

# eScholarship

## International Journal of Comparative Psychology

### Title

Misbehavior in a Neural Network Model

### Permalink

<https://escholarship.org/uc/item/3vb500tv>

### Journal

International Journal of Comparative Psychology, 28(1)

### ISSN

0889-3675

### Author

Burgos, José E

### Publication Date

2015

### DOI

10.46867/ijcp.2015.28.01.02

### Copyright Information

Copyright 2015 by the author(s). This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



## Misbehavior in a Neural Network Model

**José E. Burgos**

*CEIC, University of Guadalajara, Mexico*

This paper describes a neural network account of misbehavior with an extant neural network model of conditioning. The model makes no distinction between learning (weight-change mechanisms) in operant and Pavlovian conditioning, but preserves the standard behavioral distinctions between types of stimuli, responses, and contingencies, with connectionist interpretations of some possible neuroanatomical substrates. Misbehavior has been traditionally conceived as a species-specific response  $R^*$  that is unnecessary for a biologically significant reward  $S^*$  but interferes with another response  $R$  that is necessary for  $S^*$ . Misbehavior thus conceived has been explained as interfering Pavlovian conditioned responding. Three four-layer feedforward neural networks were designed to differ only in their output layers, as a connectionist interpretation of three hypothetical operant-Pavlovian relations in misbehavior, namely, interference (Pavlovian output to operant output lateral inhibitory connection), compatibility (Pavlovian output to operant output lateral excitatory connection), and independence (no lateral connection in the output layer). These relations are proposed as neural-network interpretations of neuroanatomical substrates of conditioning with three biologically significant stimuli, namely, food, water, and sexual mate, respectively. Each network first received pairings of contextual cues with its respective  $S^*$ , to simulate pretraining with such stimuli. Then, networks received operant contingencies where  $S^*$  was paired with the same contextual cues, as well as cues from a token dependently on  $R$  responding, defined as a minimal  $R$  activation of 0.5. Networks showed substantial misbehavior (qua conditioned  $R^*$  responding) that interfered with  $R$  to different extents, food causing the most, sexual mate the least interference. Limitations, future directions, and implications for biological constraints and the generality of learning are discussed.

This study presents a computer simulation of the learning phenomenon known as *misbehavior* using an extant neural network model. The study is intensely theoretical and conjectural. The model is described in the first section (its mathematical part is described in the appendix). The simulation is described in the second section. Implications and limitations are discussed in the last section. For this introduction, I summarize previous research on misbehavior and specify the elements that will be the focus of the study.

The term *misbehavior* was first applied to learning by Breland and Breland (1961). They trained animals of different species (e.g., chickens, raccoons, pigs) in non-experimental (but still unnatural) situations to respond in some relatively arbitrary way  $R$  (e.g., picking up some discrete object like a wooden coin and depositing it in a container) for food as a primary reinforcer ( $S^*$ ) in operant contingencies. The animals learned  $R$ . However, they also performed certain other responses  $R^*$  (e.g., rooting,

rubbing a token) that were not required for  $S^*$  but interfered with  $R$  and thus significantly delayed  $S^*$ ; hence the term *misbehavior* to refer to  $R^*$ .

Such misbehaving was consistently observed across species, for which the authors concluded that it showed “breakdowns of operant conditioned behavior” (p. 681).  $R^*$ , in fact, was reminiscent of species-specific behaviors towards food. For this reason, the authors also interpreted  $R^*$  biologically as *instinctive drift*, to mean that “learned behavior drifts toward instinctive behavior” (p. 684).

Misbehavior was thus viewed as a case of biological constraints on learning (e.g., Bitterman, 1975; Domjan & Galef, 1983; Seligman, 1970), the notion that the laws of learning depend on the specific stimuli, responses, and reinforcers used, and vary systematically across species. In the case of misbehavior, this idea challenged the generality of the law of operant (instrumental) conditioning, or “Law of Conditioning of Type R,” as formulated by Skinner (1938): “If the occurrence of an operant is followed by a presentation of a reinforcing stimulus, the strength is increased” (p. 21). Thus formulated, this law does not predict misbehavior (and other phenomena that have been taken as indicative of biological constraints on learning, such as cue-to-consequence effects and autoshaping). The law’s generality, then, is limited. This limitation has been argued to be due to the exclusion of biological factors pertaining to the adaptive value and evolutionary history of food-related behavior.

Indeed, but this proposal, as valid as it is, does not explain *how* learned behavior “drifts toward instinctive behavior.” The Brelands viewed misbehavior as *instinctive*, meaning that it is innate, not learned or acquired. And yet, it is directed towards objects other than, and in the prolonged absence of, food, its supposedly biologically specific stimulus. Misbehavior thus must involve some learning. But what kind of learning does it involve? And how does it interfere with operant responding? Neither the notion of biological constraints on learning nor the Brelands’ interpretation, answer these questions.

The reason is that these are questions about *mechanisms* as *proximate* causes of behavior in current species. Biological constraints and the Brelands’ interpretation refer to evolutionary adaptation and history, which are *distal* causes. Although mechanisms can profitably, and perhaps correctly, be hypothesized as evolutionary adaptations, they still need to be characterized. Evolutionary considerations are necessary, but insufficient for such characterization. My focus in this paper is on mechanisms as proximate causes, in abstraction of evolutionary distal causes, which I take as given, unanalyzed initial conditions, for the sake of theoretical abstraction.

Possible answers to those questions have been proposed in experimental studies of misbehavior with rats. Boakes, Poli, Lockwood, and Goodall (1978) trained rats to deliver ball-bearings (made of steel or nylon) into a hole ( $R$ ) for food or water ( $S^*$ ). The results confirmed those reported by Breland and Breland (1961) with other species: The rats tended to perform other activities ( $R^*$ ) that were unnecessary for but still biologically related to  $S^*$  (e.g., chewing), and tended to interfere with  $R$ , lengthening ball-delivery times and hence causing reinforcement delays.

To explain their results, the authors hypothesized misbehavior to be  $S^*$ -related  $R^*$  responding that is conditioned through Pavlovian ( $S$ - $S^*$ , stimulus-reinforcer)

contingencies that accompany operant ( $R-S^*$ , response-reinforcer) contingencies, where  $S$  included sensory stimulation from the balls. The authors also hypothesized such responding to interfere with the response  $R$  that is required in  $R-S^*$  contingencies. The authors took this explanation to be consistent with a stimulus-substitution interpretation of Pavlovian (classical, respondent) conditioning, but warned that such interpretation did not fit other observations (e.g., longer ball-delivery times with nylon balls).

The authors also reported that using water as a reward resulted in moderately shorter ball-delivery times (and, hence, reinforcement delays) than food. This observation is consistent with the Pavlovian-interference hypothesis, but makes it a matter of degree, being greater with food than water. Still, the authors claimed that a stimulus-substitution interpretation predicts a more pronounced difference than was observed.

In a different experimental series, Timberlake, Wahl, and King (1982, Experiments 4 and 5) also trained rats to make contact with a ball bearing for food. The rats also behaved in ways that resembled the species' typical response patterns towards food, such as gnawing, pawing, carrying, and nosing. These activities also tended to interfere with operant responding, causing reinforcement delays.

The authors also observed that explicitly pairing the ball with food in Pavlovian contingencies (Experiment 1) resulted in comparable food-related activities towards the ball, without the need of explicit operant contingencies. This result supports Boakes et al.'s (1978) hypothesis that sensory stimulation from the ball could function as a conditioned stimulus (CS) for misbehavior.

Timberlake et al. (1982) thus agreed with Boakes et al. (1978) in hypothesizing that misbehavior was conditioned Pavlovian responding, but also warned about the pitfalls of the stimulus-substitution hypothesis. Timberlake et al. (1982) also agreed in hypothesizing that misbehavior, qua food-related conditioned Pavlovian responding, interfered with operant responding in operant contingencies. But they were somewhat more precise about the kind of mechanism that could underlie such interference, namely, *inhibition*.

They also argued that misbehavior needed not be interfering, hypothesizing two other possible relations, in addition to interference: Compatibility and independence. In compatibility, the authors hypothesized that misbehavior should facilitate operant conditioning. In independence, they hypothesized that misbehavior should not affect operant conditioning. They predicted that both relations should result in less interference of operant responding by misbehavior. They also implied that misbehavior that is compatible with  $R$  should be less interfering than misbehavior that is independent of  $R$ .

These relations can help explain Boakes et al.'s (1978) observation that using water as  $S^*$  in operant contingencies results in less interfering misbehavior: Water-related responding in rats could be more compatible with  $R$  than food-related responding. Or perhaps water-related responding is independent of  $R$ . Either one would explain the observation. Unfortunately, there is no evidence to decide which one is the right explanation. A new experiment would be needed to make this determination, but I will not do this here, at least not with animals.

These studies contain all the elements, evidential and interpretive, that will be the focus of the present study. Roughly, misbehavior is a form of responding  $R^*$  (e.g., chewing) that resembles unconditioned responding to a biologically significant stimulus  $S^*$  (e.g., food, water) and is conditioned through Pavlovian,  $S$ - $S^*$  contingencies.  $R^*$  is directed towards a biologically neutral token (e.g., a wooden coin, a ball-bearing), contact with which is required as a response  $R$  to obtain  $S^*$  in operant ( $R$ - $S^*$ ) contingencies. Misbehavior thus conceived interferes with  $R$  to different degrees, depending on the type of  $S^*$  and how compatible  $R^*$  is with  $R$ . Next, I interpret these elements in terms of the model.

## The Model

The model was first proposed by Donahoe, Burgos, and Palmer (1993) as a unified connectionist interpretation of Pavlovian and operant conditioning. Pavlovian conditioning phenomena previously simulated with the model include acquisition, extinction, and reacquisition “savings” (Donahoe, Burgos, & Palmer, 1993); ISI functions (Burgos, 1997); latent inhibition (Burgos, 2003); C/T ratio effects (Burgos, 2005); context-shift effects (Burgos & Murillo-Rodriguez, 2007); simultaneous conditioning (Burgos, Flores, García, Díaz, & Cruz, 2008); second-order conditioning and resistance to extinction (Sánchez, Galeazzi, & Burgos, 2010); blocking and overshadowing (Burns, Burgos, & Donahoe, 2011).

Simulations of operant conditioning have been less extensive. They include acquisition, extinction, faster reacquisition, generalization, and discrimination (Donahoe, Burgos, & Palmer, 1993); timing in fixed-interval schedules (Burgos & Donahoe, 2000; Donahoe & Burgos, 1999); and reinforcement revaluation (Donahoe & Burgos, 2000). It remains to be seen whether the model can also simulate other operant conditioning phenomena (e.g., behavioral contrast, negative reinforcement, positive punishment, fixed-ratio performance, matching law, matching to sample, etc.).

