unconditioned stimulus typically results in connection weights increasing and its absence in connection weights decreasing. The gray regions in Fig. 1 that lead from the NPU labeled Dop indicate which connections are affected by this reinforcement signal. The dopaminergic NPU has been argued to correspond to the ventral tegmental area because the current neu-