

# Response Selection Using Neural Phase Oscillators

*J. Acacio de Barros\**, *G. Oas†*, and *P. Suppes†*

\*Liberal Studies, San Francisco State University, San Francisco, CA, 94132

†CSLI, Stanford University, Stanford, CA 94305

## Abstract

In a recent paper, Suppes et al. (2012) used neural oscillators to create a model, based on reasonable neurophysiological assumptions, of the behavioral stimulus-response (SR) theory. In this paper, we describe the model in a less mathematical and more physical and intuitive way.

## 1 Introduction

Understanding how we process information is perhaps the most challenging of the current scientific endeavors. This challenge comes from the fact that our brain is tremendously complicated, as it is constituted of many different components that are, by themselves, complex, but that also seem to sometimes interact holistically with each other. Among the approaches to try and understand the brain, the most prominent ones are the top-down and bottom-up. In the top-down approach, we start with the higher-level functions, and we try to go from them to their underlying mechanisms. So, an example of such approach would be the field of cognitive neuroscience, where often one starts with experiments in cognitive psychology and try to understand them from principles in neuroscience (Adolphs, 2003). In the bottom-up approach, one tries to start with neurophysiology, and by studying how each elementary component works, one tries to see how higher functions arise from such components or their interaction (Kandel et al., 2000).

Each of those approaches have their shortcoming. For example, one of the main issues is what we may call a problem of scale. When trying to understand a complex system, the first question that arises is how detailed we need to be. In the case of the brain, some researchers say that we need to go all the way down to the chemical reactions in the synapses. Others argue that individual neurons hold the key to understanding brain computation. Yet another view is that collections of neurons are important. So, when trying to understand how the brain works, our first problem is where to begin. Regardless of what scale is

chosen and where we start, ultimately we would need to understand the whole process if we were to claim to have understood the brain.

The main problem with connecting a higher scale with a lower one is due to its complexity. For example, evidence exists that higher cognitive processes involve tens to hundreds of thousands of neurons, interacting with each other in very complex ways. Modeling such processes require the use of powerful computers. But, even when a model is shown to work from the underlying neuronal dynamics, the use of massive computer simulations helps little in understanding, in an intuitive or conceptual way, what is actually happening. The system is simply too complex.

To deal with the issue of complexity, different approaches can be taken. One possible route is to find physically plausible arguments that impose constraints on the system's dynamics, therefore reducing it to fewer degrees of freedom. This is the approach taken by Suppes et al. (2012). In their paper, a large number of independent neurons was modeled by a single dynamical parameter determined by the phase of a neural oscillator. They then showed that under certain reasonable assumptions, the main characteristics of behavioral stimulus-response (SR) theory could be described by neural oscillators. The use of neural oscillators thus provided a significant reduction on the number of degrees of freedom, allowing for the physical interpretation of many different parameters in the model.

In this paper we present the work of Suppes et al. (2012) with emphasis on the physics and intuition behind the model. Our goal is to make this model more understandable, as many of the concepts used in our previous paper are not well-known to certain audiences. For example, while all physicists have an excellent knowledge of oscillations and interference and could easily follow the arguments leading from neurons to oscillators, only a few would feel comfortable with the mathematical learning theories used. Neuroscientists, on the other hand, would probably feel at home with neurons and learning theories, but not so much with oscillators and interference. Neither would most psychologists. Here we focus on the intuitions behind the physics, with the hopes that, in conjunction with Suppes et al. (2012), psychologists and neuroscientists could benefit more from the insights that we obtain from the model.

## 2 A Brief Review of SR theory

Stimulus-response theory (or SR theory; see Suppes and Atkinson (1960)) is one of the most successful behavioral learning theories in psychology. Though it has decreased in importance in current psychology, we chose to model SR theory for the following reasons. First, it is based on a rigid trial structure, which permits its concepts to be formally axiomatized, resulting in many important non-trivial but illuminating representation theorems (Suppes, 2002). In fact, the theory is rich enough to represent language in it. Second, despite its few parameters (the learning probability  $c$  and the number of stimuli), it has been shown to fit well empirical data in a variety of experiments. Finally, as we

showed in Suppes et al. (2012), SR theory seems to have natural counterparts at a neuronal level, and is, in some sense (though, sadly, not in its mathematical form) still used by neuroscientists.

