



Analysis of Biased Competition and Cooperation for Attention in the Cerebral Cortex

Tatyana Turova^{1,2} and Edmund T. Rolls^{3,4,5*}

¹ Mathematical Center, University of Lund, Lund, Sweden, ² IMPB - The Keldysh Institute of Applied Mathematics of the Russian Academy of Sciences, Moscow, Russia, ³ Oxford Centre for Computational Neuroscience, Oxford, United Kingdom, ⁴ Department of Computer Science, University of Warwick, Coventry, United Kingdom, ⁵ Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China

A new approach to understanding the interaction between cortical areas is provided by a mathematical analysis of biased competition, which describes many interactions between cortical areas, including those involved in top-down attention. The analysis helps to elucidate the principles of operation of such cortical systems, and in particular the parameter values within which biased competition operates. The analytic results are supported by simulations that illustrate the operation of the system with parameters selected from the analysis. The findings provide a detailed mathematical analysis of the operation of these neural systems with nodes connected by feedforward (bottom-up) and feedback (top-down) connections. The analysis provides the critical value of the top-down attentional bias that enables biased competition to operate for a range of input values to the network, and derives this as a function of all the parameters in the model. The critical value of the top-down bias depends linearly on the value of the other inputs, but the coefficients in the function reveal non-linear relations between the remaining parameters. The results provide reasons why the backprojections should not be very much weaker than the forward connections between two cortical areas. The major advantage of the analytical approach is that it discloses relations between all the parameters of the model.

Keywords: biased competition, attention, cerebral cortex, mathematical analysis, neural networks, top-down connections, bottom-up connections

OPEN ACCESS

Edited by:

Paul Miller,
Brandeis University, United States

Reviewed by:

Guy Elston,
University of South Australia, Australia

Adam Ponzi,

Okinawa Institute of Science and
Technology Graduate University,
Japan

*Correspondence:

Edmund T. Rolls
edmund.rolls@oxcns.org

Received: 08 May 2019

Accepted: 04 July 2019

Published: 31 July 2019

Citation:

Turova T and Rolls ET (2019) Analysis
of Biased Competition and
Cooperation for Attention in the
Cerebral Cortex.
Front. Comput. Neurosci. 13:51.
doi: 10.3389/fncom.2019.00051

1. INTRODUCTION

Biological systems employ a selection processing strategy for managing the enormous amount of information resulting from their interaction with the environment. This selection of relevant information is referred to as attention. One type of visual attention is the result of top-down influences on the processing of sensory information in the visual cortex, and therefore is intrinsically associated with neural interactions within and between cortical areas. Thus, elucidating the neural basis of visual attention is an excellent paradigm for understanding some of the basic mechanisms for interactions between cortical areas.

Observations from a number of cognitive neuroscience experiments have led to an account of attention termed the “*biased competition hypothesis*,” which aims to explain the computational processes governing visual attention and their implementation in the brain’s neural circuits and neural systems. According to this hypothesis, attentional selection operates in parallel by biasing an

underlying competitive interaction between multiple stimuli in the visual field toward one stimulus or another, so that behaviorally relevant stimuli are processed in the cortex while irrelevant stimuli are filtered out (Chelazzi et al., 1993; Duncan, 1996; Chelazzi, 1998; Reynolds and Desimone, 1999). Thus, attending to a stimulus at a particular location or with a particular feature biases the underlying neural competition in a certain brain area in favor of neurons that respond to the location, or the features, of the attended stimulus. This attentional effect is produced by generating signals in areas outside the visual cortex which are then fed back to extrastriate visual cortical areas, where they bias the competition such that when multiple stimuli appear in the visual field, the cells representing the attended stimulus *win*, thereby suppressing the firing of cells representing distracting stimuli (Duncan and Humphreys, 1989; Desimone and Duncan, 1995; Duncan, 1996; Reynolds et al., 1999). According to this line of work, attention appears as a property of competitive/cooperative interactions that work in parallel across the cortical modules. Neurophysiological experiments are consistent with this hypothesis in showing that attention serves to modulate the suppressive interaction between the neuronal firing elicited by two or more stimuli within the receptive field (Miller et al., 1993; Motter, 1993; Chelazzi, 1998; Reynolds and Desimone, 1999; Reynolds et al., 1999). Further evidence comes from functional magnetic resonance imaging (fMRI) in humans (Kastner et al., 1998, 1999) which indicates that when multiple stimuli are present simultaneously in the visual field, their cortical representations within the object recognition pathway interact in a competitive, suppressive fashion, which is not the case when the stimuli are presented sequentially. It was also observed that directing attention to one of the stimuli counteracts the suppressive influence of nearby stimuli.