At the intersection of operant with Pavlovian conditioning, the model also simulates autoshaping, automaintenance (Burgos, 2007), and autoshaped choice (Burgos & García-Leal, 2015). These phenomena exemplify acquisition and maintenance of emitted responding through Pavlovian contingencies, without explicit operant contingencies. Hence, explicit operant contingencies are not necessary for such acquisition and maintenance. The model’s ability to simulate this will be central to the present study. But first, I layout the model’s basic assumptions.

## Connectionist Assumption

As other neural network models, the present one adopts the general connectionist assumption that cognition (broadly conceived to include learning and behavior in nonhuman animals) is better explained in terms of the functioning of artificial neural networks, parallel distributed processing systems that consist of interconnected neuron-like units, if one is to take into account brain structure and function. The model extends this assumption to animal conditioning, Pavlovian and

operant. The assumption has the following methodological aspects that are familiar to modelers, but may not be apparent to non-modelers.

Neural network models are strongly mathematical, for the sake of clarity and precision. Explanations in terms of these models take the form of *computer simulations* that help determine whether a model has numerical solutions that are consistent with evidence about some phenomenon of interest. Computers are used for expediency, to achieve this task reliably and efficiently. This reason, however, does not imply that the phenomena being simulated are computational in nature (cf. Gallistel, 1990).

Very rarely, if ever, simulations perfectly emulate phenomena, especially complex ones, like animal learning and behavior. Rather, simulations are very rough, interpretive approximations often assessed through *ordinal comparisons*, my strategy here (quantitative fits of the data are also pursued, but I will not do it here). The reason is that models used in simulations are highly abstract and simplified in that they focus on a few aspects, purposely excluding the rest. This simplification allows using models effectively for explanation and prediction. More realistic models are more complex and hence more difficult to use.

A model, then, can hardly be rejected for excluding, as exclusion is essential to modeling. To reject a model for this reason, then, is to reject *all* models, which is too drastic, given the centrality of models in science (see Bailer-Jones, 2009). The present model is no exception to this exclusionary character of models. As will become apparent next, in the model's neuroscientific and psychological assumptions, it excludes most details to focus on a few ones, to see how they work on their own, in isolation of others. There is no universally agreed-upon criterion to determine how much exclusion is too much. Therefore, no model can be meaningfully said to exclude too much, or too little, for that matter. Whether a model excludes too much or too little, then, very much like beauty, is in the eye of the beholder. Still, the model is somewhat more strongly inspired by neuroscience than others.

## **Neuroscientific Assumptions**

The model assumes Pavlovian and operant conditioning to depend on neuroanatomical features. Only a few neuroanatomical features have been chosen, for the sake of theoretical abstraction. Following Grossberg's (e.g., 1974) method of "minimal anatomies," networks in this model are intended as minimal theoretical structures that realize the assumptions and simulate the phenomena of interest<sup>1</sup>. Networks are to be made more realistic only as dictated by the assumptions and phenomena.

Networks in this model are not intended to simulate any specific local brain circuit either, but rather global circuits that involve several brain areas and nuclei. More precisely, networks are designed after a few basic, general, well-established gross-neuroanatomical principles of organization of vertebrate brains. According to these principles (see Mesulam, 1998), all vertebrate brains have a basic anatomical structure

---

<sup>1</sup>Construction of the present model, however, did not follow another feature of this method: Inference of neuroscientific from psychological assumptions. Rather, neuroscientific assumptions were derived directly from neuroscience, and the psychological assumptions were derived from the neuroscientific assumptions.

where primary-sensory areas ( $S'$ ; e.g., primary-visual, primary-auditory) project to polysensory (sensory-association, or  $S''$ ) areas. These, in turn, project to the hippocampal system (which projects back to  $S''$ ) and motor-association (motor-association, or  $M''$ ) areas.  $M''$  areas project to dopaminergic systems ( $D$ , which project back to motor-association areas) and primary-motor ( $M'$ ) areas.

This basic  $S'-S''-H-S''-M''-D-M''-M'$  organization is common to a wide range of stimuli, reinforcers, responses, and species. It thus provides a neuroanatomical sense in which learning and behavior can be said to be fairly general. However, the model also allows for considerable variation in this basic organization across particular networks. Thus, indefinitely many particular neural networks are possible that differ in the number of units, layers, units per layer, and how they are connected. That is to say, the model allows network architecture to be an *independent variable* in simulation experiments (e.g., Burgos & Donahoe, 2000).

At the synaptic level, the model assumes that Pavlovian and operant conditioning also depend on synaptic plasticity modulated by dopaminergic ( $D$ ) and hippocampal ( $H$ ) systems. The role of dopaminergic systems in both types of conditioning has been well documented (see Schultz, 2010). The evidence indicates that such role is to provide a *discrepancy* (temporal-difference) *diffuse* signal that modulates changes in synaptic efficacies in motor areas (associative as well as primary). The model also hypothesizes that this role is also played in the sensory-motor interface.

Hippocampal systems too, influenced by dopaminergic systems, have been shown to play a role in both types of conditioning (see Everitt & Robbins, 2005). The role of hippocampal systems has not been as precisely asserted as that of dopaminergic systems, but the model generalizes from the latter role: Hippocampal systems too provide a discrepancy diffuse signal that modulates synaptic plasticity, but in *sensory association* areas.

Both modulations work in the model as follows: The larger the discrepancy, the more the weight gained by a connection, everything else being equal. This process is intended to simulate increases in sensory and motor synaptic efficacies by hippocampal and dopaminergic activity. The model assumes that these increases underlie both types of conditioning, which is not to say that the two types of conditioning do not differ in anything, whether anatomically, behaviorally (the model makes both distinctions), or genetically (e.g., Brems & Plendl, 2008). Nor does it mean that they have no other commonalities.

## **Psychological Assumptions**

Psychologically, the model was proposed as a connectionist interpretation of the unified behavioral principle of reinforcement for Pavlovian and operant conditioning, proposed by Donahoe, Crowley, Millard, and Stickney (1982). This principle states that all cues that are contiguous with a biologically significant stimulus  $S^*$  come to control all responses occurring in that moment. The model adds assumptions about some possible neural mechanisms and structures underlying this principle (e.g., dopaminergic and hippocampal functioning).

There is considerable overlap between the psychological and neuroscientific assumptions. To begin with, the model defines learning neurobiologically, in terms of

synaptic plasticity, as changes in the efficacies of certain synapses. Synaptic efficacies are modeled as connection weights. These changes are thus modeled as changes in connection weights, according to the learning rule (see Appendix, Eq. 2). The model uses the same rule to change weights for all connections, independently of stimuli, responses, contingencies, and networks. This commonality is another sense in which the model assumes that learning is general. I will argue in the discussion that this assumption does not preclude biological constraints on learning.

Performance is interpreted as effector (muscle, gland) functioning dependent on brain functioning. The model does not simulate effector functioning. The closest the model gets to performance is the simulation of primary-motor ( $M'$ ) precursors of performance. Still, the model allows for the definition of *response rules* that convert continuous output activations into discrete (e.g., binary) responses.

The model also maintains the traditional behavioral distinctions between two types of stimuli, responding, and contingencies. The first distinction is between exteroceptive sensory stimuli typically used as antecedent controlling cues (CSs in Pavlovian contingencies, discriminative stimuli in operant contingencies), and biologically significant stimuli typically used as unconditioned stimuli or USs in Pavlovian conditioning and primary reinforcers in operant conditioning, such as food and water. This distinction is interpreted as one between activations of different types of input units ( $S$  for cues,  $S^*$  for primary-reinforcers). Such activations are interpreted as primary-sensory ( $S'$  in the basic neuroanatomical organization described above) effects of stimulation.

The second distinction is between responding  $R$  that is emitted (not unconditionally elicited by  $S^*$ ) and responding  $R^*$  that is unconditionally elicited by  $S^*$ . Both can come to be controlled by  $S$ . This distinction is neuroscientifically interpreted as one between different types of output activations. These activations are not intended to simulate responses per se, but primary-motor ( $M'$ ) precursors of responses. The third distinction is between Pavlovian (response-independent) and operant (response-dependent) contingencies. This distinction is interpreted as one between independence and dependence of  $S^*$  activations on  $R$  activations.

Associations are interpreted as connections, associative strengths as weights.  $S$ - $S^*$  and  $S$ - $R^*$  associations in Pavlovian conditioning, and  $R$ - $S^*$  associations in operant conditioning, are interpreted somewhat differently from the ways they sometimes are depicted in the literature. Some authors (e.g., Holland, 1990, Figure 1) depict associations as elementary direct links that would be depicted in the model as direct input  $S$ - $S^*$ ,  $S$ - $R^*$ , or  $R$ - $S^*$  connections.

However, there are no such direct connections in this model. Instead, the model follows depictions where separate paths for CS and US converge somewhere (e.g., Wagner, 2008, Figure 3). In this spirit, different types of associations are interpreted in connectionist fashion, as distributed throughout pathways that consist of multiple connections in a network.  $S$ - $S^*$  associations are thus interpreted as pathways consisting of  $S$ - $S''$ ,  $S''$ - $M''$ ,  $M''$ - $D$ ,  $M''$ - $R^*$ ,  $S^*$ - $R^*$ , and  $S^*$ - $D$  connections, where  $S$  and  $S^*$  converge on  $R^*$  and  $D$ .  $S$ - $R^*$  associations are interpreted as pathways consisting of  $S$ - $S''$ ,  $S''$ - $M''$ , and  $M''$ - $R^*$  connections, where  $S$  can evoke  $R^*$  indirectly through  $S''$  and  $M''$ , after learning.



These  $S$ - $R^*$  connections are a proper subset of the  $S$ - $S^*$  connections. The model thus hypothesizes Pavlovian conditioning to involve both,  $S$ - $S^*$  and  $S$ - $S^*$  associations.

It is unclear whether and how the model allows for connections that could be interpreted as  $R$ - $S^*$  associations in operant conditioning, but I will not discuss this here. However, most networks used in this study, which I describe next, include  $R^*$ - $R$  connections that can be interpreted as *response-response* associations. Others (e.g., Dragoi, 1997) have postulated such associations for operant conditioning, but here I postulate them for operant-Pavlovian relations. The model also allows for  $R$ - $R$  connections in operant conditioning, but they will not be part of this study.

