Extending unified-theory-of-reinforcement neural networks to steady-state operant behavior

Olivia L. Calvin [*,1], J.J. McDowell

Department of Psychology, Emory University, Atlanta, Georgia

ABSTRACT

The unified theory of reinforcement has been used to develop models of behavior over the last 20 years (Donahoe et al., 1993). Previous research has focused on the theory's concordance with the respondent behavior of humans and animals. In this experiment, neural networks were developed from the theory to extend the unified theory of reinforcement to operant behavior on single-alternative variable-interval schedules. This area of operant research was selected because previously developed neural networks could be applied to it without significant alteration. Previous research with humans and animals indicates that the pattern of their steady-state behavior is hyperbolic when plotted against the obtained rate of reinforcement (Herrnstein, 1970). A genetic algorithm was used in the first part of the experiment to determine parameter values for the neural networks, because values that were used in previous research did not result in a hyperbolic pattern of behavior. After finding these parameters, hyperbolic and other similar functions were fitted to the behavior produced by the neural networks. The form of the neural network's behavior was best described by an exponentiated hyperbola (McDowell, 1986; McLean and White, 1983; Wearden, 1981), which was derived from the generalized matching law (Baum, 1974). In post-hoc analyses the addition of a baseline rate of behavior significantly improved the fit of the exponentiated hyperbola and removed systematic residuals. The form of this function was consistent with human and animal behavior, but the estimated parameter values were not.

© 2016 Elsevier B.V. All rights reserved.

Contents

1. Introduction

The central assertion of the unified theory of reinforcement (UTR) is that behavior in operant and respondent experiments is a result of the same neural process (Donahoe et al., 1993). This is a general theory that describes the internal biological processes that lead to both operant and respondent behavior. To evaluate

* Corresponding author at: Department of Psychology, Emory University, Atlanta, GA 30322, Georgia.
   E-mail address: ncalvin@emory.edu (O.L. Calvin).
[1] Went by the name of Nicholas T. Calvin in previously published work.

the plausibility of the UTR as an account for behavior, biologically inspired models have been developed and evaluated (Burgos, 1996, 1997, 2003, 2005, 2007; Burgos and Murillo-Rodríguez, 2007; Burgos et al., 2008; Burns et al., 2011; Calvin and McDowell, 2015; Donahoe, 2002; Donahoe and Burgos, 1999, 2000; Donahoe et al., 1993, 1997a,b; Sánchez et al., 2010). UTR-inspired models do not explicitly distinguish between respondent and operant contingencies, but in their functioning they adapt to both contingencies. The theory states that behavior adapts by adjusting the strength of neural connections in response to positive consequences. In the absence of positive consequences, neural connections slowly weaken and behavior is less likely to be observed. Through repeated interactions with the environment, UTR-inspired models adapt their behavior to environmental events.

Research with UTR-based models has focused on respondent behavior, and very little has been done to examine operant behavior. The only exceptions to this were demonstrations of operant conditioning by UTR-based models (Donahoe et al., 1993; Calvin and McDowell, 2015). That a behavior is more frequently observed when the behavior is followed by positive consequences was a minimum prerequisite for UTR-based models to be plausible, and it was important that this was demonstrated. Since the UTR must account for both operant and respondent behavior, more operant research would enhance its plausibility as an account for behavior.

An area of operant research to which previously explored UTR-based models can be applied without significant alteration is the quantitative law of effect (Herrnstein, 1970). The quantitative law of effect is a development of the matching law (Herrnstein, 1961) that added important theoretical underpinnings in order to understand behavior on single-alternative variable-interval (VI) schedules. The quantitative law of effect has been shown to describe single-alternative VI behavior of animals (e.g., Herrnstein, 1970; McSweeney et al., 1983; reviewed in McDowell, 2013) and humans (Beardsley and McDowell, 1992; Bradshaw et al., 1976, 1977, 1978; Fernandez et al., 1995; McDowell and Wood, 1984, 1985). Herrnstein developed the quantitative law of effect by making the two important assumptions that humans and animals engage in constant rates of behavior, and that all behavior is choice. These assumptions extended the matching law, which describes behavior on concurrent VIVI schedules, to single VI schedules (Herrnstein, 1970). The quantitative law of effect is a hyperbola,

