REVISTA MEXICANA DE ANÁLISIS DE LA CONDUCTA
MEXICAN JOURNAL OF BEHAVIOR ANALYSIS
26, 159-188
NÚMERO 2 (SEPTIEMBRE)
NUMBER 2 (SEPTEMBER)

SUPERSTITION IN ARTIFICIAL NEURAL NETWORKS: A CASE STUDY FOR SELECTIONIST APPROACHES TO REINFORCEMENT

SUPERSTICIÓN EN REDES NEURALES ARTIFICIALES: UN ESTUDIO DE CASO PARA APROXIMACIONES SELECCIONISTAS AL REFORZAMIENTO

JOSÉ E. BURGOS[1]
UNIVERSITY OF GUADALAJARA

ABSTRACT

The superstition phenomenon remains a crossroad of conceptual issues, especially regarding the operant-respondent distinction and the role of neural principles in understanding of behavior. In the present paper, I examine the phenomenon from the perspective of artificial neural networks, in the context of a selectionist approach to reinforcement. I define the basic phenomenon as a persisting change in a behavior that is not a conditional part of the reinforcement operation. Two computer simulations of this phenomenon were run using two feedforward and fully-connected selection networks. Superstition was obtained in both networks through the same reinforcement mechanism used to obtain Pavlovian and operant conditioning in previous simulations. Results showed that response-dependent reinforcement was not necessary to change any emitted behavior, and that superstition was maximally generalized over the networks' repertoire. A more specific form of superstition was obtained in a third simulation by using a partially-connected network. A similar result might be obtained by making different responses mutually exclusive through inhibitory connections. Also, it is likely that a form of shaping through response-dependent reinforcement will be required in order to simulate more complex environment-behavior relations in selection networks. I conclude by examining certain criticisms that have been raised towards neural-network modeling in behavior analysis and the incorporation of neural principles in our accounts of behavior.

---

[1] Centro de Estudios e Investigaciones en Comportamiento, 12 de Diciembre 204, Col. Chapalita, Zapopan, Jalisco 45030, Mexico. E-mail: jburgos@cucba.udg.mx. The author thanks John W. Donahoe and François Tonneau for useful comments to a previous draft of this manuscript.

Keywords: superstition, selection neural networks, Pavlovian conditioning, operant conditioning, neural principles, computer simulation

RESUMEN

El fenómeno de la superstición permanece como una encrucijada de problemas conceptuales, especialmente respecto a la distinción operante-respondiente y al papel de los principios neurales en el entendimiento de la conducta. En el presente trabajo se examina el fenómeno desde la perspectiva de las redes neurales artificiales, en el contexto de una aproximación seleccionista al reforzamiento. Se define el fenómeno básico como un cambio persistente en una conducta que no forma parte condicional de la operación de reforzamiento. Se corrieron dos simulaciones digitales de este fenómeno, utilizando dos redes de selección completamente connectadas de forma anterógrada. La superstición fue obtenida en ambas redes mediante el mismo mecanismo de reforzamiento utilizado para obtener condicionamiento Pavloviano y operante en simulaciones anteriores. Los resultados mostraron que el reforzamiento dependiente de la respuesta no fue necesario para cambiar conducta emitida alguna y que la superstición fue máximamente generalizada a lo largo del repertorio de las redes. Una forma más específica de superstición fue obtenida en una tercera simulación, utilizando una red parcialmente conectada. Un resultado similar podría ser obtenido haciendo que distintas respuestas sean mutuamente excluyentes mediante conexiones inhibitorias. También es posible que una forma de moldeamiento mediante reforzamiento dependiente de la respuesta sea necesaria para simular relaciones ambiente-conducta más complejas en redes neurales de selección. Se concluye examinando ciertas críticas que han sido dirigidas hacia el uso de redes neurales en el análisis conductual y la incorporación de principios neurales en las explicaciones de la conducta.

Palabras clave: superstición, redes neurales de selección, condicionamiento pavloviano, condicionamiento operante, principios neurales, simulación por computadora

---

The phenomenon of superstition remains a crossroad of conceptual issues, especially regarding the operant-respondent distinction and the incorporation of biological principles into our accounts of behavior. In the present paper, I examine the phenomenon from the perspective artificial neural networks, in the context of a selecionist approach to reinforcement. My aim is twofold. First, I want to provide a selectionist, biobehavioral interpretation of superstition, as it occurs in simulations with artificial neural networks that function according to the model proposed by Donahoe, Burgos, and Palmer (1993), and further characterized by Donahoe and Palmer (1994), and Donahoe, Palmer, and Burgos (1997). Second, I want to explore certain challenges that this result poses for said approach. I should clarify that the conclusions reached in our previous simulation research are unaffected by the present

results, and that some of the issues raised here are orthogonal to those conclusions.

In the first part of the paper, I review the behavior-analytic literature on superstition and discuss the standard interpretation of this phenomenon. In the second section, I briefly describe the model that underlies the simulations. In the third section, I describe the basic simulation result, and some of the challenges it poses. In the fourth section, I propose ways of meeting these challenges. I conclude the paper by discussing some of the criticisms that have been made towards incorporating neural principles into our accounts of behavior and towards neural-network modeling in behavior analysis.

Superstition in Behavior Analysis

Skinner (1948) was the first to report the basic phenomenon and call it "superstition". Pigeons were given a fixed-time (FT) 15-s schedule (i.e., free, unconditional access to a feeder at 15-second intervals) for an extended period of time. Although the animals did not end up pecking the key, their behavior had clearly changed, suggesting that the arrangement had had a strong effect on them. This occurred in spite of the fact that food was effectively independent of such patterns, and that they involved the emission of skeletal, striate-muscle responses. Staddon and Simmelhag (1971) replicated and extended Skinner's experiment, making a more systematic and detailed observation of the different kinds of activities displayed by the pigeons. They observed that behavior was not as idiosyncratic as Skinner had originally supposed, although it still was highly stereotyped. A similar result was observed in rats, using a FT 30-s and an hexagonal chamber where subjects could perform a variety of activities, such as drinking and running in a wheel (Staddon & Ayres, 1975).

After Skinner, other phenomena have been observed and called "superstition". For example, Wilson and Keller (1953) observed a similar effect in rats, in the form of an increase in collateral behaviors between barpressing responses under a DRL schedule. These behaviors consisted of species-specific responses such as grooming and nose-poking. Also, Morse and Skinner (1957) observed pigeons responding differentially (some with high rates, others with low rates) to a light whose occurrence was independent of the animal's behavior, in an arrangement where food depended on keypecking. Similarly, Herrnstein and Morse (reported by Herrnstein, 1966) trained pigeons to keypeck for food in a fixed-interval (FI) 11-s schedule of response-dependent reinforcement. Then, pigeons were given a FT 11-s schedule. Keypecking showed substantial maintenance under the response-independent procedure. Although the rates under FI were higher than the rates under FT, the latter were

still substantial. In a different kind of arrangement, Boren and Devine (1968) studied response chaining in monkeys. They observed the regular occurrence of components that were not required for reinforcement, a phenomenon that has been referred to as "superstitious chaining". Also, Williams and Williams (1969) observed substantial response maintenance under an arrangement where food presentations depended on the keylight, but not on keypecking, while the omission of food depended on keypecking. This phenomenon was subsequently called "negative automaintenance".

On the basis of the above phenomena, superstition can be generally defined as a persisting change in a behavior that is not a conditional component of the reinforcement operation. That is, a response pattern is considered as superstitious if it occurs regularly in arrangements where emitting it is not required for reinforcement. Whether or not the pattern is stereotyped or idiosyncratic does not define the phenomenon, for nonsuperstitious behavior can also be stereotyped and idiosyncratic. Alternatively, a response pattern is considered as superstitious if it deviates from a formulation of the Law of Effect in terms of effective dependence of reinforcers on behavior.

The last characterization makes superstition a puzzling, anomalous phenomenon. Skinner attempted to resolve the anomaly by clarifying that the term "contingency" refers only to a relationship of temporal contiguity between a response and a reinforcer. In this way, the term lost all reference to effective, causal, conditional ('if..., then...') dependence of reinforcers on responses. At best, conditionality was given a methodological role, as a way of making the temporal contiguity between a particular response and a reinforcer as close as possible. The Law of Effect that results from this conceptual restriction states that if instances of a certain response class (e.g., keypecking) are followed closely in time by instances of a certain stimulus class (e.g., food), then the probability of occurrence of future instances of the former class will increase. Whether or not reinforcers depend effectively on the responses becomes theoretically irrelevant.

This formulation yielded to the standard behavior-analytic interpretation, according to which superstition is due to adventitious reinforcement. That is, a chance or accidental occurrence of the reinforcer immediately after whatever response the organism was emitting at that moment, increased the probability of future occurrences of similar responses. Such an increase, in turn, increased the likelihood of future chance pairings of the reinforcer with new occurrences of the same response class, and so on. Adventitious reinforcement thus increases response probability in the same manner response-dependent reinforcement does. The only difference is operational, in that response-dependent reinforcement allows for a better experimental control of the temporal relationship between responses and reinforcers. But behavior change

in both procedures is explained by the occurrence of reinforcers in a close temporal contiguity relationship with responses.