Here we present the mathematical version of SR theory for a continuum of responses, formalized in terms of a stochastic process (here we follow Suppes et al., 2012). Let  $(\Omega, \mathcal{F}, P)$  be a probability space, and let  $\mathbf{Z}$ ,  $\mathbf{S}$ ,  $\mathbf{R}$ , and  $\mathbf{E}$  be random variables, with  $\mathbf{Z} : \Omega \rightarrow E^{|S|}$ ,  $\mathbf{S} : \Omega \rightarrow S$ ,  $\mathbf{R} : \Omega \rightarrow R$ , and  $\mathbf{E} : \Omega \rightarrow E$ , where  $S$  is the set of stimuli,  $R$  the set of responses, and  $E$  the set of reinforcements. Then a trial in SR theory has the following structure:

$$\mathbf{Z}_n \rightarrow \mathbf{S}_n \rightarrow \mathbf{R}_n \rightarrow \mathbf{E}_n \rightarrow \mathbf{Z}_{n+1}. \quad (1)$$

The trial structure works the following way. Trial  $n$  starts with a certain state of conditioning and a sampled stimulus. Once a stimulus is sampled, a response is computed according to the state of conditioning. Then, reinforcement follows, which can lead (with probability  $c$ ) to a new state of conditioning for trial  $n + 1$ . In more detail, at the beginning of a trial, the state of conditioning is represented by the random variable  $\mathbf{Z}_n = (z_1^{(n)}, \dots, z_m^{(n)})$ . The vector  $(z_1^{(n)}, \dots, z_m^{(n)})$  associates to each stimuli  $s_i \in S$ ,  $i = 1, \dots, m$ , where  $m = |S|$  is the cardinality of  $S$ , a value  $z_i^{(n)}$  on trial  $n$ . Once a stimulus  $\mathbf{S}_n = s_i$  is sampled with probability  $P(\mathbf{S}_n = s_i | s_i \in S) = \frac{1}{m}$ , its corresponding  $z_i^{(n)}$  determines the probability of responses in  $R$  by the probability distribution  $K(r | z_i^{(n)})$ , i.e.  $P(a_1 \leq \mathbf{R}_n \leq a_2 | \mathbf{S}_n = s_i, \mathbf{Z}_{n,i} = z_i^{(n)}) = \int_{a_1}^{a_2} k(x | z_i^{(n)}) dx$ , where  $k(x | z_i^{(n)})$  is the probability density associate to the distribution, and where  $\mathbf{Z}_{n,i}$  is the  $i$ -th component of the vector  $(z_1^{(n)}, \dots, z_m^{(n)})$ . The probability distribution  $K(r | z_i^{(n)})$  is the smearing distribution, and it is determined by its variance and mode  $z_i^{(n)}$ . The next step is the reinforcement  $\mathbf{E}_n$ , which is effective with probability  $c$ , i.e.  $P(\mathbf{Z}_{n+1,i} = y | \mathbf{S}_n = s_i, \mathbf{E}_n = y, \mathbf{Z}_{n,i} = z_i^{(n)}) = c$  and  $P(\mathbf{Z}_{n+1,i} = z_i^{(n)} | \mathbf{S}_n = s_i, \mathbf{E}_n = y, \mathbf{Z}_{n,i} = z_i^{(n)}) = 1 - c$ . The trial ends with a new (with probability  $c$ ) state of conditioning  $\mathbf{Z}_{n+1}$ .

### 3 Oscillator model

In this section we will describe intuitively the oscillator model. We start by arguing for the use of neural oscillators as a way to model the brain at a system level. We then discuss how we can represent in a mathematically sensible way these oscillators. Finally, we show how response computations and learning can be modeled using this theoretical apparatus. Readers interested in more detail are referred to Suppes et al. (2012).

There are many different ways in which researchers try to figure out how the brain works. For example, in cognitive neuroscience, among the most popular

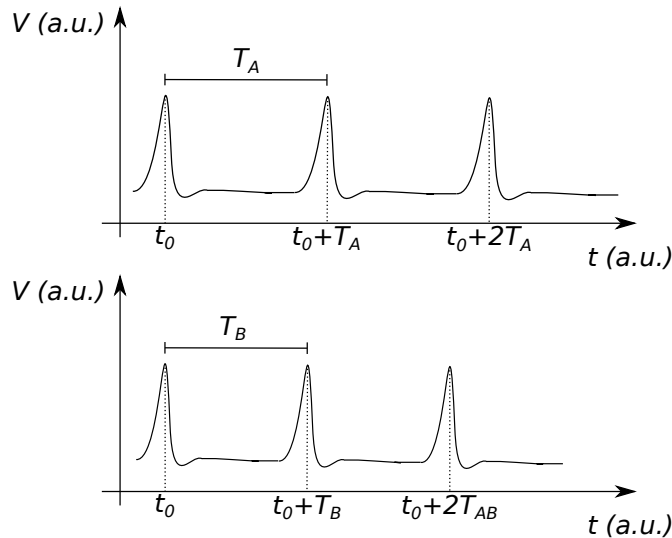


Fig. 1: Approximate shape of the action potentials  $V_A$  and  $V_B$  as a function of time  $t$  for two uncoupled neurons  $n_A$  and  $n_B$  firing periodically, with periods  $T_A$  and  $T_B$ . For simplicity, we chose a  $t_0$  when both neurons fire simultaneously.