## Neural Network Examples

In order to illustrate in more detail how these assumptions work out in the interpretation of misbehavior with the model, Figure 1 shows the three networks that were used in the simulation. They are labeled as INT for interference (top), COM for compatibility (bottom left), and IND for independence (bottom right). I intend them as neural network interpretations of the three possible relations between Pavlovian ( $R^*$ ) and operant ( $R$ ) responding that Timberlake et al. (1982) hypothesized in their account of misbehavior as Pavlovian conditioned responding.

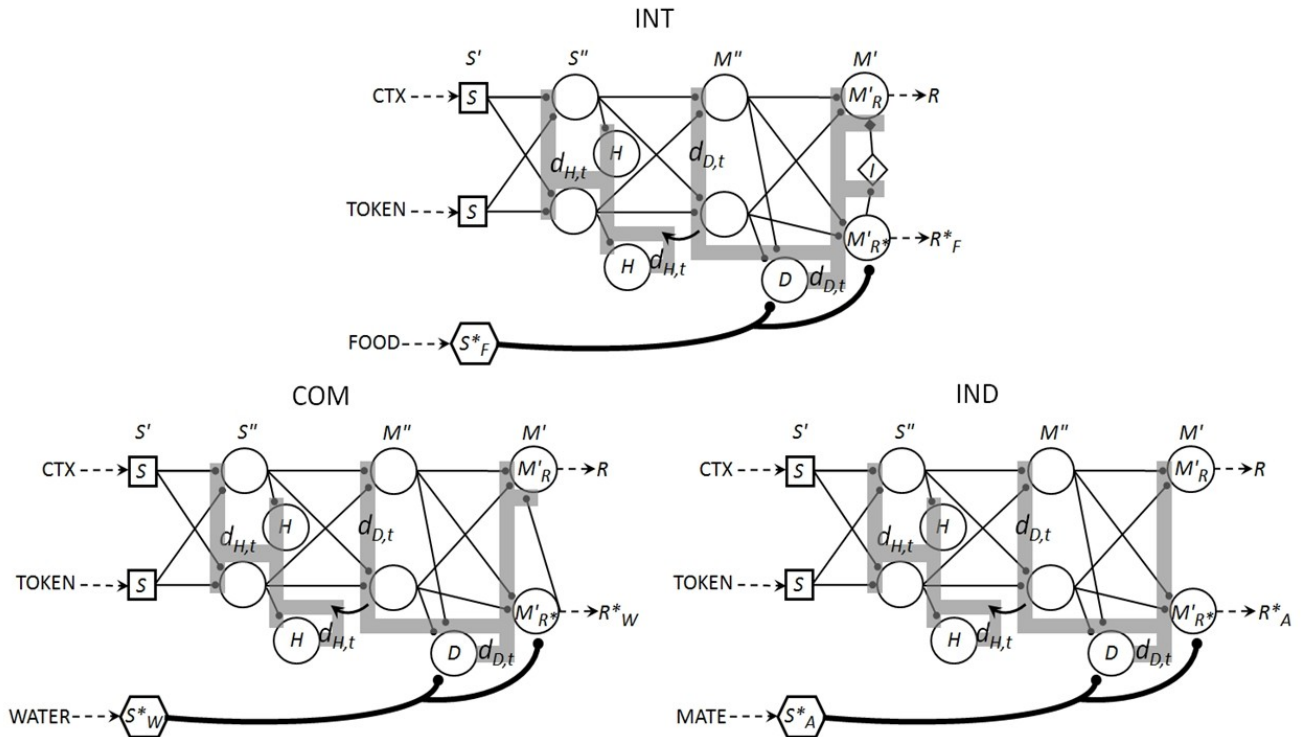
Like other networks in previous simulation research with the model, these have the same basic  $S'$ - $S''$ - $H$ - $S''$ - $M''$ - $D$ - $M''$ - $M'$  organization outlined before, with a few additional details. The networks are intended as connectionist interpretations of neural circuits underlying conditioning with three biologically significant stimuli, namely, food ( $F$ ), water ( $W$ ), and sexual mate ( $A$ ) that elicit three types of unconditioned responses  $R^*_{F}$ ,  $R^*_{W}$ , and  $R^*_{A}$ , respectively, in the same individual animal (e.g., a particular rat). Thus,  $R^*_{F}$  denotes food-related (e.g., chewing),  $R^*_{W}$  water-related (e.g., licking), and  $R^*_{A}$  mate-related (e.g., copulating) responding. Emitted responding ( $R$ ) is assumed to have the same form (e.g., making contact with a token) across all networks.

In INT,  $S^*_{F}$  activations are intended to simulate primary-sensory effects of food that unconditionally elicits food-related responding ( $R^*_{F}$ , e.g., chewing). In COM,  $S^*_{W}$  activations are intended to simulate primary-sensory effects of water that unconditionally elicits water-related responding ( $R^*_{W}$ , e.g., licking). In IND,  $S^*_{A}$  activations are intended to simulate primary-sensory effects of a mate that unconditionally elicits mate-related responding ( $R^*_{A}$ , e.g., copulating). These interpretations are closely tied to the differences in the  $M'$  layers across the three networks.

Units are organized according to a *feedforward* architecture into one input (labeled as  $S'$ , consisting of two squares labeled as  $S$  and a hexagon labeled as  $S^*_{F}$ ,  $S^*_{W}$ , or  $S^*_{A}$ ), two hidden (labeled as  $S''$  for “polysensory” or “sensory-association”, and  $M''$  for “motor-association”), and one output layer labeled as  $M'$  (for “primary-motor”). In agreement with well-established, basic gross-neuroanatomical principles (see Mesulam, 1998),  $S'$  units connect to  $S''$  units,  $S''$  units connect to  $M''$  units, and  $M''$  units connect to  $M'$  units.

I propose the differences in the  $M'$  layers as neural-network interpretations of the operant-Pavlovian relations hypothesized by Timberlake et al. (1982). More precisely, for certain biologically significant stimuli (viz., food and water), I hypothesize projections from primary-motor precursors of Pavlovian responding ( $M'_{R^*}$ ) to primary-motor precursors of emitted, operant responding ( $M'_R$ ). These projections are depicted as lateral connections in the  $M'$  layers of INT and COM. The model allows for other interpretations, but I will focus on this one.

In INT, such lateral connections involve an inhibitory unit (open diamond labeled as  $I$ ) that receives a lateral excitatory connection from  $M'_{R^*}$  and sends a lateral inhibitory connection (thin line ending with a closed diamond) to  $M'_R$ . My working hypothesis here is that interference of  $R$  by misbehavior qua conditioned  $R^*$  is due, at least in part, to lateral inhibition from primary-motor precursors of  $R^*$  to primary-motor precursors of  $R$ . For the simulation, then,  $M'_{R^*}$  activations in INT are expected to significantly lower  $M'_R$  activations. On this basis, INT should simulate misbehavior qua conditioned  $R^*$  responding that significantly interferes with  $R$  and hence reduces success in obtaining  $S^*$  in operant contingencies. My interpretation of  $R^*$  in this network as food-related follows observations that misbehavior interferes with operant responding when food is used as a reward in operant contingencies.



**Figure 1.** Neural network architectures used in the simulations, labeled as INT for “interference” (top), COM for “compatibility” (bottom left), and IND for “independence” (bottom right). The networks differ only in their output ( $M'$  for “primary-motor”) layer. INT’s  $M'$  layer has an inhibitory unit (diamond labeled as  $I$ ) that receives a lateral excitatory connection from  $M'_{R^*}$  (output unit for  $R^*$  responding, unconditionally elicited by a primary reinforcer, or conditionally evoked by a cue) and sends a lateral inhibitory connection to  $M'_R$  (output unit for  $R$ , emitted responding). COM’s  $M'$  layer has a direct lateral excitatory connection from  $M'_{R^*}$  to  $M'_R$ . IND’s  $M'$  layer has no lateral connections. CTX: Constellation of contextual cues. TOKEN: Cues from a discrete object (e.g., ball-bearing, wooden coin), a response  $R$  to which (e.g., contact) is required for an operant contingency. TOKEN is present (i.e., this  $S$  unit was activated) only when an  $R$  response occurs, defined as an  $M'_R$  activation of 0.5 or more.  $S'$ : Primary-sensory or input layer.  $S'$  units are activated

according to a training protocol that simulates a conditioning procedure, and are meant to simulate primary-sensory effects of external environmental stimuli. Squares labeled as  $S$ : Input units whose activations are intended to simulate primary-sensory effects of CTX and TOKEN. Hexagons labeled as  $S^*_F$ ,  $S^*_W$ , and  $S^*_A$ : Input units whose activations are intended to simulate primary-sensory effects of biologically significant stimuli (e.g., food, water, and mate, respectively) that unconditionally elicit three different response forms  $R^*_F$  (e.g., chewing)  $R^*_W$  (e.g., licking), and  $R^*_A$  (e.g., copulating), respectively. These responses can also be conditionally evoked by CTX and TOKEN, if the relevant connections have gained sufficient weight. Dashed arrows to inputs: Stimulus-to- $S$  transduction (not modeled). Circles: Neurocomputational units activated according to the activation rule (see Appendix, Eq. 1). Thin lines ending with buttons: Variable, initially weak connections whose weights change according to the learning rule (see Appendix, Eq. 2). Thick lines arising from hexagons: Fixed, initially strong connections.  $S''$ : Sensory-association or polysensory layer, consisting of two units.  $H$ : Units intended to simulate a hippocampal area (e.g., CA1), as a source of  $d_{H,t}$  (signal that modulates changes in the  $S$ - $S''$  and  $S''$ - $H$  connections).  $D$ : Unit intended to simulate some dopaminergic nucleus (e.g., ventral-tegmental area), unconditionally activated by  $S^*$ , and source of  $d_{D,t}$ , the discrepancy signal that modulates changes in  $S''$ - $M''$ ,  $M''$ - $D$ , and  $M''$ - $M'$  connections, according to the learning rule. Shaded rectangles: Diffuse discrepancy signals ( $d_{H,t}$  and  $d_{D,t}$ ) that modulate weight changes. Curved arrows: Amplifications of  $d_{H,t}$  by  $d_{D,t}$ .  $M''$ : Motor-association layer, consisting of two units ( $M'_{R^*}$ ,  $M'_{R^*}$ ).  $M'$ : Primary-motor (output) layer.  $M'$  activations are intended to simulate primary-motor precursors of  $R$  ( $M'_{R^*}$ ) and  $R^*$  ( $M'_{R^*}$ ) responding. Dashed arrows from outputs: Primary-motor-to-effector transduction (not modeled).

