A computational model of cognitive interference without neural inhibitory mechanisms

Serge Thill (serge.thill@his.se)
Informatics Research Centre, University of Skövde
PO BOX 401, Skövde, Sweden

Robert Lowe (robert.lowe@his.se)
Informatics Research Centre, University of Skövde
PO BOX 401, Skövde, Sweden

Abstract

Interference between one cognitive behavior or sensory stimulus and subsequent behaviors is a commonly observed effect in the study of human cognition and Psychology. Traditional connectionist approaches explain this phenomenon by mutually inhibiting neural populations underlying those behaviors. Here, we present an alternative model, relying on a more detailed use of synaptic dynamics, in which populations of purely excitatory neurons can nonetheless interfere with each other, causing inhibition of activation for a varying amount of time. The fundamental, biologically motivated, mechanism in the model relies on current "spilling over" from an active neural population into another one, thereby depleting the latter population's synaptic resources. The principles underlying the model may find applications even in the design of problem-solving artificial neural networks.

Keywords: Neural modeling; Synaptic dynamics; Cognitive Interference.

Introduction

The effects on cognitive performance of *interference* in the process of associating temporally contiguous behaviors or events is a well studied phenomenon in the research disciplines of psychology and animal learning. Simply, it consists of the effects on working memory or memory recall of the presence of stimuli (or motor activations) that are non-critical to the learning of particular response/event associations. In the case of animal learning, it is best understood as entailing distractorstimuli introduced prior to (proactive) or after (retroactive) a task stimulus designed to be reliably predictive of another (e.g. rewarding) stimulus. In human learning, interference can manifest in learning deficits subsequent to pairing either context relevant (Oliveri et al., 2004) or incongruent (Buccino et al., 2005) motor actions and verbal descriptions. In every day human activities, the interference effect has implications for recall of important events, e.g. eye witness testimony (see Bouton, 2007).

Laboratory controlled studies of interference often utilize the delayed matching-to-sample (DMTS) paradigm whereby the subject is required to produce the desired behavioral response over a pre-determined delay period (or inter-stimulus-interval). In such cases, interference is a function of the strength of a 'distractor' stimulus and may induce forgetting (*cf.* Roberts & Grant, 1978), impaired learning (Revusky, 1971) or memory retrieval deficits (Gordon et al., 1981).

Some forms of associative learning may be more or less prone to the interference effect. Recent neuro-scientific evidence has uncovered that areas of motor and premotor cortex that become active during physical movement overlap with areas activated during the reading of the specific affected movement, e.g. hand, foot (Hauk et al., 2004). Buccino et al. (2005) for instance found an interference effect when human subjects, required to produce hand or foot responses to particular verb forms, produced physical movements apt to the action described in the particular sentence. Latency of response increasedin this case as compared to when a movement was required that was inapt to the particular action described (see Chersi et al., 2010, for a more detailed discussion).

Models exist that attempt to capture empirically demonstrated features of the interference phenomenon specified at the level of both connectionist and more neurobiologically motivated levels of abstraction. A seminal model of McGeoch (1932) proffered a connectionist account of interference whereby responses learned during a given time window would compete for retrieval by way of mutual inhibition. Essentially, this offered a classical account of 'distractor' stimuli inhibiting the influence of task-specific stimuli. The learned associative strengths of the responses determined the 'winner' which was, however, premised on the biological implausability of there being independence, as opposed to overlap, between the available responses.

Mensink & Raaijmakers (1988) provided a stochastic search model of retrieval that was able to describe behavioral data accounting for many of the effects of interference, e.g. proactive inhibition, retroactive inhibition, spontaneous recovery - where previously learned associations become behaviorally extinguished but, presumably still reside in memory.