Neurodynamical models providing a theoretical framework for biased competition have been proposed and successfully applied in the context of attention and working memory (Rolls and Deco, 2002; Rolls, 2016). In the context of attention, Usher and Niebur (1996) introduced an early model of biased competition to explain the attentional effects in neural responses observed in the inferior temporal cortex, and this was followed by a model for V2 and V4 by Reynolds et al. (1999) based on the shunting equations of Grossberg (1988). Deco and Zihl (2001) extended Usher and Niebur's model to simulate the psychophysics of visual attention by visual search experiments in humans. Their neurodynamical formulation is a large-scale hierarchical model of the visual cortex whose global dynamics is based on biased competition mechanisms at the neural level. Attention then appears as an effect related to the dynamical evolution of the whole network. This large-scale formulation has been able to simulate and explain in a unifying framework visual attention in a variety of tasks and at different cognitive neuroscience experimental measurement levels, namely: single-cells (Deco and Lee, 2002; Rolls and Deco, 2002), fMRI (Corchs and Deco, 2002, 2004), psychophysics (Deco et al., 2002; Deco and Rolls, 2004), and neuropsychology (Deco and Rolls, 2002; Heinke et al., 2002). In the context of working memory, further developments (Deco and Rolls, 2003; Deco et al., 2004; Szabo et al., 2004) managed to model in a unifying form attentional

and memory effects in the prefrontal cortex integrating single-cell and fMRI data, and different paradigms in the framework of biased competition.

A detailed dynamical analysis of the synaptic and spiking mechanisms underlying biased competition was produced by Deco and Rolls (2005). However, the parameter regions within which biased competition operates were identified by a mean field analysis, which consisted of testing a set of parameters until effective regions in the state spaces were identified.

Here we treat the biased competition system analytically for the first time. This mathematical analysis complements previous numerical results and improves our understanding of the principles of operation of the system. Although the results are presented in the context of attention, they apply more generally to interactions between cortical areas. The dynamics of cortical attractor networks, and the dynamical interactions between them that enable them to interact usefully for short-term memory and attention yet maintain separate attractors have been analyzed with the rather different approaches of theoretical physics elsewhere (Treves, 1993; Battaglia and Treves, 1998; Renart et al., 1999a,b, 2000, 2001; Panzeri et al., 2001; Rolls, 2016).

2. METHODS

2.1. The Biased Competition Network

The network to be analyzed is shown in **Figure 1**, and has the same general architecture used by Deco and Rolls (2005) to investigate the mechanisms of biased competition in a range of neurophysiological experiments. The system has forward connections J_f and K_f , and top-down backprojection connections J_b and K_b , as shown in **Figure 1**. The top-down connections are weaker, so that they can bias the bottom up inputs, but not dominate them, so that the system remains driven by the world. The top-down connections in the model correspond to the backprojections found between adjacent cortical areas in a cortical hierarchy, and in the area of memory recall, to the backprojections from the hippocampus to the neocortex (Kesner and Rolls, 2015; Rolls, 2016, 2018). The anatomical arrangement that facilitates this is that the backprojections end on the apical dendrites of cortical pyramidal cells far from the cell body, where their effects can be shunted by forward inputs that terminate on the parts of the dendrite that are electrically closer to the cell body (Rolls, 2016). In both top-down attention, and in memory recall, it is important that any bottom-up inputs from the world take priority, so that the organism is sensitive to events in the world, rather than being dominated by internal processing (Rolls, 2016). Interestingly, there is evidence that this situation is less the case in schizophrenia, in which some key forward connections are reduced in magnitude relative to the backprojections (Rolls et al., 2019). The type of neurophysiological experiment for which this model was designed is described by Deco and Rolls (2005), with one of the original neurophysiological investigations performed on object-based attention in the inferior temporal visual cortex (IT) by Chelazzi et al. (1993). The overall operation of this architecture is conceptualized as follows [see Deco and Rolls