The above interpretation has been challenged on a number of grounds (see Staddon, 1977). First, it is inconsistent with certain phenomena, such as negative automaintenance and the suppression of instrumental responding maintained by response-dependent reinforcement, by occasional response-independent food deliveries. One could attempt to explain the latter phenomenon in terms of the adventitious reinforcement of responses that interfere with the one initially maintained by the response-dependent contingency. However, this latter response is presumably more probable than others, at the moment of introducing the response-independent contingency. Hence, the responses that were initially maintained by the response-dependent contingency should be more frequently followed by the response-independent reinforcer than others, which should result in their substantial maintenance. In order to explain the suppression phenomenon in terms of adventitious reinforcement, the contiguity formulation of the Law of Effect would have to be complemented the assumption that behavior is more strongly affected by reinforcement during acquisition than during maintenance. This move would also be required to explain a similar phenomenon, that is, the eventual replacement of behaviors that are more probable early in a response-independent arrangement (e.g., putting the head into the magazine) by other, initially less probable behaviors (e.g., pecking).

A second challenge to the standard interpretation of superstition is that adventitious reinforcement remains mostly an interpretative, operationally elusive concept. Therefore, it is extremely difficult to demonstrate experimentally that a certain behavioral pattern is due to chance response-reinforcer pairings. Indeed, the term "reinforcement" refers to a kind of relationship between environment and behavior. An adequate operational definition thus requires the specification of a unit of responding and a unit of reinforcement that preserve their identities across time and serve as criteria for determining whether or not reinforcement has actually occurred. If we want the concept of reinforcement to have any predictive value, then such a specification must be done a priori (i.e., before the organism is exposed to the contingencies). For example, we can predict (whether correctly or not) that the probability of occurrence of keypecking will increase after instances of this class are immediately followed by instances of food. This kind of prediction is possible if and only if we specify a priori a consequent-stimulus class, a response class, and a contingency relation. Such a specification inevitably leads to a procedure in which instances of the consequent-stimulus class will depend effectively on instances of the response class.

In contrast, the notion of adventitious reinforcement allows for an a priori specification of the stimulus class, leaving the response class and the contingency relation unspecified. In this sense, adventitious reinforcement is a postdictive, more than a predictive concept. It thus suffers from a deep logical limitation. At best, it allows for very indirect and general predictions (e.g., "any response class whose instances are immediately followed by food will become more probable"). A successful prediction of this kind, of course, does not guarantee that any observed probability changes are due to adventitious reinforcement, unless a procedure is implemented for recording chance response-reinforcer pairings. Such a procedure, however, can prove rather intricate and, hence, not susceptible of adequate experimental control. Another difficulty arises from the fact that the above kind of prediction implicitly assumes that any behavior is modifiable by reinforcement (adventitious or otherwise). This is the so-called "transituationality" assumption (Meehl, 1950), which has been challenged by the idea of biological constraints on learning (e.g., Bolles, 1970; García, McGowan, & Green, 1972; Hinde & Stevenson-Hinde, 1973; Seligman, 1970; Shettleworth, 1972). The evidence indicates that a substantial portion of superstitious behavior consists of (or, at least, relates in meaningful ways to) species-specific (unconditional, biological) responses to the reinforcer (Staddon & Simmelhag, 1971). This leads to the possibility that superstition may arise from a complex mixture of Pavlovian and operant contingencies, one that may be impossible to disentangle experimentally and theoretically (but see Skinner, 1935). The issue of whether or not superstition can be sufficiently explained without appealing to nonbehavioral events and processes is thus raised. We have raised this issue in relation to phenomena such as reinforcement revaluation (Donahoe & Burgos, in press). In fact, it can be equally raised in relation to any behavioral phenomenon.

Appealing to nonbehavioral events and processes has become an anathema in behavior analysis and radical behaviorism. A substantial portion of radical behaviorists' efforts within psychology has been to condemn as meaningless, useless, and misleading any kind of account that appeals to nonbehavioral events and processes. We must not forget, however, that this effort has been directed primarily towards rejecting explanations that appeal to *inferred* events and processes, of the kinds found in cognitivist psychology (a substantial portion of Pavlovian-conditioning research with animals included). In this particular respect, the present approach joins the radical-behaviorist effort. However, in our zeal to jettison inferred-process psychology, we have thrown out the baby together with the bathtub, ignoring the possibility that nonbehavioral events and processes can be admitted without violating our philosophical commitment against cognitivist explanations.

I am referring, of course, to events and processes that occur in the organism's nervous system. The possibility, legitimacy, and even need of incorporating such events and processes into our accounts of behavior was acknowledged by Skinner (1974):

> ... we shall eventually know much about the kinds of physiological processes, chemical or electrical, which take place when a person behaves. The physiologist of the future will tell us all that can be known about what is happening inside the behaving organism. His account will be an important advance over a behavioral analysis, because the latter is necessarily 'historical' --that is to say, it is confined to functional relations showing temporal gaps. Something is done today which affects the behavior of an organism tomorrow. No matter how clearly that fact can be established, a step is missing, and we must wait for the physiologist to supply it. He will be able to show how an organism is changed when exposed to contingencies of reinforcement and why the changed organism then behaves in a different way, possibly at a much later date. What he discovers cannot invalidate the laws of a science of behavior, but it will make the picture of human action more nearly complete (p. 215).

The above paragraph is meaningful in that it was written by someone whose work has lead to a *theoretical* distinction between operant and respondent conditioning as involving *different underlying reinforcement mechanisms*. Skinner himself changed his view on the matter throughout his life. He did reject a fundamental distinction early in his work, favoring a strictly operational distinction (see Skinner, 1935, 1937, vs. Konorski & Miller, 1937a, 1937b). However, his speculations on the evolution of behavior (Skinner, 1969, Ch. 7; 1974, Ch. 3; 1975, 1981) imply that different phylogenetic histories may have given rise to separate neural mechanisms for operant and respondent conditioning.

In any case, it is clear that a *phenomenalistic* distinction can be made. However, phenomenalistic distinctions do not necessarily imply theoretical distinctions. Newtonian mechanics makes a phenomenalistic distinction between body fall, Earth's movement around the sun, and the tides, but accounts for them theoretically in terms of a single set of principles (viz., the famous Three Laws). In the above paragraph, Skinner thus advises us to wait for the relevant physiological data, instead of speculating about underlying mechanisms (his own speculations notwithstanding), if by 'underlying mechanism' we mean 'a process that occurs inside the organism'. In this sense, talk about underlying mechanisms for Skinner amounted to talk about neural mechanisms, but none of this talk was the behavior analyst's job. A theoretical distinction between Pavlovian and operant conditioning thus would

ultimately refer to a distinction between different underlying neural mechanisms at work in each kind of phenomenon. If a single reinforcement mechanism underlies both kinds of phenomena, however, then no theoretical distinction would be justified. The present approach sustains that this is a real possibility, based on some of the relevant evidence from the neurosciences (for a review of this evidence, see Donahoe & Palmer, 1994). Pavlovian and operant conditioning may just be different phenomena, in the same sense that body fall, Earth's movement around the sun, and the tides are different phenomena in Newtonian mechanics.

But Skinner's paragraph gives us further advice. On the one hand, the aim of a science of behavior is to get as complete an account of behavior as possible. This search for completeness, of course, must be carried within the limits imposed by the inherently simplifying nature of scientific theories. To achieve such an account, we can appeal to any principles at our disposal that seem pertinent to the phenomenon of interest, as long as they have been derived through experimental analysis, and regardless of whether they are behavioral or neural. On the other hand, it is a truism that organisms have nervous and endocrine systems, and changes in these systems constitute an integral component of learning and behavior. To be sure, this truism does not force us, as a matter of logical necessity, to appeal to neural events and processes (never mind use the techniques and methods from neuroscience to obtain ordered behavioral data). Applications of the experimental method to derive behavioral principles are possible only by viewing the behavior of the whole organism as a subject matter in its own right.

However, this consideration does not necessarily mean that we must eschew any reference whatsoever to the organism's nervous system in our accounts of behavior, after we have established our behavioral principles through experimental analysis. Once a behavioral principle is available, it is legitimate (even inevitable) to ask what happens inside the organism (or, more precisely, to its nervous and endocrine systems) in particular realizations of the principle. We do not have to ask this, but not doing it will make our picture of the whole organism less complete. The whole organism is not only its behavior but also its biology. Hence, accounts that combine behavioral and neural principles are more complete than accounts that consist only of behavioral principles. I shall return to this issue in my concluding remarks.