More recently, neural models have been put forward to account for the ability of organisms to retain spatial information about stimuli over delay periods in the face of distracting (interfering) stimuli. Spencer et al. (2009) have described how the tuning of parameters of an interaction kernel on a dynamic neural field representing spatial working memory permits the development of activation peaks. These peaks are sustained through the use of tuned local excitation and global inhibition parameters on the kernel that afford more or less robustness to noise and distractor stimuli presented to the spatial field. Self-sustained activity can be achieved through bistable unit dynamics (*cf.* Amari, 1977) such that input or noise in-

duced supra-threshold individual unit activity may be maintained even following the withdrawal of the input. Neural field and bistable dynamics through the effective coupling of spatially mapped locally excited activation peaks in different fields provide mechanisms for coping with interference effects over delays between events of motor sequences to be associated.

The assumption in the above-mentioned models and theory is that interference (or distracting stimuli) induce inhibitory effects on the activity of applicable functional circuits or psycho-behavioral states whereas chaining of activations within populations of units entails excitatory activity. In dynamic field theory, for example, distracting stimuli induce elevated levels of global inhibitory activity serving to suppress existing continuous attractor states (i.e. activation peaks) potentially below threshold levels thus serving as a medium for forgetting.

Connectionist and population coding models seeking to enhance comprehension of the interference effect typically do not concern themselves with the biophysical details of the neuron units implied in the modeling approach, relying simply on 'point-to-point' synaptic transmission. However, considering that associations of activation may be somatotopically realized in the brain, i.e. via neighboring or overlapping populations of neurons (*e.g.* Chersi et al., 2010), and that current in a given population typically overlaps with or may otherwise 'spill over' into another population, it may be instructive to produce more detailed neural models taking into account these effects in order to better understand neural substrates of behavior.

A precedent for modelling the effects of a non-synaptic neuromodulatory process only recently thought to play a significant cognitive role exists. Nitric oxide (*NO*) gas is an inter-cellular signalling mechanism found in various structures of the brain. *NO* emissions affect neighbouring cells according to a slow diffusive dynamic different to standard point-to-point synaptic transmission. *NO* diffusion has been modelled (Philippides et al., 1998) and an analogue has been applied in the domain of cognitive robotics (Husbands et al., 1998). Recent evidence also suggests a functional role in homeostatic regulation of essential metabolic variables (e.g. Canabal et al. 2007).

The particular inter-cellular signalling mechanism we are concerned with here involves current that affects neighbouring regions of cells through non-standard synaptic transmission. A complete discussion of the different mechanisms that can cause current from one neural population to leak, or "spill over" into another population is beyond the scope of this paper. However, an interesting example of such a current spillover can for instance be observed when ionic neurotransmission at the synaptic cleft is not fully absorbed by the post-synaptic receptors of the receiving cell. Ions spill over the synaptic cleft and can thereby affect neighboring neurons, possibly of other populations leading to slow-rising increases in excitatory post-synaptic currents in the affected
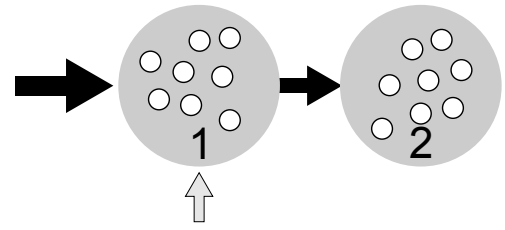

Figure 1: **Schematic of the neural model**. Two connected populations represent the neural substrate of a behavior. The behavior is triggered if the second population fires after triggering current arrives at the first one (large arrow). Weak spillover current, by itself insufficient to trigger the behavior, can also arrive at the first population (small arrow)

cells. Spillover has recently been recognized as a modulatory effect that may play a significant role in brain functioning, e.g. in the communication between the brain stem and cerebellum (Nishiyama & Linden, 2007), illustrating that neural communications do not necessarily rely solely on canonical synaptic transmission.

Here, we propose a neural model of the interference effect based primarily on synaptic dynamics. We model a sequence of two neural connected populations and show that, if spillover current from neural circuits external to the model reach the first population, activation of the second population may be prevented. Since we are mainly interested in the possible effects of the spillover current, we do not model or make assumptions on the precise underlying mechanisms. Nonetheless, we show that interference effects can be observed even though all currents are excitatory. Our model thus departs from the classically conceived models focusing on inhibitory inter-population inducement of interference. Our aim is to demonstrate that neural or neural network models of interference may be insufficient when focused solely on inter-population 'point-to-point' synaptic transmission effects. Accounting for biophysical dynamics when designing computational models or artificial neural networks may provide valuable insights to the fields of animal learning and psychology.