research techniques are fMRI (functional magnetic resonance imaging), MEG (magnetoencephalogram), and EEG (electroencephalogram). MEG and EEG measure the electrical activities in the brain, whereas fMRI measures changes in blood flow associated with higher metabolic rates. While fMRI's popularity is due to its better spatial resolution, MEG and EEG present significantly better time resolution. However, what these techniques have in common is that, in order to measure a signal from the brain, they require a large numbers of neurons to fire synchronously. To make our point, let us focus on EEG (though MEG would be adequate too). There are many experiments (see Carvalhaes et al. (2012) and references) showing that the EEG data are allow a good representation of language or visual imagery. Thus, neurophysiological evidence points toward language being an activity involving large collections of synchronizing neurons, and we will center our model exactly on this.

Before we show how to describe such collections of synchronizing neurons mathematically, it is useful to think about the physical mechanisms of synchronization. Let us look first at individual neurons, and then think about ensembles of neurons. Figure 1 shows the qualitative behavior of two neurons  $n_A$  and  $n_B$  firing periodically, with  $T_B < T_A$ . What happens if we now couple  $n_A$  to an excitatory synapse coming from neuron  $n_B$ ? Because  $t_0 + T_B < t_0 + T_A$ , the excitatory coupling will increase the membrane potential of neuron  $n_A$  before  $t_0 + T_A$ , causing  $n_A$  to fire a little earlier than it would it were not connected to  $n_B$ . So, excitatory synaptic couplings between neurons can change the timing

of coupling, and this timing is changed such that the firings of both neurons approach (in this case, the firing of  $n_A$  approaches that of  $n_B$ ). In other words, excitatory couplings push  $n_A$  and  $n_B$  toward synchronization. In fact, it is possible to prove mathematically that if the number of neurons is large enough, the sum of the many weak synaptic interactions can cause a strong effect, making all neurons fire closer together (Izhikevich, 2007); even when weakly coupled, ensembles of periodically firing neurons synchronize. It is interesting to note that the argument shown above can be scaled up to distinct collections of neurons. Imagine we have two ensembles of neurons,  $N_A$  and  $N_B$ , such that neurons in them synchronize. If neurons in  $N_A$  and  $N_B$  become coupled, then the same mechanism as above will be at play, and the ensembles will synchronize among themselves. We will come back to this point later, when we talk about response mechanisms.

We are now in good shape to introduce the intuition behind the mathematical description for the dynamics of synchronization. One of the main simplifying assumptions we make is that the relevant information coded in the brain is represented by the synchronization of an ensemble of neurons. This ensemble may include tens of thousands of neurons, but because they are synchronized, we can represent them, at least in first approximation, by a single dynamical variable. To understand this, let us think about the simplest case, where an oscillator  $O(t)$  can be represented by a sine function<sup>1</sup>:

$$O(t) = A \sin \omega t, \quad (2)$$

where  $\omega = \omega(t)$  is its time-dependent frequency. Since  $\omega$  may be a function of time, the value of  $O(t)$  is completely determined by the argument of the sine, i.e. by  $\varphi = \omega t$ . The quantity  $\varphi$  is the *phase* of the oscillator  $O(t) = A \sin \varphi(t)$ . Since collections of firing neurons have very little variability in its intensity (except, as we see below, when they interfere), we can describe a neural oscillator by its phase. The interaction of a neural oscillator with other neural oscillators may change the evolution of its phase.

We emphasize that there is a certain invariance of scale in the above argument: it somehow does not matter how many neurons we have; all it matters is that their amplitude does not vary, that their couplings are strong enough to produce synchronization, and that their dynamics is encoded in the phase. Furthermore, in the same way that individual oscillating neurons synchronize to each other, a collection of coherent neurons can also synchronize to another collection of coherent neurons. Since neurons firing coherently may be described approximately by their phase, the same we can focus on the phase dynamics, instead of being concerned about the full description of the very complex dynamical system.