$$B = \frac{kR}{R + r_e},\tag{1}$$

where B is the observed rate of a target behavior, R is the rate of obtained reinforcement, and k and r_e are estimated parameters that have important theoretical interpretations. In the theory, the k parameter is the maximum rate of behavior, which is assumed to be constant. If the organism is not engaging in a targeted operant behavior then it is assumed to be engaging in other behaviors that may result in beneficial outcomes. The unmeasured extraneous behavior is assumed to occasionally result in reinforcement, which is the r_e parameter. Mathematically, the hyperbola asymptotes at k, and r_e is the point on the x-axis that predicts a rate of behavior that is half that of k (Bradshaw et al., 1976).

An improved version of the matching law was developed by Baum (1974) to account for systematic inaccuracies in the original version, and can be used to develop a single alternative version in the same way that Eq. (1) was derived (McDowell, 1986; McLean and White, 1983; Wearden, 1981). The simplified version of this generalized quantitative law of effect is an exponentiated hyperbola,

$$B = \frac{kR^a}{R^a + r_e}.\tag{2}$$

The parameter a allows for systematic deviations from the exact matching of ratios of behavior and reinforcement, which are termed under- and over-matching. The r_e in Eq. (2) has a slightly different meaning than in Eq. (1) because its full theoretical expression is $r_e{}^a/b$, which could be interpreted as the relative value of the reinforcers obtained by unmeasured behavior. The interpretation changes because the bias parameter, b, accounts for systematic differences in the reinforcing values of measured and unmeasured reinforcers. For the purposes of fitting the equation it is simplified to a single parameter, because r_e and b cannot be independently estimated. The exponent parameter, a, adjusts the form of the function by bending it at the lowest rates of obtained reinforcement. If the value of a is greater than 1 then the function tends to flatten at the lowest rates of obtained reinforcement, and when less than 1 the function becomes steeper. While this exponentiated hyperbola is similar to the hyperbola specified by Eq. (1), it has some unique fitting characteristics, and is based on the more strongly supported generalized matching law (McDowell, 2013).

To assess the UTR's predictions it is necessary to simulate UTR-inspired neural networks. The complex and flexible behavior of these neural networks comes from the interactions of relatively simple components. At their simplest, these networks are built from two types of components: neural processing units (NPUs) and connections. These components serve different functions within the networks, with NPUs primarily determining how the network will behave in the immediate future, and connections transmitting and regulating the importance of NPU determinations. Connections are very important, because they are the components of neural networks that adapt to the environment. Connections adapt by changing their strength, which regulates how important the NPU at the beginning of the connection is to the NPU at its end point. If a positive consequence follows behavior, then the connections that previously led to that behavior are strengthened, which makes that behavior more likely to occur in the future. Detailed mathematical descriptions of the network components may be found in multiple articles (Burgos, 2003, 2007; Burgos et al., 2008; Calvin and McDowell, 2015; Donahoe et al., 1993; Sánchez et al., 2010), but their exact mathematical functioning is not a critical component of the UTR and has been omitted in this paper for the sake of concision. Of the articles that provide mathematical descriptions, Calvin and McDowell (2015) provides a particularly clear description of the networks and also includes a copy of the code that was used to conduct those experiments, which is helpful to whoever is interested in replicating UTR neural networks.

While the exact functioning of connections and NPUs are not theoretically important to the UTR, their arrangement and roles within the network are especially important to the theory (Donahoe et al., 1993). The standard UTR neural network is 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) layers. Information about environmental stimuli and the consequences of behavior are given to the neural network in the input layer, thus acting as its eyes and ears. The hippocampal and dopaminergic layers then process this information to determine how the network should behave in the environment. The output layer implements this decision by interacting with the environment. By processing information through these layers the neural network engages with and adapts to its environment.

The architecture of UTR neural networks can be subdivided into response and learning pathways that cross all four layers. The response pathway determines which, if any, behaviors are evoked or elicited by the environment, and the learning pathway adapts the response pathway to the environment by changing connection strengths. In Fig. 1, the learning pathway is shaded gray to differentiate it from the response pathway. The network is selectionist because behaviors become more likely to occur when followed