The Model

Neural principles can be incorporated into our accounts of behavior in a number of different ways. In the present approach, a model has been built and used as a basis for simulating certain behavioral phenomena in a digital

computer. The rationale for doing computer simulations in the present approach is similar to the one found in other approaches (e.g., this issue). That is, if the phenomenon of interest is successfully simulated, then the underlying model can be used to interpret the phenomenon in question. This is an example of a "formal interpretation" (Donahoe & Palmer, 1989).

A correspondence between simulations and behavioral phenomena only indicates the behavioral plausibility of a model, which we consider as necessary but not sufficient. In addition to this correspondence, the present approach also emphasizes neural plausibility. This emphasis raises an issue to which I shall in my concluding remarks. Suffice it to say at this point that the search for neural plausibility is a legitimate endeavor (at least as legitimate as rejecting it), even if current models can aspire to capture only very general properties of nervous systems. For the moment, let me describe the present model briefly and informally. For a more detailed description, see Donahoe and Burgos (1999), Donahoe, Burgos, and Palmer (1993), Donahoe and Palmer (1994), and Donahoe, Palmer, and Burgos (1997).

The model consists of two submodels, namely, a neurocomputational model and a network model. A neural network that is designed according to these submodels we call a 'selection neural network'. The term 'selection' refers to the basic, unifying concept of the general approach to reinforcement underlying the model. The use of this term arises from an analogy between learning and evolution by selection (e.g., Donahoe & Palmer, 1994; Skinner, 1981; Staddon & Simmelhag, 1971; cf. Tonneau & Sokolowski, in press). According to this analogy, behavior change in the individual organism, whether Pavlovian or operant, involves the selection of environment-behavior relations through reinforcement. Before reinforcement, an organism is capable of responding in many different ways to many different stimuli. That is, many different stimulus-response relations are possible (although not necessarily equiprobable, due to species-specific biological constraints). However, only some relations become more probable (i.e., become selected) through the action of reinforcement on the organism's nervous system. Exactly which relations are selected depends on an interaction between the contingencies and the organism's biology. The present model intends to capture just a few of the neural mechanisms that may constitute the latter.

The neurocomputational submodel

The neurocomputational submodel is a discrete-time mathematical model consisting of two sets of equations, namely, an activation rule and a learning rule. The activation rule is a function that determines the state of a neural processing element (or NPE, the fundamental structural and functional unit of

a neural network), at a given moment in time. In the present model, that state is a real number between 0 and 1. An activation state can be roughly interpreted as the firing probability of a neuron. The activation rule allows for an NPE to have a spontaneous activation, that is, an activation larger than zero in the absence of input signals.

The learning rule arose as a neural analogue of the unified reinforcement principle proposed by Donahoe, Crowley, Millard, and Stickney (1982), and it is defined as a difference equation that describes changes in connection weights across successive moments in time. A connection weight is a magnitude that represents the strength of a connection between two elements. In the present model, this magnitude is also a real number between 0 and 1. A connection weight can be neurally interpreted as the proportion of postsynaptic neurotransmitter receptors that are controlled by a given presynaptic process.

The rule has three crucial features. First, it includes a signal that modulates the amount of weight change that can occur at a moment in time. This signal is defined as the activation state of a certain kind of NPE (see below) at a given moment in time, minus its activation at the immediately preceding moment. Although the signal may vary from moment to moment, it has the same magnitude for all the weights at any given moment in time. Hence, its diffuse nature. Also, the elements whose activation gives rise to the signal are directly activated by the reinforcer (see below). We thus refer to it as a 'reinforcement signal'. We reserve the term 'discrepancy signal' for an actual larger-than-zero difference in the reinforcement signal.

A second feature is that the rule consists of two mutually exclusive modes, namely, incremental (or acquisition) and decremental (or extinction). The incremental mode is enabled whenever a discrepancy signal occurs. Otherwise, the decremental mode is enabled. If the incremental mode is enabled, any weight change that occurs at that moment is added to the respective current weight. If the decremental mode is enabled, then weight changes are subtracted. Finally, the rule includes a competition factor, meaning that presynaptic elements compete for a fixed amount of weight on the postsynaptic element. The amount of weight change for a given connection to an NPE thus will depend on the total amount of weight controlled by other connections to the same NPE.

The network submodel

The network submodel specifies a classification of the kinds of elements that may constitute a selection neural network, and certain rules about how they are to be connected. As Figure 1 shows, selection networks follow the

standard topological organization of elements into input, hidden, and output. In general, the activation of input elements represents the occurrence of an exteroceptive stimulus, while the activation of output elements represents the occurrence of a response on the network's part. As a simplifying device for our initial simulations, we have not given more specific interpretations to input and output activations. Thus, an input element may represent either a sensory modality in itself (a visual or an auditory sensor) or a sensory channel within a given modality (a visual sensor specialized for red or blue). Similarly, an output element may represent either a corticospinal tract group of neurons in primary motor cortex that controls some particular response topography (e.g., barpressing or keypecking), or a component of such a group.

In the present model, the standard classification is elaborated into more specific kinds of elements. Input elements are classified into primary-sensory, whose activation simulates the occurrence of exteroceptive stimuli (e.g., lights or tones), and reinforcer (or *S**), whose activation simulates the occurrence of a primary reinforcer (e.g., food or water). As a simplifying device, all selection networks are assumed to have only one *S** element. Strictly speaking, input elements are not NPEs, for their states are not computed through the activation rule. Rather, such states are assigned according to some prespecified training protocol (see simulations below). Only hidden and output elements qualify as NPEs, in that their states are computed through the activation rule. However, like NPEs, the activation state of input elements (primary-sensory as well *S**) is represented as a real number between 0 and 1.

Hidden elements are classified into cortical and subcortical, in an effort to capture the gross anatomical organization of the mammalian brain. Cortical elements are classified into sensory-association (*sa*) and motor-association (*ma*), while subcortical elements are classified into *ca1* (for the CA1 region in the hippocampus) and *vta* (for the ventral-tegmental area). Output NPEs are subdivided into operant (or *R*) and respondent (or *CR/UR*) NPEs. The only difference is that *CR/UR* NPEs can be activated by *S**, while *R* NPEs cannot (see below). In this manner, the distinction between operant and respondent responses in the present model arises as an anatomical distinction that leads to a functional one.

The activation of subcortical elements is the source of the reinforcement signals. Specifically, the activation of the *vta* elements is the source of the signal that modulates changes in the weights of connections to *ma*, output, and *vta* NPEs. We call this signal 'd_M'. The activation of the *ca1* elements is the source of the signal that does the same for connections to *sa* and *ca1* NPEs. We call this signal 'd_S', and it is amplified by 'd_M'. For a neural interpretation of these signals, see Donahoe and Palmer (1994).

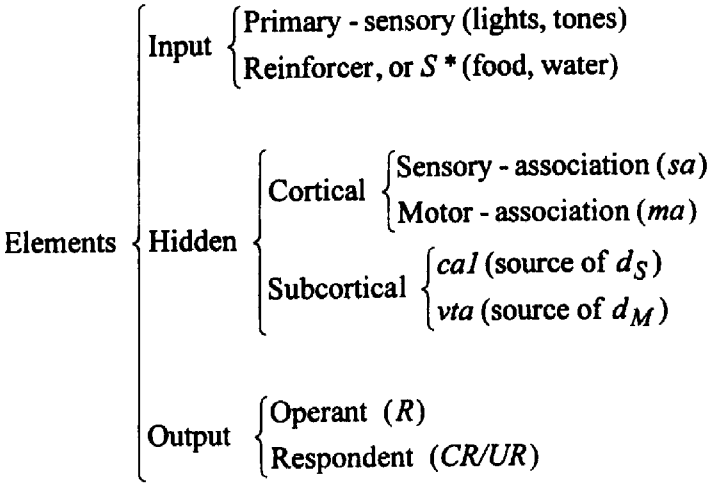

Figure 1. Classification of element types in a selection network (see text for details)

Regarding connectivity, the present model imposes very general and flexible restrictions. More specific restrictions can be used in case a particular architecture is needed. Most networks we have used in our simulation research share a basic feature (also found in other models), namely, feedforward connectivity (see Donahoe & Burgos, in press, for an exception to this feature). In this kind of connectivity, elements are first organized into layers and then connected from one layer to the next. Feedforward connections are one-directional, meaning that a signal propagates from one element to the other, but not vice versa.

One way to understand a layer is to imagine it as a set of circles representing elements and arranged on an imaginary straight line that intersects each circle's centre. Depending on the orientation of that imaginary line, layers can be arranged either vertically (one beside the other) or horizontally (one on top of the other). In our simulation research we have used a vertical arrangement, so I will use it here as well. This kind of arrangement, of course, does not intend to capture the exact topological manner in which neurons are organized in real brains. Rather, the idea is to provide a relatively clear and simple way of representing neural networks visually. In this representation, input elements typically constitute the leftmost layer (the input layer), while output elements constitute the rightmost layer (the output layer), all other elements constituting the layers that are 'hidden' between the input and the output layers.