Methods

Neural and synaptic dynamics

We model the neural and synaptic dynamics following a standard model. The synaptic dynamics in particular take into account the fact that synaptic transmitters (or simply resources) are finite and both short term facilitation and depression can result from their dynamics (See Tsodyks et al., 1998, for a detailed discussion). Briefly, depression is caused by recognizing that synaptic resources may be "active" (in the synaptic cleft or at the post-synaptic receptors), "inactive" (returning to the pre-synaptic terminals and thus unavailable) or "recovered" (at the pre-synaptic terminals and available for release into the synaptic cleft on arrival of pre-synaptic current) and making the post-synaptic current dependent on the proportion

of active resources. The corresponding mean field equations are adapted from Tsodyks et al. (1998) with minor modifications to make the bounded nature of the resources explicit:

$$\frac{d\langle\rho\rangle}{dt} = \frac{1-\langle\rho\rangle}{\tau_{rec}} - \min\left(\langle\rho\rangle, \langle U_{SE}^1\rangle \langle\rho\rangle E(t)\right) \qquad (1)$$

$$\frac{d\langle\alpha\rangle}{dt} = -\frac{\langle\alpha\rangle}{\tau_{in}} + \min\left(\langle\rho\rangle, \langle U_{SE}^1\rangle \langle\rho\rangle E(t)\right) \qquad (2)$$

where ρ and α denote recovered and active resources respectively. Only recovered resources can generate post-synaptic current (by becoming active) and active resources affect the amplitude of post-synaptic current (Eqn. 5).The firing rate $E(t)$ is discussed further below. U_{SE}^1 is a time-varying and firing-rate dependent parameter which models short term synaptic facilitation believed to be caused by residual calcium in the synaptic cleft. It is governed by the following equations:

$$\frac{d\langle U_{SE}^-\rangle}{dt} = \frac{\langle U_{SE}^-\rangle}{\tau_{facil}} + \min\left(1 - \langle U_{SE}^-\rangle, U_{SE}\left(1 - \langle U_{SE}^-\rangle\right)E(t)\right) \qquad (3)$$

$$\langle U_{SE}^1\rangle = \langle U_{SE}^-\rangle(1 - U_{SE}) + U_{SE} \qquad (4)$$

Population dynamics

To model the effect one population of excitatory neurons may have on another, we also follow the model by Tsodyks et al. (1998). The mean firing rate of a given population r is thus dependent on the incoming current from other populations r' and external current I_r arriving directly at population r:

$$\tau_e \frac{dE_r}{dt} = -E_r + g\left(\sum_{r'} J_{rr'}\alpha_{r'} + I_r\right) \qquad (5)$$

where $J_{rr'}$ denotes the absolute strength of the connections from r' to r multiplied by the average number of such connections and $\alpha_{r'}$ is given by Eqn. 2. It can be noted here that the original model is more complex since it also caters for inhibitory populations, but those aspects are not relevant to the present work. g, finally, is a transfer function, for which we use a standard sigmoid with a threshold:

$$g(x) = \max\left(0, \frac{2}{1 + e^{(4-x)/3}} - 1\right) \qquad (6)$$

Two or more populations governed by the above dynamics can then be seen to form the neural substrate of an observable behavior. In our model, the parameter choices are: $\tau_{rec} = 1000$ms, $\tau_{in} = 100$ms, $\tau_{facil} = 530$ms, $U_{SE} = 10^{-6}$ and $J = 4$. These parameters have been chosen to produce bell-shaped activation curves in the neural populations (rather than undesired firing patterns). They mostly (except where discussed below) affect the firing rates of the neural populations but the precise choices are not critical for illustrating the effect described in the present work.
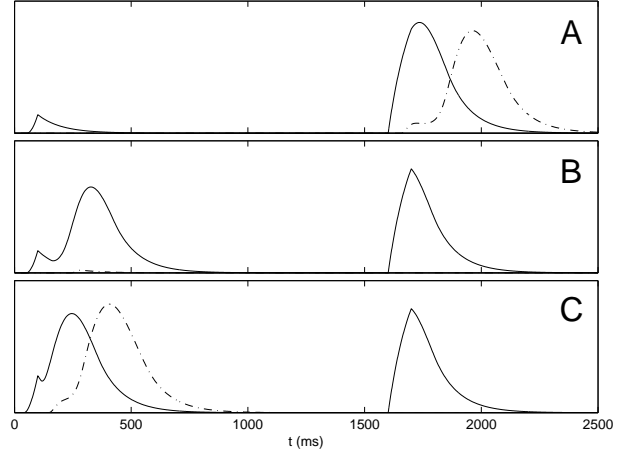

Figure 2: **The effects of weak, moderate and strong spillover current**. Solid (broken) line represents firing rate in first (second) population. Spillover current begins at t=0, behavior triggering at t=1600ms. (**A**) Spill-over current is insufficient to prevent activation of second population during behavior triggering. (**B**) Spill-over current causes significant but sub-threshold activation in the first population and prevents triggering of the second population later on. (**C**) Spillover current is sufficient to prematurely trigger the behavior.

Results

We model two connected populations of neurons (Fig. 1) which are meant to represent the neural substrate (or part thereof) of an observable cognitive behavior. Such an arrangement is for instance thought to underlie action execution in the motor cortex (Chersi et al., 2006). The behavior is "triggered" if external current arriving at the first population is of sufficient amplitude to cause activation in the second population. In other words, a behavior is successfully triggered if the second population fires after the first one was stimulated (Fig. 2A, after 1500ms). We call *triggering current* any current that, in the absence of spillover current effects, is sufficient to trigger the behavior.

Conversely, we model spillover current as a type of external current arriving at the first population but of insufficient amplitude to cause the activation of the second population (Figs. 2A and B, the first 1000ms). For the present illustrative purposes, the spillover current is modeled as lasting 100ms and increasing linearly by a small amount I_{spill} during that time. After 100ms, the current dies away instantaneously. I_{spill} has a range of possible values, with the exact choice affecting overall behavior, which is explored below. It should be noted that the observation of the reported interference effect does not critically depend on this particular choice for modeling the spillover current. Of importance is merely the fact that supra-threshold activation is generated in the first population in some way.

To illustrate the effect spillover current can have (Fig. 2),

we first determine a sufficient triggering current for the behavior in a control case with no spillover current. We then measure the post-triggering firing rate of the second population in situations where the triggering current was preceded by a spillover current δt ms earlier. Any change in firing rate compared to the control case is of interest.
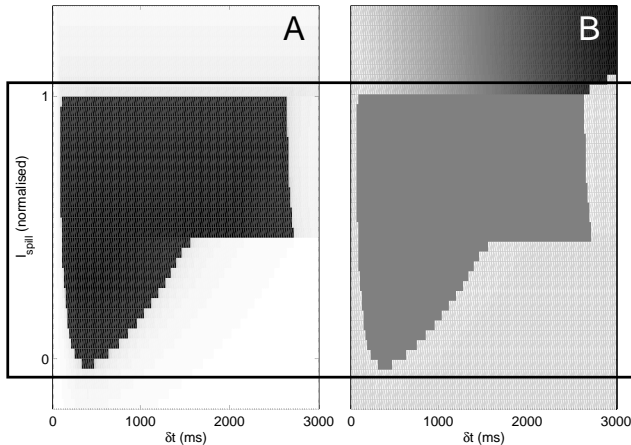
Interference without inhibition