Now, let us look a little more into the details of the mathematics of two synchronizing oscillators. Let us start with two oscillators,  $O_1(t)$  and  $O_2(t)$ , described by their phases  $\varphi_1$  and  $\varphi_2$ . If those two oscillators are uncoupled and

---

<sup>1</sup> We use a sine function for simplicity, but the following argument is valid for periodic functions.

their frequency  $\omega$  is constant, then it is clear from equation (2) that they should satisfy the following set of differential equations,

$$\frac{d\varphi_1}{dt} = \omega_1, \quad (3)$$

$$\frac{d\varphi_2}{dt} = \omega_2, \quad (4)$$

where  $\omega_i$ ,  $i = 1, 2$ , are their natural frequencies. However, if they are weakly coupled, such that their interaction does not affect the overall form of the oscillations given by  $O_1(t)$  and  $O_2(t)$  but affects their phase, then equations (3) and (4) need to be modified to include changes to the phase. Furthermore, if the underlying interaction is such that it will make the phases approach each other, such as in the case of synaptically coupled neurons, then it is possible to show that, in first approximation, the modified dynamical equations become

$$\frac{d\varphi_1}{dt} = \omega_1 - k_{12} \sin(\varphi_1 - \varphi_2), \quad (5)$$

$$\frac{d\varphi_2}{dt} = \omega_2 - k_{21} \sin(\varphi_2 - \varphi_1), \quad (6)$$

where  $k_{ij}$  are the phase coupling strengths. If we extend this to allow for  $N$  oscillators, equations (5) and (6) then become

$$\frac{d\varphi_i}{dt} = \omega_i - \sum_{j \neq i} k_{ij} \sin(\varphi_i - \varphi_j). \quad (7)$$

Equation (7) is known as Kuramoto equation (Kuramoto, 1984), and it is widely used to describe complex systems with emergent synchronization. The strength and usefulness of Kuramoto's equation comes from two main points. First, it can be solved under certain symmetric conditions and in the limit of large  $N$ , yielding significant insight into the nature of emerging synchronization. Second, a set of weakly-coupled oscillating dynamical systems close to a Andronov-Hopf bifurcation can be described, in first approximation, by Kuramoto-like equations (see Izhikevich (2007)). For our purpose, Kuramoto's equations are a good approximation for the dynamics of coupled neural oscillators.

So, we now turn into the discussion of how we can think of stimulus and response as modeled by oscillators, and in particular by Kuramoto's equations. The basic idea is simple. Once a distal stimulus is presented, the perceptual system activates an ensemble of brain neurons,  $N_s$ , associated with it. This system itself is described by Kuramoto's equations, and, because it synchronizes, we use its average phase to describe its mean dynamics. If this stimulus elicits a response, the activation of the response neurons via synaptic couplings follows. Responses, as stimuli, are also represented by synchronously firing ensemble of neurons. The selection of a particular response happens when the stimulus oscillator synchronizes in phase with it, and such phase is determined by the relative couplings between stimulus and response oscillators. Let us now look more into its detail.

The simplest stimulus-response neural oscillator model requires three oscillators,  $O_s$ ,  $O_{r_1}$ , and  $O_{r_2}$ .  $O_s$  is the oscillator representing firing neurons corresponding to the sampling of a stimulus, and  $O_{r_1}$  and  $O_{r_2}$  are the response oscillators. Their phases are  $\varphi_s$ ,  $\varphi_{r_1}$ , and  $\varphi_{r_2}$ . Before we describe their dynamics, let us go through the process of a response computation. Whenever  $O_s$  is activated, and subsequently  $O_{r_1}$  and  $O_{r_2}$ , then the intensity of firings in each response oscillator is not only due to its firing, but also to the firings of  $O_s$ . As we mentioned earlier, a collection of firing neurons may interfere, and in this case, interference means stronger firing rates when in phase, and weaker firing rates when off of phase. Let us analyze this with a mathematically simple example of equal intensity harmonic oscillators, given by

$$O_s(t) = A \cos(\omega_0 t) = A \cos(\varphi_s(t)), \quad (8)$$

$$O_{r_1}(t) = A \cos(\omega_0 t + \delta\phi_1) = A \cos(\varphi_{r_1}(t)), \quad (9)$$

$$O_{r_2}(t) = A \cos(\omega_0 t + \delta\phi_2) = A \cos(\varphi_{r_2}(t)). \quad (10)$$

Equations (8)–(10) represent the case where the oscillators are already synchronized with the same frequency  $\omega_0$  but with relative but constant phase differences  $\delta\phi_1$  and  $\delta\phi_2$ . The mean intensity give us a measure of the excitation carried by the oscillations, and for the superposition of  $O_s(t)$  and  $O_{r_1}(t)$  it is given by

$$\begin{aligned} I_1 &= \left\langle (O_s(t) + O_{r_1}(t))^2 \right\rangle_t \\ &= \langle O_s(t)^2 \rangle_t + \langle O_{r_1}(t)^2 \rangle_t + \langle 2O_s(t)O_{r_1}(t) \rangle_t, \end{aligned}$$

where  $\langle f(t) \rangle_{t_0} = \frac{1}{\Delta T} \int_{t_0}^{t_0 + \Delta T} f(t) dt$  ( $\Delta T \gg 1/\omega_0$ ) is the time average. A quick computation yields

$$I_1 = A^2 (1 + \cos(\delta\phi_1)),$$

and, similarly for  $I_2$ ,

$$I_2 = A^2 (1 + \cos(\delta\phi_2)).$$

Therefore, the intensity depends on the phase difference between the response-computation oscillators and the stimulus oscillator.