(2005)]. Activity in neuron or population of neurons L_1 have a strong driving effect on H_1 (via J_f), as stimulus 1 acting via λ_1 is the preferred (i.e., most effective) stimulus, and a weaker effect on H_2 (via the crossed connection K_f) for which stimulus 1 is the less preferred stimulus. There are weaker corresponding backprojections J_b and K_b . Correspondingly, L_2 has a strong driving effect on H_2 (via J_f) as stimulus 2 acting via λ_2 is the preferred stimulus, and a weaker effect on H_1 (via the crossed connection K_f) for which stimulus 2 is the less preferred stimulus. L_1 is in competition with L_2 (with strength c_L), and H_1 is in competition with H_2 (with strength c_H). The present model does not imply that L_1 and L_2 are in different areas in a topographical map, but that can be easily implemented by decreasing the strength of c_L if L_1 and L_2 are not close together in the map, and the same general results hold (see Deco and Rolls, 2005). It is noted that convergence in topographically mapped systems from stage to stage is important in providing one of the bases for translation invariant visual object recognition as modeled in VisNet (Rolls, 2012, 2016), and that the dendritic morphology at different stages of processing in the visual cortical hierarchy may facilitate this (Elston, 2002; Elston and Fujita, 2014).

The nodes L correspond to V4, and the nodes H to inferior temporal visual cortex, in the neurophysiological experiment of Chelazzi et al. (1993) on object-based top-down attention. The lower nodes are L_1 and L_2 , and receive inputs λ_1 and λ_2 . L_1 has forward connections of strength J_f to higher node H_1 , and L_2 has forward connections of strength J_f to higher node H_2 . The corresponding backprojections have strength J_b . In addition, there are crossed connections K_f and K_b shown with dashed lines. The rationale for this connectivity is that the preferred stimulus for L_1 has strong effects on H_1 , but weaker effects on H_2 for which it is not the preferred stimulus. In a corresponding way, the preferred stimulus for L_2 has strong effects on H_2 , but weaker effects on H_1 for which it is not the preferred stimulus. This is implemented as follows. L_1 has forward connections of strength K_f to higher node H_2 , and L_2 has forward connections of strength K_f to higher node H_1 . The corresponding backprojections have strength K_b . The K connections are weaker than the J connections. A top-down attentional bias signal λ_2^H can be applied to node H_2 , and can bias the network to emphasize the effects of λ_2 even if λ_2 is less than or equal to λ_1 . There is competition within a level, with H_1 and H_2 competing with strength c_H , and L_1 and L_2 competing with strength c_L . The conditions for these effects to occur and for the network to be stable are analyzed. In addition, all four of the nodes can have recurrent collateral connections that produce self-excitatory effects, but the effects of these in the simulations do not affect the generic results obtained, and are not included for simplicity in the mathematical analyses.

The network can operate as follows. If a weaker than λ_1 input λ_2 is applied to L_2 , then neural populations L_1 and H_1 win the competition. However, if a top-down biased competition input λ_2^H is applied to H_2 , then this can bias the network so that H_2 wins the competition over H_1 , but also L_2 has higher activity than L_1 . Here we analyse the relation between the parameters shown in **Figure 1** that enable these effects to emerge and to be stable.