A minimal selection network consists of one input layer, two hidden layers, and one output layer (see Figure 2). Following the above classification, hidden layers consist of cortical and subcortical layers. We visualize cortical layers as being on top of subcortical layers. A cortical layer may consist of either *sa* or *ma* NPEs, while a subcortical layer may consist of either *ca1* or *vta* NPEs. All NPEs that constitute a layer, then, are assumed to be of the same kind. The first hidden layer is placed immediately to the right of the input layer, while the second hidden layer is placed immediately to the right of the first hidden layer.

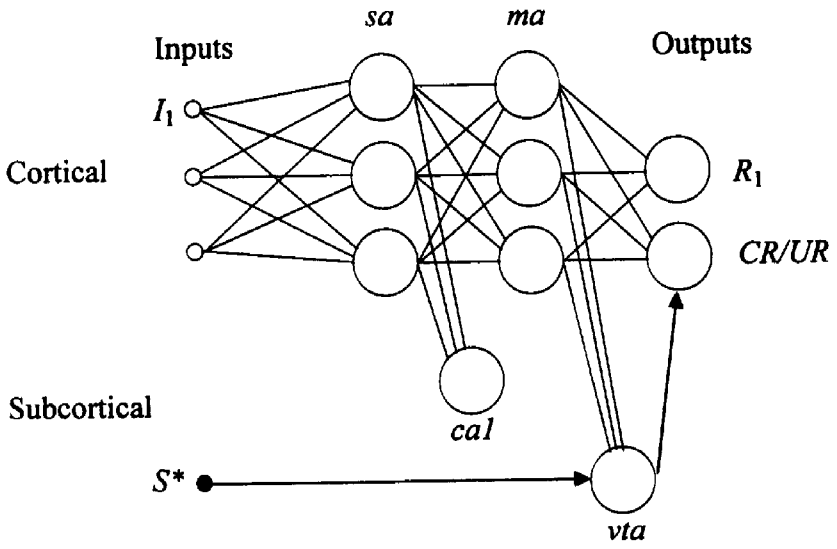

Figure 2. Selection network used in the basic superstition simulation. Small circles represent input elements. Large circles represent neural processing elements (NPEs). The activation of input elements simulated the occurrence of an exteroceptive stimulus. The activation of the output NPEs simulated the network's behavior. The activation of I_1 simulated the occurrence of an exteroceptive stimulus. The activation of *R* NPE simulated an emitted response

In the network shown in Figure 2, from left to right, the input layer consists of three elements. The first cortical hidden layer consists of three *sa* NPEs, while the second cortical hidden layer consists of three *ma* NPEs. The first subcortical layer, from left to right, consists of one *ca1* NPE, while the second consists of one *vta* NPE. Finally, the output layer (the rightmost set of elements) consists of one *R* and one *CR/UR* NPE. Each input element is connected to each *sa* NPE, and each *sa* is connected to the *ca1* and to each *ma* NPE. Also, each *ma* NPE is connected to the *vta* NPE and to each output NPE.

All the connections are plastic and initially weak, except for the two connections that constitute the *S*-vta-CR/UR* path, which are nonplastic and maximally strong. The activation of this path represents the occurrence of an unconditional reflex caused by a primary reinforcer (e.g., food, water).
Note that *CR/UR* can also be activated through the input-*sa-ma* paths. In a typical simulation, the two forms of activating *CR/UR* are mutually exclusive at any moment in time, depending on whether the *S** activation is equal to or larger than zero. If it is equal to zero (i.e., no reinforcer occurs), then *CR/UR* is activated through one or more of the possible input-*sa-ma* paths, simulating the occurrence of a conditioned response (CR). If it is larger than zero (i.e., a reinforcer occurs), then it is given priority in activating *CR/UR*, simulating the occurrence of an unconditioned response (UR) whose magnitude is identical to the activation of *S**. Hence the label '*CR/UR*'. The same strategy applies to the activation of *vta* NPEs.

In contrast, *R* NPEs can be activated only through the input-*sa-ma* paths. Their activation thus simulates a nonelicited, emitted response. In this sense, *R* and *CR/UR* NPEs constitute anatomically and functionally different response systems. A differential observation of these two systems under different contingency types (i.e., response-dependent vs. response-independent, respectively), provides one criterion for distinguishing operant from Pavlovian conditioning in the present approach (Burgos, 1999). Another criterion is the distinction between the two kinds of contingencies. The present approach, however, does not make any fundamental distinction between two *reinforcement mechanisms* underlying each form of conditioning.

Simulations, Interpretations, and Challenges

The basic superstition simulation

The network shown in Figure 2 was used. The initial weights for all plastic connections were set to .01. The activation and learning free parameters were the same as those used in previous simulations. The network was given an *ABA'* sequence of treatments. During Phase *A* (acquisition), the network was given 300 reinforced acquisition trials, according to a forward-delay Pavlovian procedure. A trial was defined as the activation of the input element labeled as I_1 with a magnitude of 1 for 6 time-steps (ts). Reinforcement was defined as the activation of *S** with a magnitude of 1 at ts = 6, regardless of the state of the output element labeled as R_1. Hence, reinforcement did not depend on the network's behavior, in that the activation of R_1 was not a conditional component of the reinforcement relation. During Phase *B* (extinction), the network was given 300 nonreinforced trials (i.e., with

an S^* activation of 0). In Phase A' (reacquisition), reinforcement was reinstated. Intertrial intervals were not explicitly simulated. Rather, they were assumed to be sufficiently long to allow for the activation of all NPEs to decrease to a near-zero level.

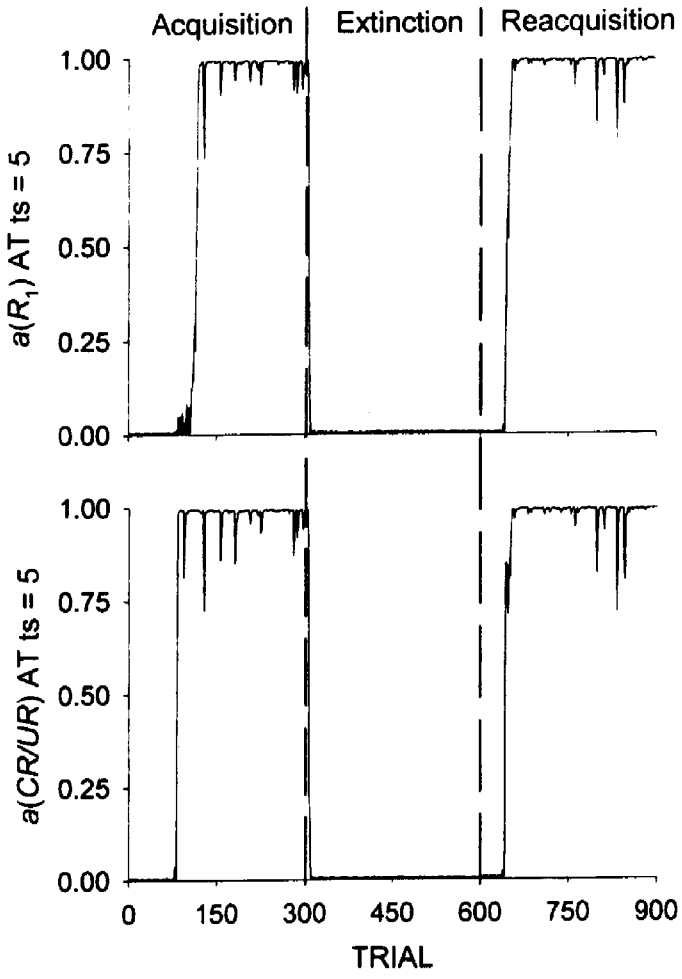

Figure 3. Basic superstition simulation. The R activation (upper panel) increased across trials during acquisition and reacquisition to near-maximal levels, in spite of its being unnecessary for reinforcement. Only activations at the second-to-last time-step (ts = 5) are shown. The *CR/UR* activation (lower panel) increased to near-maximal levels, which simulated Pavlovian conditioning

Figure 3 shows changes in the activation of R_1 (top panel) and *CR/UR* (bottom panel) at ts = 5 (the moment immediately before reinforcement), across trials. At the beginning of the acquisition phase, both NPEs showed near-zero activations. Then, after a number of reinforced trials, the activation of both elements increased to a near-maximal level. Note that the activation of *CR/UR* reached this level a few trials before the activation of R_1. Early in the extinction phase, the activation of both NPEs decreased rapidly to a near-zero level, remaining at this level throughout the entire phase. Then, a number of trials after reinforcement was reinstated (fewer trials than acquisition), both activations once again increased to near-maximal levels.