Now, the maximum intensity of  $I_1$  and  $I_2$  is  $2A^2$ , whereas their minimum intensity is zero. If we think of  $I_1$  and  $I_2$  as competing possible responses, the maximum difference between them happens when one of their relative phases (with respect to the stimulus oscillator) is zero while the other is  $\pi$ . It is standard to use the contrast, defined by

$$b = \frac{I_1 - I_2}{I_1 + I_2}, \quad (11)$$

as a measure of how different the intensities are. From its definition,  $b$  takes values between  $-1$  and  $1$ . When  $I_1$  and  $I_2$  are as different as possible,  $|b| = 1$ ; if, on the other hand,  $I_1$  and  $I_2$  are the same,  $b = 0$ .

The contrast provides us with a useful way to think about responses that are between  $r_1$  and  $r_2$ . To see this, let us impose

$$\delta\phi_1 = \delta\phi_2 + \pi \equiv \delta\phi, \quad (12)$$

which results in

$$I_1^L = A^2 (1 + \cos(\delta\phi)), \quad (13)$$

and

$$I_2^L = A^2 (1 - \cos(\delta\phi)). \quad (14)$$

In this case, the single parameter  $\delta\phi$  is sufficient to determine the contrast, as

$$b = \cos(\delta\phi), \quad (15)$$

$0 \leq \delta\phi \leq \pi$ . So, the phase difference  $\delta\phi$  between stimulus and response oscillators codes a continuum of responses between  $-1$  and  $1$  (more precisely, because  $\delta\phi$  is a phase, the interval is in the unit circle  $\mathbb{T}$ , and not in a compact interval in  $\mathbb{R}$ ). For arbitrary intervals  $(\zeta_1, \zeta_2)$ , all that is required is a re-scaling of  $b$ .

To summarize the above arguments. When a stimulus and response oscillators fire periodically, their couplings lead to synchronization. This synchronization does not have to be in phase. But since they synchronize, their phase relation is constant after activation. This phase relation causes interference, which in turn determines the relative strength of the intensities for each response. Thus, responses are determined by the interference of oscillators, which is itself affected by the neural oscillators' couplings.

We now examine in more detail the mathematics of the stimulus and response model. Let us look at each step of (1).

### Sampling

When a stimulus is sampled, a collection of neurons start firing synchronously, corresponding to the activation of a neural oscillator,  $s_n$ . Such activation leads to a spreading of activation to oscillators coupled to the stimulus oscillator, including the response ones,  $r_1$  and  $r_2$ . Since the selection and activation of  $s_n$  involves the perceptual system, we do not attempt to model with neural oscillators this step, but simply assume their activation in a way that is consistent with the stochastic process represented in SR theory by the random variable  $\mathbf{S}_n$ . Furthermore, though it would be important to develop a detailed theory of spreading activation, we do not, as for our current purposes it suffices to simply assume the activation of  $r_1$  and  $r_2$ .

### Response

After the stimulus  $s_n$  is sampled, the active oscillators evolve for the time interval  $\Delta t_r$ , the time it takes to compute a response, according to the following set of differential equations, known as the Kuramoto equations (Kuramoto, 1984).

$$\frac{d\varphi_i}{dt} = \omega_i - \sum_{i \neq j} k_{ij} \sin(\varphi_i - \varphi_j + \delta_{ij}), \quad (16)$$



where  $k_{ij}$  is the coupling constant between oscillators  $i$  and  $j$ , and  $\delta_{ij}$  is an anti-symmetric matrix representing phase differences, and  $i$  and  $j$  can be either  $s_n$ ,  $r_1$ , or  $r_2$ . Equation (16) can be rewritten as

$$\frac{d\varphi_i}{dt} = \omega_i - \sum_j [k_{ij}^E \sin(\varphi_i - \varphi_j) + k_{ij}^I \cos(\varphi_i - \varphi_j)], \quad (17)$$

where  $k_{ij}^E = k_{ij} \cos(\delta_{ij})$  and  $k_{ij}^I = k_{ij} \sin(\delta_{ij})$ , which has an immediate physical interpretation:  $k_{ij}^E$  corresponds to excitatory couplings, whereas  $k_{ij}^I$  corresponds to inhibitory ones. These are the  $4N$  excitatory and inhibitory coupling strengths between oscillators.