To simulate Boakes et al.'s (1978) observation of less interfering misbehavior when  $R$  is rewarded with water, I hypothesize that  $R^*_W$  is more compatible with  $R$  than  $R^*_F$ . My rationale for this hypothesis is that water is far more biologically significant than food, as dehydration usually causes death much faster than starvation. On this basis, I speculate that  $R^*_W$  is sufficiently faster than  $R^*_F$  as to make  $R^*_W$  at least less interfering with  $R$ . Moreover, to simulate Timberlake et al.'s (1982) interpretation of operant-Pavlovian compatibility as being facilitating, I speculate that  $R$ - $R^*_W$  compatibility could be excitatory in nature (but see below). The clearest way to translate this hypothesis into the model is by placing an excitatory connection from  $M'_{R^*}$  to  $M'_{R^*}$ , as shown in COM. This network should thus simulate misbehavior that interferes less with  $R$  than in INT.

Lastly, IND's  $M'$  layer has no lateral connections, which I intend to simulate a primary-motor independence between  $R$  and  $R^*$  responding, as a neural network interpretation of Timberlake et al.'s (1982) hypothesized independence between misbehavior qua conditioned Pavlovian responding and operant responding. A behavioral interpretation of this network is less forthcoming, but I venture to postulate mate-related responding ( $R^*_A$ ) as a possible candidate, although a good rationale for this hypothesis is elusive. The best rationale I can offer is that  $R^*_A$  is sufficiently different from  $R$ ,  $R^*_F$ , and  $R^*_W$  as to be neurally more independent of  $R$  than  $R^*_F$  and  $R^*_W$ . IND should thus simulate misbehavior that is less interfering with  $R$ .

From my hypothesis that compatibility is excitatory, COM might be expected to simulate less interfering misbehavior than IND. However, this implication might not obtain. The reason is that the learning rule includes a factor where units that are connected to the same unit compete for a limited amount of weight available on this unit (this competition occurs only between units of the same type, either excitatory or inhibitory, not between different types). As connections gain weight, then, there is less weight to be gained by other connections on the same unit. Such is the case of COM, where the  $M'_{R^*}$  and  $M''$  units compete for the weight available on  $M'_{R^*}$ . As a result, activations of  $M'_{R^*}$  by either the  $M'_{R^*}$  or  $M''$  units might be weaker. There is no such

competition or inhibition in IND's  $M'$  layer. IND should thus simulate less interfering misbehavior than COM and INT.

The three networks are otherwise identical. The  $S'$  layer has three units (squares labeled as  $S$  and hexagon labeled as  $S^*_{F}$ ,  $S^*_{W}$ , or  $S^*_{A}$ ). These units are intended to simulate primary-sensory neuronal groups and are activated according to a predesigned training protocol that simulates a conditioning procedure of interest. Activations of the  $S$  units are intended to simulate primary-sensory effects of CSs, discriminative stimuli, and contextual cues. For this study, I assume that  $S$  in Pavlovian  $S$ - $S^*$  contingencies (through which misbehavior is supposedly acquired and maintained) not only denotes sensory cues from a discrete object (TOKEN, contact with which is required for  $S^*$  in operant,  $R$ - $S^*$  contingencies), but also contextual cues (CTX).

$S^*_{F}$ ,  $S^*_{W}$ , and  $S^*_{A}$  activations are intended to simulate primary-sensory effects of three biologically significant stimuli (e.g., food, water, and mate, respectively) used as response-independent USs in Pavlovian contingencies, or response-dependent primary rewards in operant contingencies. The dashed arrows to the  $S'$  units depict sensory transduction processes, not simulated in the model.

Both  $S$  units connect to, and hence can activate, both  $S''$  units in the first hidden layer. These connections constitute only part of the  $S'$ - $S''$  segment of the basic  $S'$ - $S''$ - $H$ - $S''$ - $M''$ - $D$ - $M''$ - $M'$  architecture described above. The hexagons, albeit  $S'$  units, do not project to the  $S''$  layer, so they are not part of the  $S'$ - $S''$  segment. Hence, this segment is more precisely symbolized as  $S$ - $S''$ .

Each  $S''$  unit, in turn, connects to, and hence can activate its own  $H$  unit. These connections constitute the  $S''$ - $H$  segment of the basic architecture. Both  $S''$  units also connect to, and hence can activate the two  $M''$  units in the second hidden layer. These connections constitute the  $S''$ - $M''$  segment of the basic architecture, and are intended to simulate a sensory-motor interface. Both  $M''$  units, in turn, connect to the  $D$  unit (comprising the  $M''$ - $D$  segment), and both  $M'$  (primary-motor, output) units (comprising the  $M''$ - $M'$  segment). These networks, then, are fully connected from layer to layer (cf. Sánchez et al., 2010). Because connections are initially weak, they must gain sufficient weight for units to activate their targets (e.g., an  $S$  unit to activate its target  $S''$  units). Two  $H$  and one  $D$  unit were used based on a previous study (Sánchez et al., 2010).

Changes in connection weights are modulated by diffuse discrepancy signals (shaded rectangles arising from  $H$  and  $D$  units). The signal that arises from the  $H$  units ( $d_{H,t}$ ) modulates changes in the weights of the  $S$ - $S''$  and  $S''$ - $H$  connections. The signal that arises from  $D$  ( $d_{D,t}$ ) modulates changes in the weights of the  $S''$ - $M''$ ,  $M''$ - $D$ , and  $M''$ - $M'$  connections and amplifies  $d_{H,t}$  (curved arrow). These signals are assumed to be carried by connections that simulate axo-axonic, modulatory synapses that feed back into the feedforward synapses, and are assumed to be fixed and maximally strong.

The key to how such signals increase weights is the unconditional activation of  $D$  by  $S^*$ , which is intended to simulate unconditional activations of dopaminergic neurons by primary reinforcers. If the  $S^*$  unit is maximally activated (with level of 1.0) at  $t$ , then the  $D$  unit will also be activated at  $t$  with the same level as  $S^*$ . In a naïve network, initial weights are too low to allow  $S$  units to strongly activate  $D$  units via the  $S''$  and  $M''$  units. Hence,  $D$ 's activation at  $t-1$  will be close to 0.0 (lower than a pre-established

discrepancy threshold of 0.05, which puts the learning rule in weight-gain mode). When  $D$  is unconditionally and strongly activated by  $S^*$  at  $t$ , the difference between this activation and the one at  $t-1$  will be large. This difference will make all connections gain weight at  $t$  (which makes it a diffuse signal). After several iterations of this process, weights become sufficiently large for the  $S$  units to start *conditionally* activating the  $D$  units, which also results in relatively large discrepancies (not as large as in unconditional activations of  $D$  by  $S^*$ ). These discrepancies act as a kind of internal conditioned reinforcement mechanism that allows for further weight gain in the absence of  $S^*$ . In this way, learning can keep going for sometime without primary reinforcement.

$M'_R$  and  $M'_{R^*}$  activations are intended to simulate primary-motor precursors of two response forms  $R$  (emitted) and  $R^*$  (elicited), respectively. These units are not intended to simulate effectors (muscles, glands), nor are their activations intended to simulate responses. The model only assumes that certain appropriate effectors are in place and functioning properly, such that there is a sufficiently high correlation between primary-motor activations and responses. Therefore, the model does not simulate any manipulations that are designed to disrupt effector functioning directly without disrupting primary-motor areas. The only way the model can simulate such disruption is by simulating disruptions of primary-motor areas.

$M'_R$  can be activated only by the  $S$  via the  $S''$  and  $M''$  units, if the corresponding connections ( $S$ - $S''$ ,  $S''$ - $M''$ , and  $M''$ - $M'_R$ ) gain sufficient weight.  $M'_R$  cannot be activated by an  $S^*$  unit.  $M'_R$  activations are thus intended to simulate primary-motor precursors of a response  $R$  that is emitted in that it is not unconditionally elicited by  $S^*$ . Still,  $M'_R$  must be activated by the  $S$  units.  $R$  thus simulates responding that is controlled by  $S$ , even if not unconditionally elicited by  $S^*$ . All emitted responding in this model is thus hypothesized to have some antecedent controlling (but not eliciting) stimulation.

$M'_{R^*}$ , in contrast, can be activated either unconditionally by  $S^*$  or conditionally by the  $S$  units via the  $S''$  and  $M''$  units (which requires sufficient weight gain by the  $S$ - $S''$ ,  $S''$ - $M''$ , and  $M''$ - $M'_{R^*}$ ). An unconditional activation occurs whenever an  $S^*$  activation is greater than 0.0, which simulates the occurrence of a biologically significant stimulus, such as food, water, or a mate. In this case,  $M'_{R^*}$ 's activation will be equal to the  $S^*$  activation (see Appendix, Eq. 1), which would simulate an unconditionally elicited response. Unconditional activations are carried by a fixed and maximally strong connection (thick line) from  $S^*$  to  $M'_{R^*}$ .  $D$  too receives a direct connection from  $S^*$ . The  $S^*$ - $M'_{R^*}$  and  $S^*$ - $D$  connections should thus be added to the basic architecture.

These fixed connections are intended to simulate hardwired, innate, species-specific dendritic/somatic synaptic clusters that mediate automatic activations of dopaminergic systems (activation of  $D$  by  $S^*$ ) and primary-motor precursors of responding that is unconditionally and specifically elicited by  $S^*$  (activation of  $M'_{R^*}$  by  $S^*$ ).  $M'_{R^*}$  activations by  $S^*$  are thus intended to simulate neural substrates of an unconditioned reflex  $S^*$ - $R^*$ .

Conditional activations of  $M'_{R^*}$  are like those of  $D$  and  $M'_R$ . They can be caused only by the  $S$  via the  $S''$  and  $M''$  units (only if the  $S^*$  unit activation is 0.0, which simulates the absence of  $S^*$ , and the  $S$ - $S''$ ,  $S''$ - $M''$ , and  $M''$ - $M'_{R^*}$  connections have gained sufficient weight). The  $S$  units can thus activate the  $S''$  units, the  $S''$  units can activate the  $H$  and  $M''$  units, the  $M''$  units can activate the  $D$  and  $M'$  units. In sum,  $R$  and  $R^*$  are

intended to simulate two different response systems, one emitted ( $R$ ), the other elicited ( $R^*$ ), both controlled by certain environmental stimuli ( $S$  in the case of  $R$ ,  $S$  or  $S^*$  in the case of  $R^*$ ).