The increase in the activation of the *CR/UR* NPE simulates Pavlovian conditioning in that both (simulation and phenomenon) involve a change in a stimulus' function due to response-independent reinforcement. During the first acquisition trials, the activation of I_1 did not cause any significant activation of the *CR/UR* NPE, so the former activation effectively was a neutral stimulus for the network. However, after a few reinforced trials, the activation of the same input element eventually came to cause a significant activation of the *CR/UR* NPE. In this sense, the former activation became effectively a 'conditioned stimulus' (CS) for the network. The extinction and reacquisition curves confirm the fact that reinforcement was the critical event for the network to learn to respond during acquisition.

The way Pavlovian conditioning occurred in the present selection network is roughly as follows. At the outset of training, connection weights were too small for the activation of I_1 to cause any significant activation of the *sa*, *ma*, and, hence, the *CR/UR* NPEs. The activation of *S** at the last ts of the first trial produced a large positive difference in the activation of *vta*, between that and the second-to-last ts. This enabled the learning algorithm's incremental mode, thus causing a large positive change in all the relevant weights (i.e., the weights for those connections between coactive elements). As more reinforced trials were presented to the network, weights became increasingly larger, until they allowed the activation of I_1 to propagate throughout the *sa* and *ma* NPEs and eventually cause a significant activation of *CR/UR*.

The increase in the activation of R_1 is the basic simulation superstition, for it occurred in spite of the fact that it was not a conditional component of reinforcement. So the result does satisfy the definition provided in the introduction. The mechanism through which the increase occurred is exactly the same as the one described for *CR/UR*. The only difference was that R_1 did not belong in the *S*-vta-CR/UR* path. Hence, the activation of R_1 was not elicited. However, this does not mean that it could not be under control by an antecedent exteroceptive stimulus (the activation of I_1), for R_1 belonged in the

same path as the activated input element. Reinforcement did not only caused weight increases that eventually allowed I_1 to activate *CR/UR*, but also weight increases that allowed for the same input element to activate R_1. [For explanations of other features of the present results, see Donahoe, Burgos, and Palmer (1993), Donahoe and Palmer (1994), Donahoe, Palmer, and Burgos (1997), and Donahoe and Burgos (in press)].

A selectionist interpretation

The above mechanism provides a basis for an interpretation of what happens to an organism's nervous system when the organism is exposed to response-independent periodic reinforcement. The central notion is that nonelicited behavior may come under the control of antecedent exteroceptive stimulation as much as elicited behavior. This notion could be challenged by arguing that in the procedure described above, reinforcement occurred at the end of discrete trials (i.e., the activation of I_1), according to a Pavlovian procedure. Hence, the argument goes, the procedure is more similar to an autoshaping/automaintenance than to the typical superstition one, for the latter does not involve the presentation of discrete trials. The difference between free-operant and discrete-trial procedures, however, is one of experimental control, rather than one of underlying mechanisms. Consequently, behaviors that are learned and maintained under free-operant arrangements must also be under the control of antecedent exteroceptive stimulation. The main difference is that, in the free-operant case, such stimulation is not under the experimenter's control.

On this basis, superstition can be interpreted in terms of the selection of stimulus-response relations through the same reinforcement mechanism that underlies the selection of stimulus-response relations in Pavlovian and operant procedures. This selection is physically implemented through changes in the appropriate synaptic efficacies in the organism's nervous system. The mechanism in question would be similar to the one described by the learning algorithm. This interpretation makes no fundamental distinction between superstition, Pavlovian, and operant conditioning, at least regarding the underlying reinforcement mechanism and the controlling exteroceptive causal factors. All learning, be it superstitious, operant, or Pavlovian, can be understood in terms of a selection of stimulus-response relations that are biologically implemented in a nervous system.

Challenges

Direct observations of particular response topographies, across different species and with different kinds of reinforcers, have revealed that superstition (like other kinds of schedule-induced behavior, such as adjunctive drinking and adjunctive attack) consist largely of species-specific responses to periodic reinforcement (Schwartz & Gamzu, 1977; Staddon, 1977). Any model that simulates superstition should capture this feature of the phenomenon, which poses a first challenge. In general, the present approach hypothesizes that different realizations of the basic superstition phenomenon arise from differences in particular biological implementations, through different nervous systems with different phylogenetic histories. In order to incorporate this hypothesis into the present model, a more specific semantics of input and output signals is required. This semantics may depend critically on network architecture. We have reported some research on the role of network architecture (e.g., Burgos, 1996, 1997; Donahoe & Burgos, in press). However, said semantics remains undefined. So, for the moment, I will simply acknowledge this challenge and concentrate on two other, more general ones.

The two challenges in question are closely related, and are better introduced through the following simulation. The network shown in Figure 4 was given the same treatment as the previous network. The only difference was that the new network had an additional *R* output NPE (labeled as R_2). The new network, thus, was behaviorally more complex. This could mean that it was capable of emitting either a different kind of behavior or more complex forms of the same kind of behavior (in a moment I shall argue that this is a crucial distinction). The results are depicted in Figure 5, which shows changes in the activation of R_1 (upper panel) and R_2 (lower panel). The activation of *CR/UR* in the present network (not shown) increased in a similar manner as the one in the previous network. Also, like the previous network, the activation of R_1 in the present network increased to a near-maximal level. However, the activation of R_2 also increased to a near-maximal level.

A new network with three *R* NPEs was given the same treatment and their activations also increased to a near-maximal level. If a new network with four *R* NPEs had been trained, their activations would most likely have increased as well, and so on. In general, response-independent reinforcement would seem to induce a change in all the behaviors that a selection network can possibly emit. Superstition in these networks thus becomes maximally generalized over their behavioral repertoires. However, superstition in real organisms is quite specific, in that only some of the organism's possible responses increase in frequency. So a first challenge to the present approach is how to make superstition in selection networks more specific.

A second challenge has to do with the role of response-dependent reinforcement. The present results show that this procedure was not necessary to increase the probability of occurrence of emitted responses in the networks used. This implies that response-independent contingencies are necessary and sufficient, while response-dependent contingencies are unnecessary in selection networks. However, in real organisms, response-independent contingencies seem to be insufficient (and even unnecessary), while response-dependent contingencies seem to be necessary and sufficient for certain kinds of behavior change (e.g., keypecking or barpressing) to occur under certain circumstances (viz., free-operant arrangements). This apparent inconsistency poses the challenge of designing simulations in which response-dependent reinforcement is necessary, without the need to postulate any fundamental separation between Pavlovian and operant conditioning. This challenge, of course, takes us outside the realm of superstition, insofar as it demands a role for response-dependent reinforcement. Nonetheless, it is worth addressing. More challenges can be posed, but I shall concentrate on these two.

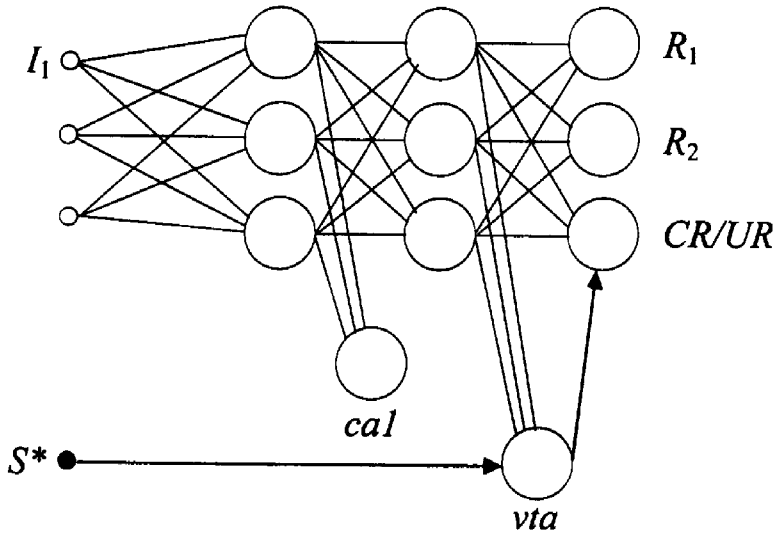

Figure 4. Selection network used in the second superstition simulation. The only difference with respect to the previous network is that the present one has two *R* elements (R_1 and R_2)