$$\begin{aligned} \frac{d\varphi_{s_n}}{dt} = & \omega_0 - k_{s_n, r_1}^E \sin(\varphi_{s_n} - \varphi_{r_1}) \\ & - k_{s_n, r_2}^E \sin(\varphi_{s_n} - \varphi_{r_2}) \\ & - k_{s_n, r_1}^I \cos(\varphi_{s_n} - \varphi_{r_1}) \\ & - k_{s_n, r_2}^I \cos(\varphi_{s_n} - \varphi_{r_2}), \end{aligned} \quad (18)$$

$$\begin{aligned} \frac{d\varphi_{r_1}}{dt} = & \omega_0 - k_{r_1, s_n}^E \sin(\varphi_{r_1} - \varphi_{s_n}) \\ & - k_{r_1, r_2}^E \sin(\varphi_{r_1} - \varphi_{r_2}) \\ & - k_{r_1, s_n}^I \cos(\varphi_{r_1} - \varphi_{s_n}) \\ & - k_{r_1, r_2}^I \cos(\varphi_{r_1} - \varphi_{r_2}), \end{aligned} \quad (19)$$

$$\begin{aligned} \frac{d\varphi_{r_2}}{dt} = & \omega_0 - k_{r_2, s_n}^E \sin(\varphi_{r_2} - \varphi_{s_n}) \\ & - k_{r_2, r_1}^E \sin(\varphi_{r_2} - \varphi_{r_1}) \\ & - k_{r_2, s_n}^I \cos(\varphi_{r_2} - \varphi_{s_n}) \\ & - k_{r_2, r_1}^I \cos(\varphi_{r_2} - \varphi_{r_1}), \end{aligned} \quad (20)$$

where  $\varphi_{s_n}$ ,  $\varphi_{r_1}$ , and  $\varphi_{r_2}$  are their phases, and  $\omega_0$  their natural frequency. The solutions to (18)–(20) and the initial conditions randomly distributed at activation give us the phases at time  $t_{r,n} = t_{s,n} + \Delta t_r$ . The coupling strengths between oscillators determine their relative phase locking, which in turn corresponds to the computation of a given response, according to equation (11).

### Reinforcement and Conditioning

As we saw above, the computation of a response depends on the inhibitory and excitatory couplings between neural oscillators. Therefore, when an effective reinforcement  $\mathbf{Y}_n$  corresponding to changes in the conditioning  $\mathbf{Z}_{n+1}$  occurs, the coupling strengths change. As with stimulus and responses, we represent a reinforcement by a neural oscillator. Such oscillator, with frequency  $\omega_e$ , is activated during reinforcement, and we assume that it forces the reinforced response-computation and stimulus oscillators to synchronize with the same

phase difference of  $\delta\varphi$ , while the two response-computation oscillators are kept synchronized with a phase difference of  $\pi$ . Let the reinforcement oscillator be activated on trial  $n$  at time  $t_{e,n}$ ,  $t_{r,n+1} > t_{e,n} > t_{r,n}$ , for an interval of time  $\Delta t_e$ . Let  $K_0$  be the coupling strength between the reinforcement oscillator and the stimulus and response-computation oscillators. In order to match the probabilistic SR axiom governing the effectiveness of reinforcement, we also assume that there is a normal probability distribution governing the coupling strength  $K_0$  between the reinforcement and the other active oscillators with density

$$f(K_0) = \frac{1}{\sigma_{K_0} \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma_{K_0}^2} (K_0 - \bar{K}_0)^2 \right\}. \quad (21)$$

When a reinforcement is effective, all active oscillators phase-reset at  $t_{e,n}$ , and during reinforcement the phases of the active oscillators evolve according to the following set of differential equations.

$$\begin{aligned} \frac{d\varphi_{s_n}}{dt} = & \omega_0 - k_{s_n, r_1}^E \sin(\varphi_{s_n} - \varphi_{r_1}) \\ & - k_{s_n, r_2}^E \sin(\varphi_{s_n} - \varphi_{r_2}) \\ & - k_{s_n, r_1}^I \cos(\varphi_{s_n} - \varphi_{r_1}) \\ & - k_{s_n, r_2}^I \cos(\varphi_{s_n} - \varphi_{r_2}) \\ & - K_0 \sin(\varphi_{s_n} - \omega_e t), \end{aligned} \quad (22)$$

$$\begin{aligned} \frac{d\varphi_{r_1}}{dt} = & \omega_0 - k_{r_1, s_n}^E \sin(\varphi_{r_1} - \varphi_{s_n}) \\ & - k_{r_1, r_2}^E \sin(\varphi_{r_1} - \varphi_{r_2}) \\ & - k_{r_1, s_n}^I \cos(\varphi_{r_1} - \varphi_{s_n}) \\ & - k_{r_1, r_2}^I \cos(\varphi_{r_1} - \varphi_{r_2}) \\ & - K_0 \sin(\varphi_{r_1} - \omega_e t - \delta\varphi), \end{aligned} \quad (23)$$