$R$  simulates another feature of emitted responding, namely, directedness (see Burgos 2007). This feature was simulated by making activations of the TOKEN  $S$  input unit depend on  $R$  responses during operant contingencies. These activations were intended to simulate physical contact with a token. Such contact, in turn, was hypothesized to provide sensory (visual, tactile, etc.) feedback from the token's features. These features are thus hypothesized to come to function as antecedent controlling stimulation of  $R$  and  $R^*$ , as a result of learning.

### **Some Comparisons with Other Models**

Detailed and exhaustive formal comparisons would take too long. Instead, I will only make some general informal comparisons. One important similarity is the use of a temporal difference or "error" in successive activations of  $D$  to modulate learning (see Appendix, Eq. 2). Many models of conditioning (e.g., Rescorla & Wagner, 1972) use this magnitude, and the present one is no exception. The model also makes a learning-performance distinction that favors a definition of learning in terms of underlying mechanisms, not performance. And like some other models (e.g., Schmajuk, Lam, & Gray, 1996), the present one is explicitly connectionist. As mentioned before, the model can also simulate many phenomena that other models simulate, and a few that they cannot, such as autoshaping and automaintenance.

I have pointed out some differences above. The model, in contrast to others, postulates a common learning mechanism for Pavlovian and operant conditioning, while making standard behavioral and neuroanatomical distinctions. Other models of Pavlovian conditioning (e.g., Gibbon & Balsam, 1981; Rescorla & Wagner, 1972; Schmajuk et al., 1996; Stout & Miller, 2007; Sutton & Barto, 1981; Wagner, 1981) do not predict operant conditioning, as they do not include emitted responding. Other models of operant conditioning (e.g., Dragoi & Staddon, 1999; Grossberg, 1974; Killeen, 1994, 2011; Machado, 1997; Staddon & Zhang, 1991) include emitted responding but do not simulate Pavlovian conditioning. In contrast to the present model, they also assume that acquisition and maintenance of emitted responding requires operant contingencies. Also, the model is more inspired by neuroscience than most others. At least, they do not make the kinds of neuroscientific assumptions laid out above.

To finish this section, a previous computational study of misbehavior (Dayan, Niv, Seymour, & Daw, 2006) deserves mention. The model in that study, like the present one, uses a temporal difference to modulate learning. However, the authors' interpretation of misbehavior is quite different. They view failures to obtain reinforcers in misbehavior as a kind of omission training. This interpretation is valid, but not the only possible one. Mine is more akin to other studies of misbehavior (Boakes et al., 1978; Timberlake et al., 1982). It also makes behavioral (e.g., CS-US, and emitted-elicited distinctions) and neuroscientific (e.g., roles of dopaminergic and hippocampal systems in conditioning) assumptions that are not made in that model. Not only do

such assumptions make the present study somewhat closer to previous behavioral work and known neural substrates of conditioning, but also allow for novel predictions.

## A Simulation

The purpose is to simulate misbehavior toward a token in operant contingencies only, where contact with the token is necessary for reinforcement, using the model described above. The three networks depicted in Figure 1 were naïve in that the initial weights of all their variable connections were relatively low, set to 0.1 (the minimum weight is 0.0, the maximum weight is 1.0).

All networks first received 100 CTX- $S^*$  Pavlovian conditioning pairings. CTX was defined as a maximal (1.0) activation of the top  $S$  unit throughout the entire phase. This activation was intended to simulate primary-sensory effects of contextual cues, permanently present throughout training.  $S^*$  was defined as the maximal activation of the  $S^*_F$ ,  $S^*_W$ , and  $S^*_A$  units every 5 moments. This training protocol thus simulated a 5-moment fixed-time schedule in a context. The other  $S$  (TOKEN) unit was never activated during this phase.

This pretraining is necessary in these networks to promote sufficient weight gains to facilitate activations throughout the network and thus increase the likelihood that the response criterion during operant reinforcement is met. Such learning provides a Pavlovian kick-start that yields some preparatory stimulus control of  $R$ . In the case of food and water, this pretraining could be plausibly interpreted as simulating training with a food or water dispenser, pretraining that has been used in misbehavior studies. There still have not been systematic studies of sexual misbehavior. However, some studies of sexual instrumental conditioning (e.g., Everitt, Fray, Kostarczyk, Taylor, & Stacey, 1987) with rats have included some pretraining with sexual mates (e.g., visual exposure), before operant conditioning.

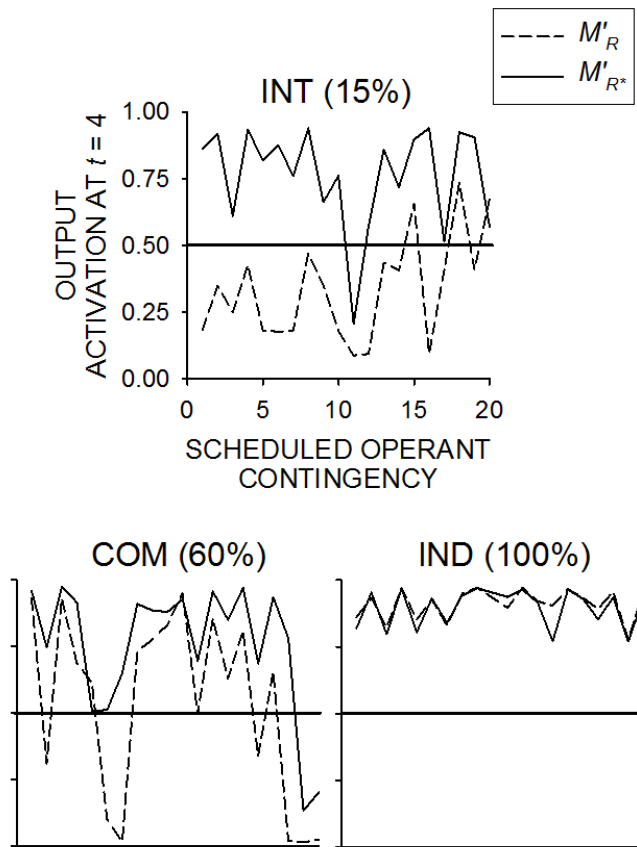
Then, in a second phase, the three networks were scheduled to receive 20 operant contingencies where the  $S^*$  units were maximally activated every five moments only if  $M'_R$ 's activation at the last moment was 0.5 or more. This threshold can also be used to define a response rule for misbehavior, applied to  $M'_{R^*}$  activations. This protocol thus simulated a five-moment fixed-interval schedule with contextual cues. Also during this phase, the TOKEN  $S$  unit (second square from top to bottom in the networks in Figure 1) was maximally activated whenever an  $R$  response thus defined occurred. This activation was intended to simulate physical contact with a token, a contact that I hypothesize to provide sensory feedback from the object that can also function as antecedent controlling stimulation (Burgos, 2007), in addition to contextual cues.

For simplicity, all  $S$  activations (CTX and TOKEN) were simulated as discrete trials. No explicit intertrial intervals were simulated. Instead, they were assumed to be sufficiently long for all activations to decrease to near-zero values at the beginning of each trial. All activation and learning free parameters were the same as in previous simulation research with this model (see Appendix), for all units.

The results of the first phase are not shown, as output activations did not increase to response levels until about 90 to 100 trials. Pavlovian contextual conditioning thus took this long to occur, with INT showing only  $R^*$  responding, and COM and IND also showing some  $R$  responding. The results of the second phase (operant conditioning) are given in Figure 2.

The figure depicts the activations of  $M'_R$  (dashed curve, representing a primary-motor precursor of operant responding) and  $M'_{R^*}$  (solid curve, representing a primary-motor precursor of Pavlovianly conditioned misbehavior) at the moment immediately before a scheduled operant contingency ( $t = 4$ ) for all 20 scheduled contingencies and the three networks (INT: top panel, COM: bottom left panel, IND: bottom right panel). The percentage of reinforcers obtained (in parentheses) was used as a measure of success in obtaining reinforcers during this phase.

These results are from a single simulation run (i.e., they simulate a study with three different individual subjects of the same species; e.g., *Rattus norvegicus*). The simulation was repeated 100 times, to determine reliability. Differences were observed across runs (due to stochastic noise in the model; see Appendix), but casual visual inspection suggested that the results shown here are reasonably representative. Quantitative analyses remain to be done to assess the statistical properties of the model's simulation of the phenomenon. I did not do such analyses for this study because they would have complicated it well beyond what I intended. My aim was first to test the model with individual networks, a common practice in connectionist modeling.





**Figure 2.** Activations of  $M'_R$  (dashed curve) and  $M'_{R^*}$  (solid curve) at the moment immediately before a scheduled operant contingency, across all scheduled contingencies, during operant conditioning, after a pretraining in Pavlovian contextual conditioning during feeder/magazine training, for the three networks used, INT (top), COM (bottom left), and IND (bottom right). The horizontal line at 0.5 indicates the output activation criterion for responses. Percentages of reinforcers obtained during the phase are shown in parentheses.

The networks showed different levels of  $M'_R$  and (conditioned)  $M'_{R^*}$  activations, to different degrees of stability. Except for a few occasions, all  $M'_{R^*}$  activations in the three networks were above the activation criterion for responses (indicated by the horizontal line at 0.5), and from the beginning of the phase. This result means that all networks simulated misbehavior qua conditioned  $R^*$  responding most of the time (assuming properly functioning effectors) and from the outset. Such early onset of misbehavior is not surprising, as, again, the networks had received in the previous phase 100 CTX- $S^*$  trials, which promoted strong Pavlovian contextual control of  $R^*$ .

These results thus simulate the acquisition of misbehavior qua conditioned  $R^*$  responding acquired through Pavlovian contextual conditioning during pretraining with food, water, or a sexual mate. Still,  $R^*$  occurred to different degrees of stability across the three networks, with IND being the most stable (all of IND's  $M'_{R^*}$  activations were well above the 0.5 criterion). INT and COM simulated comparably less stable misbehavior. The reason for these differences has to do with the  $M'_R$  activations, which I describe next.

The most pronounced differences across the networks were in their  $M'_R$  activations, intended to simulate primary-motor precursors of  $R$  responding. As expected, INT showed the fewest  $M'_R$  activations above the 0.5 response criterion (only three towards the end of the phase). INT thus simulated the weakest operant responding. This result was due to  $M'_R$ 's inhibition by  $M'_{R^*}$  in INT's output layer. This inhibition can be appreciated in the opposite tendencies often observed between INT's  $M'_R$  and  $M'_{R^*}$  activations, although their Spearman rank order correlation ( $\rho$ ) was below 0.3 and statistically non-significant.