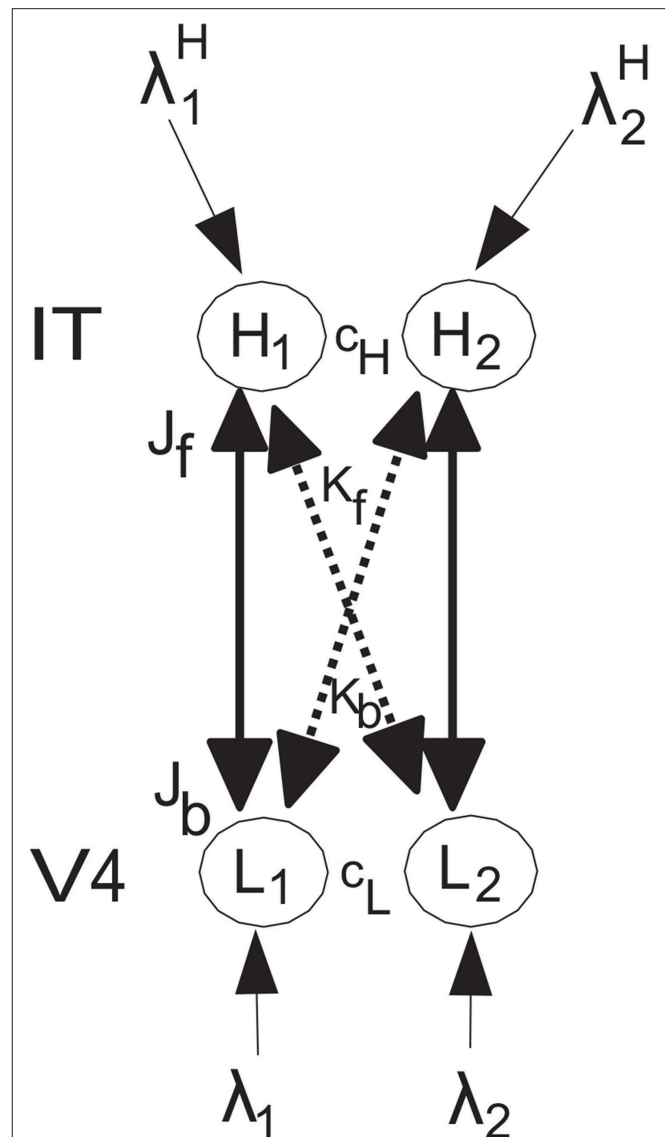


FIGURE 1 | Model architecture. The lower nodes are L_1 and L_2 , and receive inputs λ_1 and λ_2 . L_1 has forward connections of strength J_f to higher node H_1 for which stimulus 1 acting via λ_1 is the preferred input, and L_2 has forward connections of strength J_f to higher node H_2 for which stimulus 2 acting via λ_2 is the preferred input. The corresponding backprojections have weaker strength J_b . In addition, there are crossed connections shown with dashed lines. In particular, L_1 has forward connections of weak strength K_f to higher node H_2 as stimulus 1 is not the preferred stimulus for H_2 , and L_2 has forward connections of strength K_f to higher node H_1 as stimulus 2 is not the preferred stimulus for H_1 . The corresponding backprojections have weaker strength K_b . A top-down attentional bias signal λ_2^H can be applied to node H_2 , and can bias the network to emphasize the effects of λ_2 even if λ_2 is less than or equal to λ_1 . There is competition within a level, with H_1 and H_2 competing with strength c_H , and L_1 and L_2 competing with strength c_L . The conditions for these effects to occur and for the network to be stable are analyzed. In addition, all four of the nodes can in the simulations have recurrent collateral connections that produce self-excitatory effects, consistent with cortical architecture, and these produce the expected effects, but are not considered further here because the mathematical analysis focusses on the tractable case in which they are not present. The nodes L correspond to V4, and the nodes H to inferior temporal visual cortex, in the neurophysiological experiment of Chelazzi et al. (1993) on object-based top-down attention.

A feature of the model described here and also by Deco and Rolls (2005) is that local attractor dynamics are implemented within each of the neuronal populations, as recurrent collateral connections are a feature of the cerebral neocortex, and are one way to incorporate non-linear effects into the operation of the system. The new results are derived here analytically, and are illustrated with simulations with the parameters originating from the formulas obtained.

2.2. The Attentional Effects to be Modeled: Investigation 1

One top-down biased competition effect to be considered is as follows. If λ_1 is greater than λ_2 , then the activity of L_1 will be greater than the activity of L_2 . But if we apply a top-down biased competition input λ_2^H to H_2 , then at some value of λ_2^H the top-down bias will result in L_2 having as much activity as L_1 . We seek to analyse the exact conditions under which this biased competition effect occurs, in terms of all the parameters of the system.

This system has been studied previously as follows, but with a mean-field analysis to search the parameter space, rather than the analytic approach described here. **Figure 1** shows a design associated with a prediction that can be made by setting the contrast-attention interaction study in the framework of the experimental biased competition design of Chelazzi et al. (1993) involving object attention. Deco and Rolls (2005) modeled this experiment by measuring neuronal responses from neurons in neuronal population (or pool) H_1 in the inferior temporal cortex (IT) to a preferred and a non-preferred stimulus simultaneously presented within the receptive field. They manipulated the contrast or efficacy of the stimulus that was non-preferred for the neurons H_1 . They analyzed the effects of this manipulation for two conditions, namely without object attention, or with top-down object attention on the non-preferred stimulus 2 that produced input λ_2 , implemented by adding an extra bias λ_2^H to H_2 . In the previous integrate-and-fire simulations (Deco and Rolls, 2005), top-down biased competition was demonstrated, and it was found that the attentional suppressive effect implemented through λ_2^H on the responses of neurons H_1 to the competing non-preferred stimulus (2) was higher when the contrast of the non-preferred stimulus (2) was at intermediate values. However, the operation of this type of network was not examined analytically, which is the aim of the present investigation.