Figure 3: **Interference effect**. Y-axes indicate spill-over current strength I_{spill}, normalised so that values of interest fall between 0 and 1. Rectangle indicates this region of interest (bounds of I_{spill}). X-axes indicate values for δt. the waiting time between end of spillover and start of behavior-triggering current. Figures are grayscale ranging from black (0) to white (maximal values of the plotted parameters). (**A**) Firing rate of the second population determined by chosen values of I_{spill} and δt. Black region indicates no firing (and therefore interference). Other regions show firing rates all at similar, close to maximal levels. The interference effect thus either causes a strong suppression of firing rate or no significant effect at all. Further, the δt values for which the interference effect is observed depend on the value of I_{spill} (see text). (**B**) Time delay between peak of activation in first population and corresponding peak in second population. If the second peak was inhibited, this information does not exist (solid gray area). Region with $I_{spill} > 1$ shows premature activation (little to no time delay, dark colors) of second population due to excessively high values of I_{spill} (see Fig. 2C). Region with $I_{spill} < 0$ shows normal separation between peaks (see Fig. 2A). Region within rectangle ($0 \leq I_{spill} \leq 1$) shows separation similar to the normal case with $I_{spill} < 0$ but not to the premature activations observed when $I_{spill} > 1$. Thus, if both populations fire, I_{spill} does not significantly affect the timing between peaks in the region of interest (rectangle).

Since spillover current that is too low (Fig. 2A) or too high Fig. (2C) is not going to cause any interesting effects, we define lower and upper bounds of I_{spill} as follows: the spillover current should be strong enough to cause some measurable

effect during an attempt at triggering the behavior but weak enough not to cause this triggering by itself (*e.g.* Fig. 2B). We define "measurable effect" simply as a difference in time-course and/or peak values in the firing rate of the second population, thus not excluding the possibility of a facilitation effect.

We find, however, that any spillover current sufficient to cause a measurable effect prevents activation of the second population (Fig. 3). The duration of this interference can vary and depends on the strength of the spillover current (Fig. 3A). For values near the lower boundary, the effect disappears if the behavior is triggered around 460ms or later after termination of the spillover current. Near the upper boundary, the interference window can last up to about 2800ms. For very small values of the spillover current, it is possible to avoid the interference effect if the behavior is triggered very shortly after the end of the spillover current (up to 340ms in the best case), since synaptic resources are depleting more slowly.

The maximal duration of the interference window is mostly affected by the choice of τ_{rec}. Interestingly, however, it is not reached monotonically. Rather, as can be seen in Fig. 3A, a threshold value for spillover current exists below which the interference effect disappears after a fraction of its maximal effect. Above the threshold, the interference effect lasts for its entire possible duration.

It would theoretically be possible for the spillover current to cause a delayed activation in the second population, rather than complete inhibition. This would be apparent if the time between the peak activation of both populations was a function of the strength of the spillover current. However, at least within the context of the work presented here, no such effect was found. Fig. 3B shows that, if the spillover current is within its bounds, it will either cause complete interference or, with a sufficient waiting period between spillover and behavior-triggering current, no effect at all. It should be noted however, that on a behavioral level, delays can still be observed. This would correspond to a control mechanism which re-triggers the behavior after noticing that the initial attempt was not succesful. Modeling these control mechanisms in detail is, however, beyond the scope of this work.