Relevant final weights of this phase for INT were about 0.39 for each  $M''-M'_R$ , 0.85 for the  $I-M'_R$ , and 0.53 for the  $M'_{R^*}-I$  connection. Thus, inhibition of  $M'_R$  by  $M'_{R^*}$  was quite strong, which is a way to simulate the interference that Boakes et al. (1978) and Timberlake et al. (1982) have hypothesized, according to which food-related misbehavior exerts on operant responding. In sum, INT simulated considerable interference of operant responding by misbehavior resulting in the most failures to obtain reinforcers (only a success of 15%), in ordinal agreement with evidence of misbehavior where food has been used as a reward.

COM simulated comparable levels of misbehavior, but noticeably more  $R$  responding and, hence, less interfering misbehavior than INT, dramatically increasing COM's success to obtain reinforcers (60% vs. INT's 15%). Such success was due to  $M'_{R^*}$ 's excitatory connection to  $M'_R$ , which allowed  $M'_{R^*}$  activations to contribute to  $M'_R$  activations. The two were much more highly correlated ( $\rho = 0.86, p < .0001$ ). Relevant final relevant weights of this phase for COM were about 0.07 for each  $M''-M'_R$  and 0.38 for the  $M'_{R^*}-M'_R$  connection. This difference means that the  $R^*-R$  association was stronger than the  $S-R$  associations, due to a competition between  $M'_{R^*}$  and the  $M''$  for the

weight available on  $M'_R$ .  $M'_{R^*}$  had competitive advantage, due to its stronger activation by  $S^*$  at the moment of reinforcement. Still, the  $R^*-R$  connection did not gain sufficient weight for  $M'_{R^*}$  to activate  $M'_R$  more stably. COM's simulation thus simulates less interference of operant responding by misbehavior than INT, but perhaps less than what Timberlake et al. (1982) appeared to have predicted for compatibility. This result is also ordinally comparable to Boakes et al.'s (1978) result of less interfering misbehavior when water is used as a reward, although quantitatively the present results simulate a more pronounced difference.

The relative instability of misbehavior in INT and COM was due to the instability of  $R$  responding, which, in turn, reduced the frequency of  $S^*$  in the operant contingencies. The source of such instability was different for each network. In INT, the source was the inhibition of  $M'_R$  by  $M'_{R^*}$ . In COM, the source was the competition between the  $M'_{R^*}$  and  $M''$  units for the weight available on  $M'_R$ . Both caused a reduction in the  $S^*$  frequency, which resulted in occasional weight losses in the CTX- $R^*$  paths, which occasionally decreased  $R^*$  control by CTX. On other occasions,  $M'_R$  activations increased sufficiently to cause  $S^*$  and weight regain. The net effect of this was instability in the  $M'_R$  and  $M'_{R^*}$  activations in INT and COM.

As expected, IND showed the highest, most stable misbehavior and operant responding, and hence, the most successes in obtaining reinforcers (100%), with an  $R-R^*$  correlation comparable to COM's ( $\rho = 0.87, p < .0001$ ). Relevant final weights of this phase for IND were about 0.4 for each  $M''-M'_R$  connection, indicating a much stronger control of  $R$  by  $S$  than in COM. These weights were comparable to the corresponding ones in INT. However, IND's  $M'_R$  activation was not damped by any inhibition, nor was there the sort of competition that interfered with COM's  $R$  responding. All this helped IND's  $R$  responding to be more strongly controlled by the  $S$  (CTX and TOKEN). This result is a novel prediction, so there is no evidence with which to compare it. It remains to be seen whether it is confirmed in future research with animals.

## General Discussion

In terms of my proposed interpretation of the three networks as stimulations of neural circuits underlying different reinforcement and response systems, the results predict an order of interference of operant responding by misbehavior that relates to different biologically significant stimuli that unconditionally elicited their different responses. From the most to the least interfering, the predicted order is food, water, and mate. This order is predicted to depend on certain neuroanatomical features of those circuits, in particular, connectivity among primary-motor precursors of misbehavior and operant responding. It remains to be seen whether other models that have been used to simulate misbehavior (e.g., Dayan et al., 2006) also make these predictions.

The first and second tiers of this ordering, corresponding to food and water, respectively, are ordinally comparable to Boakes et al.'s (1978) observation of less interfering misbehavior with water than food. However, the differences between food- and water-related misbehavior were quantitatively more pronounced in the simulation.

As for the third tier, corresponding to mate-related misbehavior, there is no evidence to assess the simulation's prediction. Studies on sexual instrumental conditioning (e.g., Beck, 1971; Everitt et al., 1987; Gilbertson, 1975; Michael & Keeverne, 1968; Sevenster, 1973) have not yet investigated misbehavior systematically. More experimental research is thus needed for a more detailed behavioral characterization of mate-related misbehavior in sexual instrumental conditioning, and how much it interferes with operant responding, compared to misbehavior with food and water.

The simulation also predicts the occurrence of misbehavior qua conditioned  $R^*$  responding for all the networks from the very first scheduled operant contingency. However, once again, this effect resulted after about 100 trials of Pavlovian pretraining (not shown), which promoted Pavlovian contextual conditioning of  $R^*$  responding in all the networks. Unfortunately, there is little evidence to which this result can be directly compared, as no systematic data on Pavlovian contextual conditioning during pretraining have yet been reported. The closest is Timberlake et al.'s (1982, Experiment 1, Figure 1, upper left panel) results for acquisition speed of misbehavior through Pavlovian contingencies using a token as a CS. Their results indicate that acquisition of misbehavior through Pavlovian contingencies is relatively fast (about five trials). The present results depart considerably from this particular evidence.

A possible reason is that the initial connection weights in the networks were too low. Perhaps, faster acquisition of misbehavior could be simulated with larger initial weights. In any case, the present results are consistent with a key implication of Timberlake et al.'s (1982) study: Contrary to the conventional wisdom at the time, operant contingencies are *not* necessary for the acquisition of misbehavior.

Other limitations stem from the great deal of abstraction involved in the model and simulation. To begin with, misbehavior was simulated in a very simplified way as consisting of just one behavior form  $R^*$  (cf. Timberlake et al., 1982, Table 1). Misbehavior in animals is far more complex than this, involving many  $R^*$  behavior forms that can vary and relate in many different ways. Also, the networks used were exceedingly simple (far simpler than the simplest known natural brain circuit). Natural brain circuits that underlie misbehavior, whatever they are, are likely to be far more complex than this.

Another limitation was the necessity of a Pavlovian training previous to operant conditioning for these networks to simulate some  $R$  responding that could be used in operant contingencies. For INT and COM, I interpreted such training as corresponding to pretraining with food and water dispensers. There are no studies of misbehavior with sexual mates, but this interpretation can be applied to IND as well. The relevant prediction here is that some sort of non-operant training with biologically significant stimuli, prior to operant training, is necessary for misbehavior to occur. It remains to be seen whether this prediction is empirically confirmed.

Also, contextual cues are tonic, continuous signals, but were simulated here as discrete CSs. This simplification has been convenient for theoretical simplification, but it ultimately is too simple. The present model allows for a somewhat more realistic simulation of contextual cues (e.g., Burgos & Murillo-Rodríguez, 2007) where they have a less discrete, more continuous character. It remains to be seen whether and how this way to simulate contextual cues affects the present results.

No individual or species differences were simulated either. Only differences in response and reinforcement systems within the same individual were simulated. It remains to be seen whether and how the model predicts species-specific differences in misbehavior, although such differences have not been experimentally investigated either. In principle, the model at least allows for manipulations of neural network architectures that could simulate species-specific neuroanatomical differences underlying species-specific learning differences. Because such differences obtain at the neuroanatomical and behavioral levels, they are compatible with a general mechanism of changes in synaptic efficacies that obtains at the synaptic level. Generality and specificity thus become compatible by obtaining at different levels of organization.

In a related simplification, the networks were handcrafted ad hoc to simulate the phenomenon of interest. I intended this abstraction only as a conjecture about possible neuroanatomical substrates of misbehavior, also to be tested empirically. Still, the abstraction refers only to proximate causes of behavior. The study of misbehavior and other biological constraints on learning has been squarely framed within a strong evolutionary, adaptive thinking that also seeks for distal, ancestral causes of behavior. Aside from the assumption that unconditioned  $D$  and  $M'_{R^*}$  activations by  $S^*$  simulate primary-motor precursors of innate, species-specific responding, no other evolutionary considerations were made here. However, this does not mean that they are impossible or unimportant.

Neuroanatomical architecture, in particular, is as much a phenotypic trait as height, skin color, and behavior. The networks used here can thus be hypothesized to simulate phenotypic traits that are proximate effects of genetically and environmentally driven developmental processes. They can also be hypothesized to be adaptive adaptations as distal effects of phylogenetic processes. More comparatively, they could be hypothesized as homologies among present vertebrate species. Species similarities are as important as species differences. A previous study with the model (e.g., see Burgos, 1997) shows none of this is beyond the grasp of computational modeling, despite its generalist character.

Finally, the study illustrates how computational modeling can make the idea of biological constraints congruent with a general view of learning. This congruency is made possible by theorizing at different levels of organization of brains, namely, synaptic, neuroanatomical, and behavioral. Learning is hypothesized to be general at the synaptic level by using the same learning rule to change weights across all connections, stimuli, reinforcers, responses, contingencies, and network. Learning is also hypothesized to be biologically specific at the neuroanatomical and behavioral levels, by using different network architectures that are intended to simulate not only differences within the same individual, but also across individuals and even species. Hence, no conflict need obtain between biological constraints and a generalist view of learning.

## Acknowledgments

The author wishes to thank Robert Boakes, Michael Domjan, and an anonymous reviewer for useful commentaries to a previous draft.