$$\begin{aligned} \frac{d\varphi_{r_2}}{dt} = & \omega_0 - k_{r_2, s_n}^E \sin(\varphi_{r_2} - \varphi_{s_n}) \\ & - k_{r_2, r_1}^E \sin(\varphi_{r_2} - \varphi_{r_1}) \\ & - k_{r_2, s_n}^I \cos(\varphi_{r_2} - \varphi_{s_n}) \\ & - k_{r_2, r_1}^I \cos(\varphi_{r_2} - \varphi_{r_1}) \\ & - K_0 \sin(\varphi_{r_2} - \omega_e t - \delta\varphi + \pi). \end{aligned} \quad (24)$$

The excitatory couplings are reinforced if the oscillators are in phase with each other, according to the following equations.

$$\frac{dk_{s_n, r_1}^E}{dt} = \epsilon(K_0) [\alpha \cos(\varphi_{s_n} - \varphi_{r_1}) - k_{s_n, r_1}^E], \quad (25)$$

$$\frac{dk_{s_n, r_2}^E}{dt} = \epsilon(K_0) [\alpha \cos(\varphi_{s_n} - \varphi_{r_2}) - k_{s_n, r_2}^E], \quad (26)$$

$$\frac{dk_{r_1, r_2}^E}{dt} = \epsilon(K_0) [\alpha \cos(\varphi_{r_1} - \varphi_{r_2}) - k_{r_1, r_2}^E], \quad (27)$$

$$\frac{dk_{r_1, s_n}^E}{dt} = \epsilon(K_0) [\alpha \cos(\varphi_{s_n} - \varphi_{r_1}) - k_{r_1, s_n}^E], \quad (28)$$

$$\frac{dk_{r_2, s_n}^E}{dt} = \epsilon(K_0) [\alpha \cos(\varphi_{s_n} - \varphi_{r_2}) - k_{r_2, s_n}^E], \quad (29)$$

$$\frac{dk_{r_2, r_1}^E}{dt} = \epsilon(K_0) [\alpha \cos(\varphi_{r_1} - \varphi_{r_2}) - k_{r_2, r_1}^E]. \quad (30)$$

Similarly, for inhibitory connections, if two oscillators are perfectly off sync, then we have a reinforcement of the inhibitory connections.

$$\frac{dk_{s_n, r_1}^I}{dt} = \epsilon(K_0) [\alpha \sin(\varphi_{s_n} - \varphi_{r_1}) - k_{s_n, r_1}^I], \quad (31)$$

$$\frac{dk_{s_n, r_2}^I}{dt} = \epsilon(K_0) [\alpha \sin(\varphi_{s_n} - \varphi_{r_2}) - k_{s_n, r_2}^I], \quad (32)$$

$$\frac{dk_{r_1, r_2}^I}{dt} = \epsilon(K_0) [\alpha \sin(\varphi_{r_1} - \varphi_{r_2}) - k_{r_1, r_2}^I], \quad (33)$$

$$\frac{dk_{r_1, s_n}^I}{dt} = \epsilon(K_0) [\alpha \sin(\varphi_{r_1} - \varphi_{s_n}) - k_{r_1, s_n}^I], \quad (34)$$

$$\frac{dk_{r_2, s_n}^I}{dt} = \epsilon(K_0) [\alpha \sin(\varphi_{r_2} - \varphi_{s_n}) - k_{r_2, s_n}^I], \quad (35)$$

$$\frac{dk_{r_2, r_1}^I}{dt} = \epsilon(K_0) [\alpha \sin(\varphi_{r_2} - \varphi_{r_1}) - k_{r_2, r_1}^I]. \quad (36)$$

In the above equations,

$$\epsilon(K_0) = \begin{cases} 0 & \text{if } K_0 < K' \\ \epsilon_0 & \text{otherwise,} \end{cases} \quad (37)$$

where  $\epsilon_0 \ll \omega_0$ ,  $\alpha$  and  $K_0$  are constant during  $\Delta t_e$ , and  $K'$  is a threshold constant throughout all trials. We can think of  $K'$  as a threshold below which the reinforcement oscillator has no (or very little) effect on the stimulus and response-computation oscillators. For large enough values of  $\Delta t_e$ , the behavioral probability parameter  $\theta$  of effective reinforcement is, from (21) and (37), reflected in the equation:

$$\theta = \int_{K'}^{\infty} f(K_0) dK_0. \quad (38)$$

This relationship comes from the fact that, if  $K_0 < K'$ , there is no effective learning from reinforcement, since there are no changes to the couplings due to (25)–(36), and (18)–(20) describe the oscillators' behavior. Intuitively  $K'$  is the effectiveness parameter: the larger it is, the smaller the probability of effective reinforcement.