Fundamental cause

Since the behavior of the system described here is modulated only by synaptic dynamics, the cause for the observed interference effect is also found therein and illustrated in Fig. 4. Any activity within the first population will cause a reduction of recovered synaptic resources (as they become active). Since the amount of synaptic resources activated by incoming current is proportional to the recovered resources, fewer recovered resources mean smaller increase in current. If I_{spill} is very small, recovered resources do not deplete drastically during spillover current (Fig. 4A) and a following triggering current can have normal effects. If I_{spill} is larger, the recovered resources do deplete drastically but over a relatively long time-course (Fig. 4B). This slow depletion allows active resources to inactivate quickly enough to keep the proportion
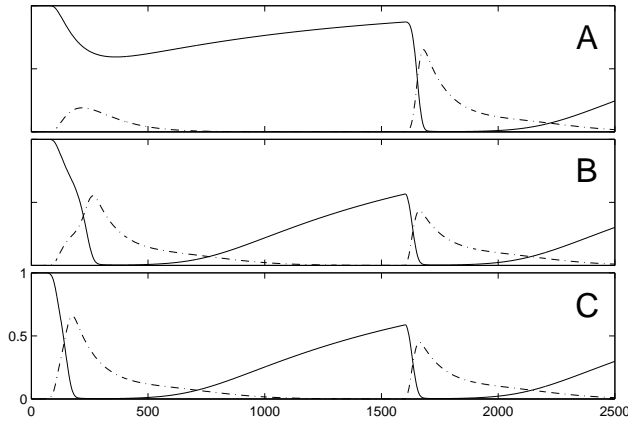
Figure 4: **Synaptic resources**. Solid (broken) line represents recovered (active) synaptic resources from the first population corresponding to the firing rates seen in Fig. 2. Spillover current begins at t=0, behavior triggering at t=1600ms. (**A**) Spill-over current causes a small decrease in recovered synaptic resources but the triggering current can activate sufficient amounts to cause firing in the second population. (**B**) Spill-over current causes complete but slow depletion of recovered resources. Not enough resources can recover and the fraction activated by the triggering current is insufficient to cause firing in the second population. (**C**) Spill-over current causes complete and fast depletion of recovered resources. Consequently, the proportion of active resources becomes sufficiently high to trigger the behavior prematurely.

of active resources below the necessary threshold for triggering the second population. At the same time, the depletion is significant enough that a later triggering current cannot activate a sufficient proportion of resources either - we observe interference. Finally, a very large value of I_{spill} works just like a triggering current: recovered resources activate quickly enough to push the proportion of active resources over the triggering threshold before it can decrease again due to inactivation.

Thus, the interference effect described here relies on a slow but significant depletion of synaptic resources. In theory, the effect of reduced available resources could be offset by the synaptic facilitation mechanism implemented here. However, since τ_{facil} is usually shorter than τ_{rec}, this is not observed in the present model.

Effects of parameter choices

Naturally, the exact values, most notably for the lower and upper boundaries of the spillover current, depend on the values chosen for the synaptic parameters in the model. The most important ones are the synaptic strength and the proportion of synaptic resources liberated. We do not address these effects in detail here but did find in a brief exploration that, as long as parameters are kept within ranges that allow a bell-shaped activation of both populations as seen in Fig. 2A after

the 1500ms mark (as opposed to, e.g. self-sustaining, chaotic or oscillatory behavior), spillover current always appears to cause interference effect.

Discussion

The model presented in this paper departs from the more classical artificial neural network models in its use of more detailed biophysical dynamics. By taking into account the fact that synaptic resources are finite, we have been able to inhibit the execution of a behavior even though all currents within the model are excitatory. While our model merely provides an alternative account compared to those relying on inhibitory dynamics, it does not necessarily replace them. However, it does illustrate the power of more detailed biophysical dynamics in a model. There is therefore a necessity to move beyond simple point-to-point artificial neural networks if the purpose of such a network is to explain cognitive phenomena.

Although we do not provide an extensive parameter exploration here, the findings are rather robust. The parameters of the synaptic model affect the firing behavior of the populations more than the effect of the spillover current (the main exceptions to this are of course τ_{rec} and τ_{facil}). Likewise, we do not need to formulate any strong assumptions on the precise nature of the spillover current because the critical aspect is merely the activation generated within the first population. The effect is thus general but further work would be needed to explore the effects of different values for τ_{rec} and τ_{facil} respectively. For instance, one could discover values for which the spillover current causes both facilitation and interference (or only facilitation). However, it should be noted that this would mainly be interesting from a theoretical perspective, since typical short term facilitation time-courses tend to be faster than depletion ones (Tsodyks et al., 1998). In fact, related work (Chersi et al., 2010) which is concerned with modeling both interference and facilitation effects simultaneously has found that in such cases, neural dynamics including inhibitory currents may provide a better explanation.