## References

- Bailer-Jones, D. (2009). *Scientific models in philosophy of science*. Pittsburgh, PA: University of Pittsburgh Press.
- Beck, J. (1971). Instrumental conditioned reflexes with sexual reinforcement in rats. *Acta Neurobiologiae Experimentalis*, *31*, 251-262.
- Bitterman, M. E. (1975). The comparative analysis of learning: Are the laws of learning the same for all animals? *Science*, 699-709.
- Boakes, R. A., Poli, M., Lockwood, M. J., & Goodall, G. (1978). A study of misbehavior: Token reinforcement in the rat. *Journal of the Experimental Analysis of Behavior*, *29*, 115-134. <http://dx.doi.org/10.1901/jeab.1978.29-115>
- Breland, K., & Breland, M. (1961). The misbehavior of organisms. *American Psychologist*, *16*, 681-684. <http://dx.doi.org/10.1037/h0040090>
- Brembs, B., & Plendl, W. (2008). Double dissociation of pkc and ac manipulations on operant and classical learning in drosophila. *Current Biology*, *18*, 1168-1171.
- Burgos, J. E. (1997). Evolving artificial neural networks in Pavlovian environments. In J. W. Donahoe, & V. Packard-Dorsel (Eds.), *Advances in psychology 121: Neural- network models of cognition: Biobehavioral foundations* (pp. 58-79). Amsterdam: Elsevier. [http://dx.doi.org/10.1016/S0166-4115\(97\)80090-8](http://dx.doi.org/10.1016/S0166-4115(97)80090-8)
- Burgos, J. E. (2003). Theoretical note: Simulating latent inhibition with selection neural networks. *Behavioural Processes*, *28*, 183-192.
- Burgos, J. E. (2005). Theoretical note: the C/T ratio in artificial neural networks. *Behavioural Processes*, *69*, 249-256. <http://dx.doi.org/10.1016/j.beproc.2005.02.008>.
- Burgos, J. E. (2007). Autoshaping and automaintenance: A neural-network approach. *Journal of the Experimental Analysis of Behavior*, *88*, 115-130. <http://dx.doi.org/10.1901/jeab.2007.75-04>.
- Burgos, J. E. (2010). The operant/respondent distinction: A computational neural-network analysis. In N. Schmajuk (Ed.), *Computational models of conditioning* (pp. 244-271). Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511760402>.
- Burgos, J. E., & Donahoe, J. W. (2000). Structure and function in selectionism: implications for complex behavior. In J. Leslie, & D. Blackman (Eds.), *Issues in experimental and applied analyses of human behavior* (pp. 39-57). Reno, NV: Context Press.
- Burgos, J. E., & García-Leal, Ó. (2015). Autoshaped choice in artificial neural networks: Implications for behavior economics and neuroeconomics. *Behavioural Processes*, *114*, 63-71. <http://dx.doi.org/10.1016/j.beproc.2015.01.010>
- Burgos, J. E., Flores, C., García, O., Díaz, C., Cruz, Y. (2008). A simultaneous procedure facilitates acquisition under an optimal interstimulus interval in artificial neural networks and rats. *Behavioural Processes*, *78*, 302-309. <http://dx.doi.org/10.1016/j.beproc.2008.02.018>
- Burgos, J. E., & Murillo-Rodríguez, E. (2007). Neural-network simulations of two context-dependence phenomena. *Behavioural Processes*, *75*, 242-249. <http://dx.doi.org/10.1016/j.beproc.2007.02.003>.
- Burns, R., Burgos, J. E., & Donahoe, J. W. (2011). Pavlovian conditioning: Pigeon nictitating membrane. *Behavioural Processes*, *86*, 102-108. <http://dx.doi.org/10.1016/j.beproc.2010.10.004>.
- Dayan, P., Niv, Y., Seymour, B., & Daw, N. D. (2006). The misbehavior of value and the discipline of will. *Neural Networks*, *19*, 1153-1160. <http://dx.doi.org/10.1016/j.neunet.2006.03.002>
- Domjan, M., & Galef, Jr., B. G. (1983). Biological constraints on instrumental and classical conditioning: Retrospect and prospect. *Animal Learning & Behavior*, *11*, 151-161. <http://dx.doi.org/10.3758/BF03199643>

- Donahoe, J. W., & Burgos, J. E. (1999). Timing without a timer. *Journal of the Experimental Analysis of Behavior*, *71*, 257-263. <http://dx.doi.org/10.1901/jeab.1999.71-257>.
- Donahoe, J. W., & Burgos, J. E. (2000). Behavior analysis and reevaluation. *Journal of the Experimental Analysis of Behavior*, *74*, 331-346. <http://dx.doi.org/10.1901/jeab.2000.74-331>.
- Donahoe, J. W., Burgos, J. E., & Palmer, D. C. (1993). A selectionist approach to reinforcement. *Journal of the Experimental Analysis of Behavior*, *60*, 17-40. <http://dx.doi.org/10.1901/jeab.1993.60-17>
- Donahoe, J. W., Crowley, M. A., Millard, W. J., & Stickney, K. A. (1982). A unified principle of reinforcement: Some implications for matching. In M. L. Commons, R. J. Herrnstein, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 2. Matching and maximizing accounts* (pp. 493-521). Cambridge, MA: Ballinger.
- Dragoi, V. (1997). A dynamic theory of acquisition and extinction in operant learning. *Neural Networks*, *10*, 201-229. [http://dx.doi.org/10.1016/S0893-6080\(96\)00067-6](http://dx.doi.org/10.1016/S0893-6080(96)00067-6)
- Dragoi, V., & Staddon, J. E. R. (1999). The dynamics of operant conditioning. *Psychological Review*, *106*, 20-61. <http://dx.doi.org/10.1037/0033-295X.106.1.20>
- Everitt, B. J., Fray, P., Kostarczyk, E., Taylor, S., & Stacey, P. (1987). Studies of instrumental behavior with sexual reinforcement in male rats (*Rattus norvegicus*): I. Control by brief visual stimuli paired with a receptive female. *Journal of Comparative Psychology*, *101*, 395-406. <http://dx.doi.org/10.1037/0735-7036.101.4.407>
- Everitt, B. J., & Robbins, T. W. (2005). Natural systems of reinforcement for drug addiction: From actions to habits to compulsions. *Nature Neuroscience*, *8*, 1481-1489. <http://dx.doi.org/10.1038/nn1579>
- Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.
- Gibbon, J., & Balsam, P. D. (1981). Spreading association in time. In Locurto, C., Terrace, H. S., & Gibbon, J. (Eds.), *Autoshaping and conditioning theory* (pp. 219-253), New York, NY: Academic Press.
- Gilbertson, D. W. (1975). Courtship as a reinforcement for key pecking in the pigeon, *Columbia livia*. *Animal Behaviour*, *23*, 735-744. [http://dx.doi.org/10.1016/0003-3472\(75\)90101-3](http://dx.doi.org/10.1016/0003-3472(75)90101-3)
- Grossberg, S. (1974). Classical and instrumental learning by neural networks. *Progress in Theoretical Biology*, *3*, 51-141.
- Holland, P. C. (1990). Event representation in Pavlovian conditioning: Image and action. *Cognition*, *37*, 105-131.
- Killeen, P. (1994). Mathematical principles of reinforcement. *Behavioral and Brain Sciences*, *17*, 105-172. <http://dx.doi.org/10.1017/S0140525X00033628>
- Killeen, P. R. (2011). Models of trace decay, eligibility for reinforcement, and delay of reinforcement gradients, from exponential to hyperboloid. *Behavioural Processes*, *87*, 57-63. <http://dx.doi.org/10.1016/j.beproc.2010.12.016>
- Machado, A. (1997). Learning the temporal dynamics of behavior. *Psychological Review*, *104*, 241-265. <http://dx.doi.org/10.1037/0033-295X.104.2.241>
- Mesulam, M.-M. (1998). From sensation to cognition. *Brain*, *121*, 1013-1052. <http://dx.doi.org/10.1093/brain/121.6.1013>
- Michael, R. P., & Keeverne, E. B. (1968). Pheromones in the communication of sexual status in primates. *Nature*, *218*, 746-749. <http://dx.doi.org/10.1038/218746a0>
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.), *Classical Conditioning II: Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.
- Sánchez, J. M., Galeazzi, J. M., & Burgos, J. E. (2010). Some structural determinants of Pavlovian conditioning in artificial neural networks. *Behavioural Processes*, *84*, 526-535. <http://dx.doi.org/10.1016/j.beproc.2010.01.018>
- Schmajuk, N. A., Lam, Y. -W., & Gray, J. A. (1996). Latent inhibition: A neural network approach. *Journal of Experimental Psychology: Animal Behavior Processes*, *22*, 321-349. <http://dx.doi.org/10.1037/0097-7403.22.3.321>

- Schultz, W. (2010). Dopamine signals for reward value and risk: Basic and recent data. *Behavioral and Brain Functions*, 6, 1–9. <http://dx.doi.org/10.1186/1744-9081-6-24>
- Seligman, M. E. P. (1970). On the generality of the laws of learning. *Psychological Review*, 78, 459–486. <http://dx.doi.org/10.1037/h0029790>
- Sevenster, P. (1973). Incompatibility of response and reward. In R. A. Hinde, & J. Stevenson-Hinde (Eds.), *Constraints on Learning: Limitations and Predispositions* (pp. 265–283). London, UK: Academic Press.
- Staddon, J. E. R., & Zhang, Y. (1991). On the assignment-of-credit problem in operant learning. In M. L. Commons, S. Grossberg, & J. E. R. Staddon (Eds.), *Neural network models of conditioning and action. Quantitative analyses of behavior series* (pp. 279–293). Hillsdale, NJ: Erlbaum.
- Stout, S. C., & Miller, R. R. (2007). Sometimes competing retrieval (SOCR): A formalization of the comparator hypothesis. *Psychological Review*, 114, 759–783. <http://dx.doi.org/10.1037/0033-295X.114.3.759>.
- Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, 88, 135–170. <http://dx.doi.org/10.1037/0033-295X.88.2.135>.
- Timberlake, W., Wahl, G., & King, D. (1982). Stimulus and response contingencies in the misbehavior of rats. *Journal of Experimental Psychology: Animal Behavior Processes*, 8, 62–85. <http://dx.doi.org/10.1037/0097-7403.8.1.62>
- Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear and R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5–47). Hillsdale, NJ: Erlbaum.
- Wagner, A. R. (2008). Evolution of an elementary theory of Pavlovian conditioning. *Learning & Behavior*, 36, 253–265. <http://dx.doi.org/10.3758/LB.36.3.253>

## Appendix

Computationally, the model consists of an activation rule (or function) and a learning rule. The activation rule is used to compute the level of activation of every neurocomputational unit (circles in Figure 1) at every moment of every trial. The learning rule is used to change weights of variable connections (thin lines ending with buttons in Figure 1). All activations and weights are numbers between 0.0 and 1.0. For both rules, time is conceived as consisting of discrete moments, occasions, epochs, or timesteps of indefinite but relatively short duration (no particular duration is used for computational purposes). Each rule is described in turn. Their neuroscientific rationale has been described in detail in earlier papers, so it will only be briefly described here.