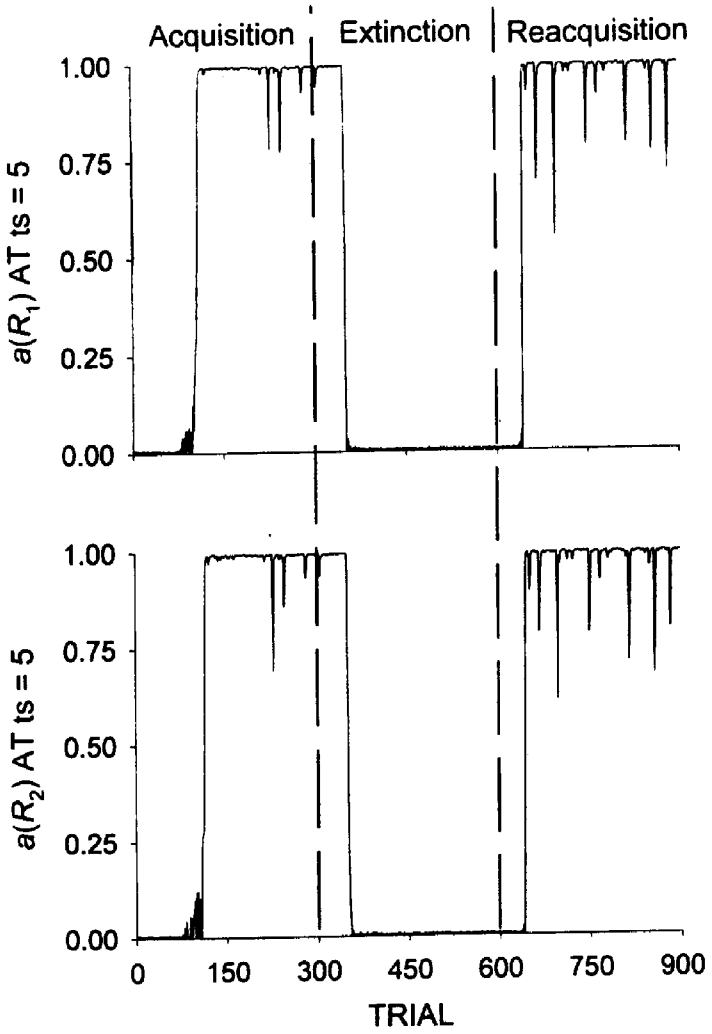

Figure 5. Results of the second superstition simulation, showing how superstition (simulated by an increase in the activation of both *R* elements) becomes maximally generalized over the repertoire of the network depicted in Figure 4

Meeting Challenges

Superstitious behavior is notably constituted by species-specific responses to the reinforcer (Staddon, 1977). Members of the same species

thus tend to display remarkably similar response patterns under periodic response-independent reinforcement, although different patterns are observed across different species. Hence, superstition is not as idiosyncratic as it was initially thought. Yet, superstition tends to be quite specific in that only a few response patterns predominate after extended exposure the procedure. Such specificity must depend critically on the organism's biology, in particular, the characteristics of its nervous system. On this basis, one way of meeting the first challenge is to consider the network's architecture as a determining factor.

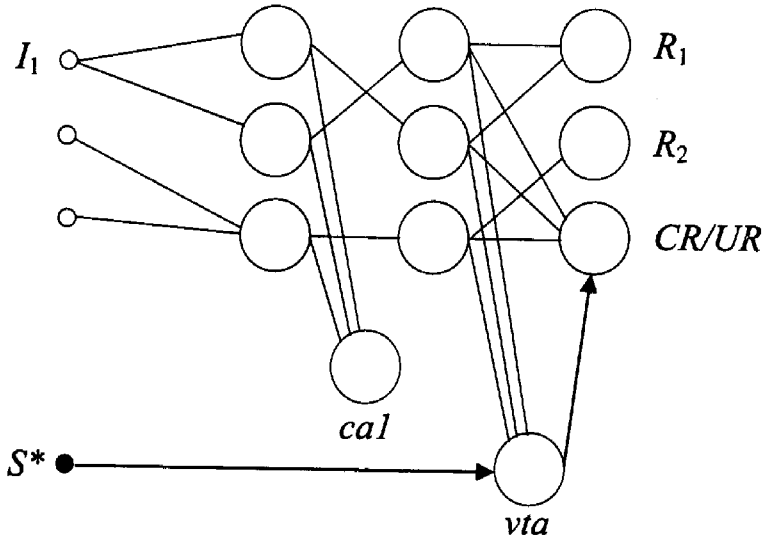

Figure 6. Selection network used in the third superstition simulation. In contrast to the network shown in Figure 4, the connectivity in the present one was such that only one *R* element (R_1) belonged in the path activated by I_1

The present model allows for an unequivocal distinction between elicited and emitted responses. The distinction between different systems of *emitted* responses is less obvious. In principle, however, this latter distinction can also be made in terms of the network's connectivity. On this basis, we can stipulate that different *R* NPEs in a network constitute different emitted-response systems to the extent that they belong in different input-*sa*-*ma* paths. This stipulation is consistent with gross anatomical features of the mammalian brain. According to this criterion, R_1 and R_2 in the second network did not constitute two, but rather one and the same response system. This is because both NPEs were activated through the same input-*sa*-*ma* paths, due to the

network's full connectivity, which explains why both increased their activations in practically the same manner. Such activations thus did not represent two different kinds of behavior, but rather two different aspects or components of the same kind of behavior. The activation increases observed in R_1 and R_2, then, simulated superstitious conditioning of a single response topography.

For R_1 and R_2 to constitute separate response systems, a different network architecture is needed, one whose connectivity allows for an anatomical separation between the two NPEs and, hence, a functional distinction between their activations. Figure 6 shows an example of such a network. The only difference with respect to the previous networks is that the connectivity in the present one is partial, such that only R_1 belongs in the path activated by I_1.

This network was given the same treatment as the other two. As Figure 7 shows, only the activation for R_1 increased substantially (although in a less stable manner), while the one for R_2 remained at a near-zero level. This result demonstrates that selection networks are capable of simulating stimulus-specific superstition, which meets the first challenge.

Stimulus-specific superstition in a selection network could be achieved through other means. For instance, the activation of different *R* NPEs could be made mutually exclusive through inhibitory connections. We still have to explore this possibility empirically, so a brief description will have to do for the moment. The basic idea is that inhibitory connections seem to play an important role in multi-*R* richly-connected selection networks that are exposed to certain arrangements, such as intradimensional discrimination (Burgos, 1996). In the present case, a more specific form of superstition could be achieved by adding an inhibitory NPE, such that the activation of a given *R* NPE inhibits other *R* NPEs. The activation of the inhibited NPEs would be expected not to show any substantial increment. Alternatively, the activations of different *R* NPEs can be made mutually exclusive. It is not obvious, however, whether or not such a network would show any substantial increase in the activation of any of its *R* NPEs. One difficulty with inhibitory connections is that in order to be beneficial they must be in the right places.

The second challenge demands a role for response-dependent reinforcement in selection networks, which, as I mentioned, takes us out of the realm of superstition. Work in this respect with selection networks remains to be done. However, it is clear that response-dependent reinforcement remains the best procedure available for making temporal contiguity between specific stimuli (antecedent as well as consequent) and specific responses as close as possible. Selection networks may not be the exception. A sort of shaping procedure through response-dependent reinforcement thus might be needed in order to simulate more complex and specific environment-behavior relations

with multi-*R* selection networks (see Gullapalli, 1997, for initial work in this direction with a different model). To this extent, the present approach will meet the second challenge, while remaining uncommitted to a fundamental distinction between Pavlovian and operant reinforcement mechanisms.

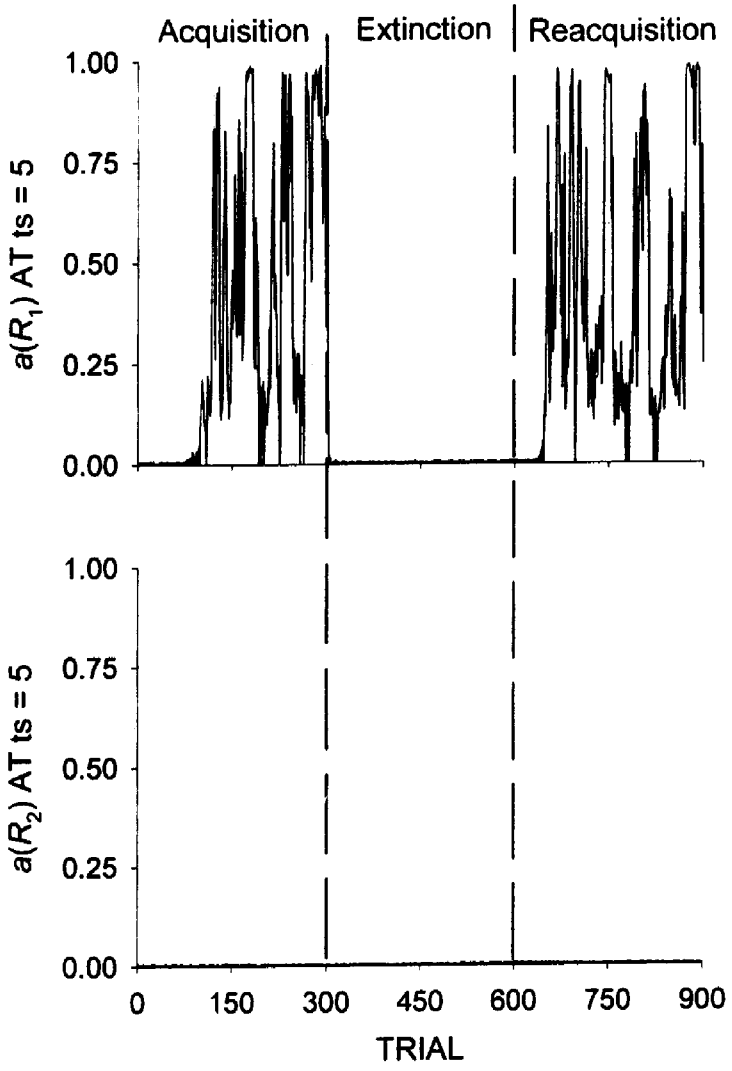

Figure 7. Results of the third superstition simulation, showing that superstition in the network depicted in Figure 6 was stimulus-specific