## 4 Final remarks

In this paper we described the neural oscillator model presented in Suppes et al. (2012), with particular emphasis to the physics and intuition behind many of the processes represented by equations (18)–(36). To summarize it, the coded phase differences were used to model a continuum of responses within SR theory in the following way. At the beginning of a trial a stimulus oscillator is activated, and with it the response oscillators. Then, the coupled oscillator system evolves according to (18)–(20) if no reinforcement is present, and according to (22)–(36) if reinforcement is present. The coupling constants and the conditioning of stimuli are not reset at the beginning of each trial, and changes to couplings correspond to effective reinforcement. Because of the finite amount of time for a response, the probabilistic characteristics of the initial conditions lead to the smearing of the phase differences after a certain time, with an effect similar to that of the smearing distribution in the SR model for a continuum of responses (Suppes, 1959).

SR theory has enjoyed tremendous success in the past, and, in a certain sense, its main features are still present in modern day neuroscience. We believe that by showing how neurons may result in theoretical structures that are somewhat similar to SR ones, as done in Suppes et al. (2012), we can provide the basis for an extension of SR theory that could be considered more realistic. For example, in our model, many parameters, such as time of response, frequency of oscillations, coupling strengths, etc., were fixed based on reasonable assumptions. However, a more detailed and systematic study should be able to relate such parameters to either underlying physiological constraints or to behavioral variations, thus opening up the possibilities for new empirical studies that go beyond SR theory. Also, in our model we postulated many features without showing or proving their dynamics from underlying neuronal dynamics. This was the case for the activation of a stimulus and the spreading of activation of a stimulus and responses. A more detailed theory based on neural oscillators of such dynamics would certainly provide with interesting empirical tests.

Finally, the use of neural oscillators and interference may also help explain certain aspects of cognition that are considered “non-classical” (de Barros and Suppes, 2009). For example, a well studied quantum-like decision making process is the violation of Savage’s sure-thing principle, shown in a series of experiments by Tversky and Shafir (Shafir and Tversky, 1992; Tversky and Shafir, 1992). Similar violations do not need any quantum-like representation in the form of a Hilbert space, as proposed in the literature, but instead can be obtained by interference of neural oscillators (de Barros, 2012b). Furthermore, the use of

neural oscillator interference even leads to predictions that are not compatible with a Hilbert space structure (de Barros, 2012a), suggesting that the use of quantum-like processes is not as quantum as many would wish.

J. Acacio de Barros and G. Oas dedicate this paper to Pat Suppes, mentor and friend, on the occasion of his 90th birthday.

## References

- Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience*, 4(3):165–178.
- Carvalhoes, C. G., de Barros, J. A., Perrau-Guimaraes, M., and Suppes, P. (2012). Using the scalp electric field to improve recognition of brain processes. *In preparation*.
- de Barros, J. A. (2012a). Joint probabilities and quantum cognition. In Khrennikov, A., editor, *Proceedings of Quantum Theory: Reconsiderations of Foundations - 6*, Växjö, Sweden. Institute of Physics. To appear.
- de Barros, J. A. (2012b). Quantum-like model of behavioral response computation using neural oscillators. *arXiv:1207.0033*.
- de Barros, J. A. and Suppes, P. (2009). Quantum mechanics, interference, and the brain. *Journal of Mathematical Psychology*, 53(5):306–313.
- Izhikevich, E. M. (2007). *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. The MIT Press, Cambridge, Massachusetts.
- Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (2000). *Principles of neural science*. McGraw-Hill, New York, NY, 4th edition.
- Kuramoto, Y. (1984). *Chemical Oscillations, Waves, and Turbulence*. Dover Publications, Inc., Mineola, New York.
- Shafir, E. and Tversky, A. (1992). Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology*, 24(4):449–474.
- Suppes, P. (1959). Stimulus sampling theory for a continuum of responses. In Arrow, K., Karlin, S., and Suppes, P., editors, *Mathematical Methods in the Social Sciences*, pages 348–365. Stanford University Press, Stanford, CA.
- Suppes, P. (2002). *Representation and Invariance of Scientific Structures*. CSLI Publications, Stanford, California.
- Suppes, P. and Atkinson, R. C. (1960). Markov learning models for multiperson interactions.
- Suppes, P., de Barros, J. A., and Oas, G. (2012). Phase-oscillator computations as neural models of stimulus–response conditioning and response selection. *Journal of Mathematical Psychology*, 56(2):95–117.

Tversky, A. and Shafir, E. (1992). The disjunction effect in choice under uncertainty. *Psychological Science*, 3(5):305–309.