### Activation Rule

This rule (see Eq. 1) has two modes: unconditional (automatic, “innate”) and conditional (“acquired,” “learned”). Unconditional activation does not require any learning (weight change) and obtains when the activation of a certain input unit ( $S^*$  units in Figure 1) at  $t$  is larger than 0.0, and the unit whose activation is being computed ( $j$ ) is a  $D$  (dopaminergic unit) or  $R^*$  (Pavlovian output unit;  $M'_{R^*}$  in Figure 1). Otherwise, conditional activation obtains. Conditional activation requires learning (weight change, usually gain) and has two modes: reactivation and decay. These modes relate mathematically according to Eq. 1:

$$a_{j,t} = \begin{cases} a_{S^*,t}, & \text{if } a_{S^*,t} > 0 \text{ and } j \text{ is } D \text{ or } R^* \text{ (unconditional activation); otherwise,} \\ L(exc_{j,t}) + \tau_j L(exc_{j,t-1}) [1 - L(exc_{j,t})] - L(inh_{j,t}), & \\ \quad \text{if } L(exc_{j,t}) > L(inh_{j,t}) \text{ and } L(exc_{j,t}) \geq \theta_t \text{ (reactivation)} \\ a_{j,t-1} - \kappa_j a_{j,t-1} (1 - a_{j,t-1}) - L(inh_{j,t}), & \quad 22 \\ \quad \text{if } L(exc_{j,t}) > L(inh_{j,t}) \text{ and } L(exc_{j,t}) < \theta_{j,t} \text{ (decay)} \\ 0, & \text{if } L(exc_{j,t}) \leq L(inh_{j,t}) \text{ (deactivation)} \end{cases}$$

(Eq. 1),

where  $t$  is a moment in time and

$$L(x) = \frac{1}{1 + e^{-\frac{x+\mu}{\sigma}}}$$

is the logistic function with constant mean  $\mu = 0.5$  and standard deviation  $\sigma$  (a spontaneous activation free parameter). In this function,

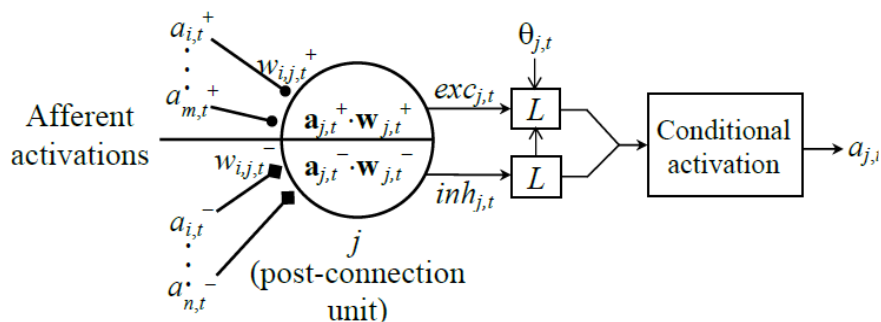
$$x = \sum_{i=1}^m a_{i,t}^+ w_{i,j,t}^+$$

for  $exc_{j,t}$ , and

$$x = \sum_{i=1}^n a_{i,t}^- w_{i,j,t}^-$$

for  $inh_{j,t}$ , where  $m$  denotes the total number of excitatory units connected to  $j$  and  $n$  the total number of inhibitory units connected to  $j$ . In INT in Figure 1, there was only one inhibitory unit (open diamond labeled as  $I$ ) connected to only one output unit ( $M'_R$ ).  $\sigma = 0.1$  for all units.

Figure 3 shows a generic neurocomputational unit (zoom-in on any circle in Figure 1) for conditional activation, intended to simulate a relatively small neuronal group. The unit receives a finite number of afferent excitatory ( $a_{1,t}^+, \dots, a_{i,t}^+, \dots, a_{m,t}^+$ ) and/or inhibitory activations ( $a_{1,t}^-, \dots, a_{i,t}^-, \dots, a_{n,t}^-$ ) from one or more (up to  $m$  or  $n$ , respectively) excitatory (+) or inhibitory (-) afferents.



**Figure 3.** A generic neurocomputational unit, labeled as  $j$ . Unit  $j$  receives connections from afferent units that can be excitatory (+) or inhibitory (-), and influence  $j$  separately at  $t$ .  $a_{i,t}^+$ : Activation of excitatory afferent  $i$  at  $t$ .  $a_{i,t}^-$ : Activation of inhibitory afferent  $i$  at  $t$ .  $w_{i,j,t}^+$ : Weight of a connection from excitatory afferent  $i$  to  $j$  at  $t$ .  $w_{i,j,t}^-$ : Weight of a connection from inhibitory afferent  $i$  to  $j$  at  $t$ .  $\mathbf{a}$ : Afferent activation vector.  $\mathbf{w}$ : Weight vector.  $\mathbf{a} \cdot \mathbf{w}$ : Dot product of activation and weight vectors.  $exc_{j,t}$ : amount of excitation on  $j$  at  $t$ .  $inh_{j,t}$ : amount of inhibition on  $j$  at  $t$ .  $L$ : Logistic function.  $\theta_{j,t}$ : Gaussian threshold on  $L(exc_{j,t})$  ( $L(inh_{j,t})$  is another threshold).  $a_{j,t}$ : Activation of  $j$  at  $t$ . All activations and weights are numbers between 0.0 and 1.0.



Each afferent unit  $i$  sends to  $j$  (the post-connection unit) a connection with a variable excitatory (e.g.,  $w_{i,j,t}^+$ ) or inhibitory (e.g.,  $w_{i,j,t}^-$ ) weight. The post-connection unit  $j$  computes a product of activation and connection vectors  $\mathbf{a}_{j,t}$  and  $\mathbf{w}_{j,t}$ , respectively, separately for excitatory and inhibitory activations and weights. This product is the net amount of excitation ( $exc_{j,t}$ ) or inhibition ( $inh_{j,t}$ ) on  $j$  and  $t$ , passed as an argument  $x$  to a logistic function ( $L$ ).

Whether the rule is in reactivation or decay mode at  $t$  depends on a Gaussian threshold ( $\theta_{j,t}$ ), a random number generated according to a Gaussian distribution with a mean of 0.2 and standard deviation of 0.15.  $\theta_{j,t}$  is dynamical, as it is generated at every moment for every computational unit. Two other activation free parameters are temporal summation ( $\tau_j$ ) and decay ( $\kappa_j$ ). For all units,  $\tau_j = 0.1$  and  $\kappa_j = 0.1$ .

## Learning Rule

This rule is used to change the weight  $w$  of every connection from unit  $i$  (afferent) to unit  $j$  (target) at  $t$  ( $w_{i,j,t}$ ) at every moment, separately for excitatory and inhibitory afferent units. A connection is intended to simulate a relatively small synaptic group, and a weight to simulate the efficacy of that group to mediate the activation of a post- by an afferent neuronal group. A weight can be interpreted as the proportion of transmitter receptors on  $j$  that are controlled by  $i$ . The rule to change weights is as follows:

$$\Delta w_{i,j,t} = \begin{cases} \alpha_j a_{j,t} d_t p_{i,t} r_{j,t}, & \text{if } d_t \geq 0.05 \\ -\beta_j w_{i,j,t-1} a_{i,t} a_{j,t}, & \text{otherwise} \end{cases} \quad (\text{Eq. 2}),$$

where  $\alpha$  (rate of weight increment) and  $\beta$  (the rate of weight decrement) denote the two free parameters of the rule ( $\alpha = 0.5$  and  $\beta = 0.1$  for all connections). The other terms of the rule are:

$a_{i,t}$ : activation of afferent unit ( $i$ ), either excitatory or inhibitory;

$a_{j,t}$ : activation of target unit ( $j$ );

$d_t = d_{H,t} = |a_{H,t} - a_{H,t-1}| + d_{D,t}(1 - d_{H,t-1})$ , if  $j$  is an  $S''$  or  $H$  unit;

$d_t = d_{D,t} = a_{D,t} - a_{D,t-1}$ , if  $j$  is an  $M''$ ,  $D$ , or  $M'$  unit (see Fig. 1);

$$p_{i,t} = \frac{a_{i,t} w_{i,j,t-1}}{N}, \text{ where } N = exc_{j,t} \text{ or } N = inh_{j,t}$$

depending on whether  $i$  is excitatory or inhibitory, respectively;

$$r_{j,t} = 1 - \sum_{i=1}^n w_{i,j,t}$$

The key factor is  $d_t$ , a signal that modulates changes of all weights, inspired by the roles of hippocampal (e.g., CA1) and dopaminergic (e.g., ventral-tegmental) areas in conditioning. In this sense,  $d_t$  is a diffuse signal. It also is a discrepancy signal in that it is defined as a temporal difference between the activations of certain units ( $H$ ,  $D$ ; see Fig. 1) in successive pairs of moments. In early simulations, the  $d_t$  threshold was 0, but was increased to 0.001 to simulate latent inhibition (see Burgos, 2003). After this, it was further increased to 0.05 to simulate other phenomena.

The  $p_{i,t}$  and  $r_{j,t}$  factors introduce a “rich get richer, poor get poorer” sort of competition among connections for a limited amount of weight on a common target unit. In the network architectures used for the simulation (e.g., Figure 1), this competition took place on units that received two connections, as was the case for all units, except for the  $H$  units, which received only one connection. The  $p_{i,t}$  factor, like some other models, includes a Hebbian component where connection weights partly depend on the activations of the connected (afferent and target) units.

In general, connections tend to gain weight (to a greater or lesser degree, depending on how much weight they have gained) when  $S^*$  (see Figure 1) is activated, and lose weight when  $S^*$  is not activated. Successive timesteps with a zero  $S^*$  activation thus promote weight loss. The same learning rule is used to modify connection weights across all times, connections, networks, units (whether they simulate emitted or elicited responding), and training protocols (whether they simulate operant or Pavlovian contingencies).

All activations and weights are updated at each moment  $t$  according to an asynchronous random procedure. In this procedure, a randomly-ordered list of all units (or connections) is generated at  $t$ , and new activations (or weights) are computed in that order (according to Equation 1 for activations, or Equation 2 for weights). The activations (or weights) from  $t-1$  are immediately replaced by the new activations at  $t$ . Hence, by chance, the activation of a unit at  $t$  could depend on activations at  $t-1$ .

**Financial conflict of interest:** No stated conflicts.  
**Conflict of interest:** No stated conflicts.

**Submitted:** December 3<sup>rd</sup>, 2014  
**Resubmitted:** February 17<sup>th</sup>, 2015  
**Accepted:** April 5<sup>th</sup>, 2015