Besides their role as explanatory tools for cognitive phenomena, neural networks also find applications as computational problem-solving tools. By illustrating the effects synaptic dynamics can have on the overall output of our model, we show that moving beyond the traditional connectionist models of nodes simply connected by a signed weight can be worth considering. While this will not extend the set of computations that a neural network can perform, it may simplify the topology or facilitate training. Such benefits have for instance been previously found in GasNets (Husbands et al., 1998). These networks have proven particularly amenable to efficient search of task solution space as cognitive robotics controllers situated according to spatial and temporal environmental constraints. This adaptive potential is tapped using a diffusive, non-purely point-to-point synaptic modulatory network. Exploration of the interaction of classically conceived synaptic transmission and less orthodox means of inter-cellular communication may provide scope to investi-

gate spatial and temporal interactions relevant to the study of cognitive phenomena particularly in an embodied context (*cf* Parisi, 2004). Again, these are possibilities that need to be explored further in future work.

Conclusions

We have presented a model that can explain temporal interference effects without relying on inhibitory dynamics in the underlying neural circuitry. Rather, the behavior is explained solely by synaptic dynamics which are modeled in a simple yet biologically plausible way. The contributions of this work are twofold: (1) We provide an alternative explanation for a range of interference effects which does not rely on explicit inhibitory dynamics. (2) We highlight the benefits of modeling synaptic and biophysical dynamics in more detail, both as a computational tool which may find applications even in artificial neural networks and as an explanatory mechanism as illustrated in the present paper.

References

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, *27*, 77-87.

Bouton, M. E. (2007). *Learning and behavior: A contemporary synthesis*. Sinauer Associates, Inc. Publishers.

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action related sentences modulates the activity of the motor system: A combined tms and behavioral study. *Cognitive Brain Research*, *24*, 355-63.

Chersi, F., Mukovskiy, A., Fogassi, L., Ferrari, P. F., & Erlhagen, W. (2006). A model of intention understanding based on learned chains of motor acts in the parietal lobe. In *Proceedings of the 15th annual computational neuroscience meeting.* Edinburgh, UK.

Chersi, F., Thill, S., Ziemke, T., & Borghi, A. M. (2010). Sentence processing: linking language to motor chains. *Frontiers in Neurorobotics*, doi:10.3389/fnbot.2010.00004.

Gordon, W. C., Taylor, J. R., & Mowrer, R. R. (1981). Enhancement of short-term retention in rats with pretest cues: Effects of the training-cueing interval and the specific cueing treatment. *American Journal of Psychology*, *94*, 309-322.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*, 201-307.

Husbands, P., Smith, T., Jakobi, N., & O'Shea, M. (1998). Better living through chemistry: Evolving gasnets for robot control. *Connection Science*, *10*, 185-210.

McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review*, *39*, 352-370.

Mensink, G. J.-M., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review*, *95*, 434-455.

Nishiyama, H., & Linden, D. J. (2007). Pure spillover transmission between neurons. *Nature Neuroscience*, *10*, 675-677.

Oliveri, M., Finocchiaro, C., Shapiro, K., Gangitano, M., Caramazza, A., & Pascual-Leone, A. (2004). All talk and no action: A transcranial magnetic stimulation study of motor cortex activation during action word production. *Journal of Cognitive Neuroscience*, *16*, 374-381.

Parisi, D. (2004). Internal robotics. *Connection Science*, *16*, 325-338.

Philippides, A., Husbands, P., & O'Shea, M. (1998). Neural signaling - it's a gas! In L. N. M. Boden & T. Ziemke (Eds.), *Proceedings of the 8th international conference on artificial neural networks.* London: Springer-Verlag.

Revusky, S. H. (1971). Animal memory. In W. K. Honig & P. H. R. James (Eds.), (chap. The role of interference in association over a delay). New York: Academic Press.

Roberts, W. A., & Grant, D. S. (1978). An analysis of light-induced retroactive inhibition in pigeon short-term memory. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*, 247-260.

Spencer, J. P., Perone, S., & Johnson, J. S. (2009). Toward a unified theory of development. In J. P. Spencer, M. S. C. Thomas, & J. L. McClelland (Eds.), (p. 86-118). Oxford.

Tsodyks, M., Pawelzik, K., & Markram, H. (1998). Neural networks with dynamic synapses. *Neural Computation*, *10*, 821-835.