Concluding Remarks

I have proposed a biobehavioral interpretation of superstition based on computer simulations with selection neural networks. The interpretation endorses the existence of a single reinforcement mechanism for Pavlovian conditioning, superstition, and operant conditioning, maintaining only a phenomenalistic distinction among the three. Such an endorsement involves the incorporation of neural principles into our accounts of behavior through neural-network modeling. To conclude the paper, I want to discuss briefly two common criticisms towards this kind of modeling in behavior analysis in general. One criticism arises from causality arguments. The other arises from model-plausibility arguments. To be sure, these arguments raise hard, extensive, and largely unresolved philosophical issues whose proper analysis goes well beyond the limits of this paper. However, a study of the history of science reveals that there usually is a philosophical disagreement lurking behind most scientific disputes, especially within disciplines in their preparadigmatic stages, like behavior analysis and psychology in general. So the issues (or at least some of them) must be acknowledged, if only to prompt future discussions. It is often the case that a scientific dispute can be resolved through a careful analysis of the underlying philosophical issues. Acknowledging them is the first step towards such an analysis.

According to the first criticism, the incorporation of neural principles into our accounts of behavior transfers causality claims from the environment back into the inner organism. Hence, the argument goes, although neural principles may constitute an advance over inferred processes (in that the former are derived through experimental analysis), they still result in an internalist psychology, that is, a psychology that puts the causes of behavior inside the organism. This argument arises from the assumption that taking the nervous system into account forces us to attribute a causal role to neural events, processes, and even structures in order to explain behavior. This assumption underlies claims about such events, processes, and structures constituting "neural bases of behavior", and about nervous systems "being responsible for behavior" or "controlling behavior", insofar as "basis", "responsible", "control", and similar terms have a causal load.

Although many (perhaps most) neuroscientists adopt the assumption in question, it is not a logical necessity. Certain philosophical accounts of causality and explanation (e.g., Salmon, 1998; von Wright, 1974) allow for an incorporation of neural principles without falling into an internalist psychology. The basic idea here would be to view neural events and processes as components of our *descriptions* (as opposed to explanations) of what happens to organisms when they are reinforced. In this sense, neural events and

processes would provide us with a more complete, more elaborate description of the *effects* of the reinforcement contingencies (Burgos & Donahoe, in press). This is the idea underlying a radical-behaviorist interpretation of notions such as motivation and emotion as referring to *by-products* of the environmental circumstances (Skinner, 1974). In this manner, the external environment would retain its causal status. Another possibility is to remain neutral with respect to the causality issue, and regard neural events and processes as *participating* in environment-behavior relations. This strategy advises postponing a discussion of whether such participation constitutes a cause of behavior, an effect of the environment, or a complex mixture of both, until philosophical disagreements on explanation and causality are settled.

Similar considerations apply to a second criticism, which arises from plausibility arguments. The basic criticism goes like this (e.g., Hutchison, 1997; Marr, 1997, this issue; Smolensky, 1990; Uttal, 1990, 1993). Neural networks reflect very little, if anything, about the actual neural processes that occur in learning and memory. In fact, no available neural-network model has been able to simulate any real nervous system at any level of organization. At best, the criticism goes, neural networks reflect very general dynamical properties that can be found in many systems other than nervous systems. As a result, we need know very little (if anything) about real nervous systems in order to model (predictively or postdictively) at least certain kinds of behavior. Therefore, the criticism concludes, neural plausibility constitutes a weak justification for incorporating neural principles into our accounts of behavior through neural-network modeling.

Once again, a number of extensive, difficult, and unresolved philosophical issues are raised. One issue is the nature and purpose of scientific models. A proper justification of the criticism in question must arise from a careful examination of this issue. Otherwise, the criticism will amount to little more than an intuitive speculation. Not that this is inherently wrong, of course. However, much more is at stake than an innocent philosophical exercise. Promising theories and, with them, potentially fruitful avenues of scientific exploration could be prematurely dismissed.

One could reject the appeal to philosophical research on the basis that these issues remain unresolved and scientists cannot wait for philosophers to resolve them. However, the mere search for a resolution has generated a considerable arsenal of analytical tools that are at the scientists' disposal for stronger, more informed, and systematic philosophical discussion. For example, many scientists (if not most) still adopt the standard, logical-empiricist view of scientific theories, unaware of its deep logical and pragmatic difficulties. This view has been rejected within the philosophy of science, in favor of less problematic, more fruitful views (e.g., Suppe, 1977, 1989; Stegmüller, 1973,

1979; Balzer, Moulines, & Sneed, 1987), according to which scientific theories are something more than mere collections of statements. Moreover, some of these views have been applied to the problem of psychoneural reduction, with results that have implications for the validity of the criticism under consideration (see Bickle, 1998).

Another issue has to do with how theoretical understanding relates to empirical knowledge. There are several considerations to be made in this regard. To begin with, there is no doubt that *current* neural-network models are extremely simple, relative to the extreme complexity of real brains. Our empirical knowledge of brains and behavior is certainly running well ahead of our theoretical understanding of the brain-behavior relation. However, this is not a necessarily permanent situation. We must be patient, wait for the relevant evidence, and give modeling work a fair chance to catch up with the experimental work. Given the potential complexity of the brain-behavior relation, this process may take much longer than anyone expects. In the meantime, the present approach advises us to keep searching for neurally more plausible models, even if the best we can do for the moment is to obtain very implausible ones. Implausible models are the first steps towards plausible models.

Impossibility axioms are often postulated, according to which our theoretical understanding of the brain-behavior relation is doomed to forever run behind our empirical knowledge of brains and behavior. Such axioms, however, are little more than hypotheses about the evolution of science and the nature of scientific change (yet another philosophical issue), and are largely motivated by intellectual impatience. There is nothing inherently erroneous with formulating this kind of hypothesis, of course. What is questionable is to regard them as demonstrated, incontrovertible truths, and use them as valid criteria for swift verdicts on scientific theories. For better or for worse, theory choice is a long and tortuous process that does not admit quick judgments.

Another difficulty with this kind of impossibility axiom is that it assumes that either our empirical knowledge of brains and behavior will grow without end, or our theoretical understanding of the brain-behavior relation will always be incomplete, even if we reject the first assumption. The second assumption, however, amounts to the truism that models are inherently simplifying devices. Our theoretical understanding thus will necessarily run behind our empirical knowledge, no matter how complete the latter. There will always be a gap between our empirical knowledge of brains and behavior, and our theoretical understanding of the brain-behavior relation. The issue, then, is not whether a gap will exist, but how wide it will be. According to the criticism under consideration, the current gap is abysmal, and one cannot disagree with this assessment. However, the gap does not necessarily have to remain equally

abysmal in the future. There simply is no valid criterion for predicting how wide it will be. In this sense, the criticism in question commits the *historicist fallacy*, the assumption that the current state of scientific knowledge provides a valid basis for inferring its future states (see Popper, 1957, 1982).

The argument under consideration functions by discouraging the search for neural plausibility as hopeless, on the basis that the resulting models are too simplistic (as if there were a valid criterion to decide between acceptably and unacceptably simplistic models) and that this simplicity is insurmountable in principle (as if there were a valid criterion to predict future states of scientific knowledge from present ones). The effectiveness of this strategy arises partially from construing model plausibility in terms of completeness relative to our current empirical knowledge. A more encouraging and fruitful approach is to construe model plausibility as a matter of degree relative to other available models. How wide the gap is between our theoretical understanding and our empirical knowledge thus becomes less important than how wide it is between different models. We believe that the present model is somewhat neurally more plausible than others. We also acknowledge that some models are more closely guided by hard experimental evidence on specific nervous systems. However, these models tend to lack the kind of behavioral plausibility we seek. This suggests that both kinds of plausibility may very well be antagonistic (too much of one kind of plausibility may be detrimental for the other), another possibility that must be taken into account when neural-network models are evaluated. On this basis, our strategy has been to aim at a balance between the two kinds of plausibility.

In conclusion, an emphasis on neural plausibility might indeed be a weak justification for doing neural-network modeling, but only for the time being (a time that certainly promises to be prolonged). In view of the above reflections, such an emphasis remains a legitimate endeavor, in spite of the extreme simplicity of current neural networks (relative to real brains) and the gap that separates our empirical knowledge from our theoretical understanding of the brain-behavior relation. This emphasis provides us with yet another demarcation criterion, in addition to subsymbolic-distributed representation and parallel processing, for separating between neural-network models and cognitivist, inferred-process ones. In this manner, neural plausibility will eventually allow us to decide between behaviorally underdetermined models, to the extent that subsymbolic-distributed representation and parallel processing result in models that are equally plausible at the behavioral level.

I have explored only the surface of criticisms against neural-network modeling in behavior analysis, and their underlying philosophical issues. A lesson to be derived from these final reflections is that proper evaluations of any kind of modeling must take into account the details of the philosophical

issues involved. We must be careful not to dismiss any modeling strategy lightly, on the basis of a few general philosophical intuitions. Otherwise, potentially fruitful research lines might be prematurely abandoned. To use Sober's (1993) fortunate phrasing about sociobiology models in the present context, "there is no 'magic bullet' that shows that [neural-network models are] and must remain bankrupt, nor any that shows that [they] must succeed. Any discussion of the adequacy of [neural-network] models inevitably must take the models one by one and deal with details" (p. 185). We must not let our criticisms towards neural-network modeling (or any other kind of modeling) be motivated by the assumption that there is a magic bullet.

REFERENCES

Balzer, W., Moulines, C. U., & Sneed, J. D. (1987). *An architectonic for science*. Dordrecht: Reidel.

Bickle, J. (1998). *Psychoneural reduction: The new wave*. Cambridge, MA: MIT Press.

Bolles, R. C. (1970). Species-specific defense reactions and avoidance learning. *Psychological Review, 77*, 32-48.

Boren, J. J., & Devine, D. D. (1968). The repeated acquisition of behavioral chains. *Journal of the Experimental Analysis of Behavior, 11*, 651-660.

Burgos, J. E. (1996). *Computational explorations of the evolution of artificial neural networks in Pavlovian environments*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Burgos, J. E. (1997). Evolving artificial neural networks in Pavlovian environments. In J. W. Donahoe, & V. P. Dorsel (Eds.), *Neural-network models of cognition: Biobehavioral foundations* (pp. 58-79). Amsterdam, Netherlands: Elsevier Science Press.

Burgos, J. E. (1999). Una reconstrucción neurocomputacional del problema de los dos tipos de condicionamiento. In A. L. Rangel, L. M. Sánchez, M. Lozada, & C. Silva (Eds.), *Contribuciones a la psicología en Venezuela, Tomo III* (pp. 215-248). Caracas: Fondo Editorial de Humanidades, Universidad Central de Venezuela.

Burgos, J. E., & Donahoe, J. W. (in press). Structure and function in selectionism: Implications for complex behavior. In J. Leslie, & D. Blackman (Eds.), *Issues in experimental and applied analyses of human behavior*. Reno, NE: Context Press.

Donahoe, J. W., & Burgos, J. E. (1999). Timing without a timer. *Journal of the Experimental Analysis of Behavior, 71*, 257-263.

Donahoe, J. W., & Burgos, J. E. (in press). Behavior analysis and revaluation. *Journal of the Experimental Analysis of Behavior*.

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.

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.

Donahoe, J. W., & Palmer, D. C. (1989). The interpretation of complex human behavior: Some reactions to parallel distributed processing, edited by J. L. McClelland, D. E. Rumelhart, & the PDP Research Group. *Journal of the Experimental Analysis of Behavior, 51*, 399-416.

Donahoe, J. W., & Palmer, D. C. (1994). *Learning and complex behavior*. Boston: Allyn & Bacon.

Donahoe, J. W., Palmer, D. C., & Burgos, J. E. (1997). The S-R issue: Its status in behavior analysis and in Donahoe and Palmer's *Learning and complex behavior*. *Journal of the Experimental Analysis of Behavior, 67*, 193-211.

García, J., McGowan, B. K., & Green, K. F. (1972). Biological constraints on conditioning. In A. H. Black, & W. F. Prokasy (Eds.), *Classical conditioning, Vol. II* (pp. 3-27). New York: Appleton-Century-Crofts.

Gullapalli, V. (1997). Reinforcement learning of complex behavior through shaping. In J. W. Donahoe, & V. P. Dorsel (Eds.), *Neural-network models of cognition: Biobehavioral foundations* (pp. 302-314). Amsterdam, Netherlands: Elsevier Science Press.

Herrnstein, R. J. (1966). Superstition: A corollary of the principles of operant conditioning. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application* (pp. 33-35). New York: Appleton-Century-Crofts.

Hinde, R., & Stevenson-Hinde, J. (1973). *Constraints on learning*. London: Academic Press.

Hutchison, W. R. (1997). We also need complete behavioral models. *Journal of the Experimental Analysis of Behavior, 67*, 224-228.

Konorski, J., & Miller, S. (1937a). On two types of conditioned reflex. *Journal of General Psychology, 16*, 264-272.

Konorski, J., & Miller, S. (1937b). Further remarks on two types of conditioned reflex. *Journal of General Psychology, 17*, 405-407.

Marr, M. J. (1997). The eternal antithesis: A commentary on Donahoe, Palmer, and Burgos. *Journal of the Experimental Analysis of Behavior, 67*, 232-235.

Meehl, P. E. (1950). On the circularity of the law of effect. *Psychological Bulletin, 47*, 52-75.

Morse, W. H., & Skinner, B. F. (1957). A second type of superstition in the pigeon. *American Journal of Psychology, 70*, 308-311.

Popper, K. R. (1957). *The poverty of historicism*. London: Routledge & Kegan Paul.

Popper, K. R. (1982). *The open universe: An argument for indeterminism* (Vol. 2 of the postscript to *The logic of scientific discovery*). London: Hutchinson & Company.

Salmon, W. C. (1998). *Causality and explanation*. Oxford University Press.

Schwartz, B., & Gamzu, E. (1977). Pavlovian control of operant behavior: An analysis of autoshaping and its implications for operant conditioning. In W. K. Honig,

& J. E. R. Staddon (Eds.), *Handbook of operant behavior* (pp. 53-97). Englewood Cliffs, NJ: Prentice-Hall.

Seligman, M. E. P. (1970). On the generality of the laws of learning. *Psychological Review, 77*, 406-418.

Shettleworth, (1972). Constraints on learning. In D. S. Lehrman, R. A. Hinde, & E. Shaw (Eds.), *Advances in the study of behavior* (Vol. 4) (pp. 1-68). New York: Academic Press.

Skinner, B. F. (1935). Two types of conditioned reflex and a pseudo type. *Journal of General Psychology, 12*, 66-77.

Skinner, B. F. (1937). Two types pf conditioned reflex: A reply to Konorski and Miller. *Journal of General Psychology, 16*, 272-279.

Skinner, B. F. (1948). "Superstition" in the pigeon. *Journal of Experimental Psychology, 52*, 270-277.

Skinner, B. F. (1969). *Contingencies of reinforcement: A theoretical analysis*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1974). *About behaviorism*. New York: Random House.

Skinner, B. F. (1975). The shaping of phylogenic behavior. *Journal of the Experimental Analysis of Behavior, 24*, 117-120.

Skinner, B. F. (1981). Selection by consequences. *Science, 213*, 501-504.

Smolensky, P. (1990). Connectionism and the foundations of AI. In D. Partridge, & Y. Wilks (Eds.), *The foundations of artificial intelligence* (pp. 306-326).

Sober, E. (1993). *Philosophy of biology*. Boulder, CO: Westview.

Staddon, J. E. R. (1977). Schedule-induced behavior. In W. K. Honig, & J. E. R. Staddon (Eds.), *Handbook of operant behavior* (pp. 125-152). Englewood Cliffs, NJ: Prentice-Hall.

Staddon, J. E. R., & Ayres, S. L. (1975). Sequential and temporal properties of behavior induced by a schedule of periodic food delivery. *Behaviour, 54*, 26-49.

Staddon, J. E. R., & Simmelhag, V. L. (1971). The "superstition" experiment: A re-examination of its implications for the principles of adaptive behavior. *Psychological Review, 78*, 3-43.

Stegmüller, W. (1973). *Theorienstrukturen un theoriendynamik*. Heidelberg: Springer-Verlag. Spanish translation: *Estructura y dinámica de las teorías*, by C. U. Moulines (1983). Barcelona: Ariel.

Stegmüller, W. (1979). *The structuralist view of theories: A possible analogue of the Bourbaki programme in physical science*. Heidelberg: Springer-Verlag.

Suppe, F. (1977). The search for philosophic understanding of scientific theories. In F. Suppe (Ed.), *The structure of scientific theories* (pp. 3-232). University of Illinois Press.

Suppe, F. (1989). *The semantic conception of theories and scientific realism*. University of Illinois Press.

Tonneau, F., & Sokolowski, M. B. C. (in press). Pitfalls of behavioral selectionism. To appear in F. Tonneau, & N. S. Thompson (Eds.), *Perspectives in ethology: Vol. 13. Evolution, culture, and behavior*. New York: Plenum.