2.3. The Attentional Effects to be Modeled: Investigation 2

A second top-down competition effect might be considered as follows. If λ_1 is greater than λ_2 and both are applied simultaneously, then the activity of H_1 will be greater than the activity of H_2 . However, if we apply top-down bias λ_2^H to H_2 , then we can influence the activity of H_2 and H_1 (through all the connections in the system) until H_2 and H_1 have the same activity (and at the same time there will be an effect on L_1 and L_2). We wish to quantify these effects analytically in terms of all the parameters of the system.

3. DYNAMICS

We shall use the notation that $[x]^+ = \max\{x, 0\}$ for any rational value of x . We assume that $L_1(t), L_2(t), H_1(t), H_2(t)$ are defined by the following system of recurrent equations in discrete time $t = 0, 1, \dots$:

$$L_i(t+1) = \left[L_i(t) + \lambda_i(t) + \sum_{k=1,2} w_{ik}^b H_k(t) - c_L L_j(t) - \beta_L L_i(t) + \mathbf{I}\{L_i(t) > T_L\}(T_L - \alpha_L L_i(t)) \right]^+, \quad (3.1)$$

$$H_i(t+1) = \left[H_i(t) + \lambda_i^H(t) + \sum_{k=1,2} w_{ik}^f L_k(t) - c_H H_j(t) - \beta_H H_i(t) + \mathbf{I}\{H_i(t) > T_H\}(T_H - \alpha_H H_i(t)) \right]^+, \quad (3.2)$$

$$i = 1, 2, \quad i \neq j.$$

One can also use here any time step $t = 0, \tau, 2\tau, \dots$, treating τ as another parameter. Here \mathbf{I} denotes the indicator function, i.e.,

$$\mathbf{I}\{H_i(t) > T_H\} = \begin{cases} 1, & \text{if } H_i(t) > T_H, \\ 0, & \text{otherwise,} \end{cases}$$

and similarly for $\mathbf{I}\{L_i(t) > T_L\}$. This means that the terms in (3.1) and (3.2) which involve these indicator functions for the recurrent dynamics are present only if the activity in $L_i(t)$ or $H_i(t)$ is greater than the corresponding threshold T_L or T_H . In the analysis described here, we assumed that T_L and T_H were infinity, so that the recurrent dynamics were not in operation, but the recurrent dynamics were tested in the simulations.

The constants w_{ij}^f represent the synaptic weight of the connection from node L_j to node H_i , while w_{ij}^b is the synaptic weight of the connection from node H_j to node L_i . We shall assume that

$$w_{ii}^b = J_b, \quad w_{ij}^b = K_b, \quad w_{ii}^f = J_f, \quad w_{ij}^f = K_f,$$

for all $i = 1, 2, i \neq j$, where indexes b and f correspond to "back" and "forward". Furthermore, we assume that for some non-negative coefficient q (typically $0 \leq q < 1$, see more comments below)

$$J_b \geq K_b, \quad J_f \geq K_f, \quad \text{and } J_b = qJ_f, \quad K_b = qK_f. \quad (3.3)$$

Functions $\lambda_i(t), \lambda_i^H(t)$ represent the external inputs.

All the remaining constants in the system (3.1), (3.2) are free parameters; they are assumed to be non-negative, as they represent the following characteristics:

β_L is the decay term for the L nodes,
 c_L is the competition between the L nodes,
 T_L is the threshold at which an L node enters its attractor dynamics as was explained above,
 α_L is the gain factor for self-excitation in an L node for attractor dynamics,
 β_H is the decay term for the H nodes,
 c_H is the competition between the H nodes,
 T_H is the threshold at which an H node enters its attractor dynamics,
 α_H is the gain factor for self-excitation in an H node for attractor dynamics.

Note that when all the constant parameters of the system are set to be zero, including the input functions λ_i, λ_i^H , then the system

$$(L_1(t), L_2(t), H_1(t), H_2(t)) \quad (3.4)$$

remains at the initial state $(L_1(0), L_2(0), H_1(0), H_2(0))$.

We shall assume that

$$(L_1(0), L_2(0), H_1(0), H_2(0)) = (0, 0, 0, 0), \quad (3.5)$$

and that all the input functions are non-negative constants, i.e.,

$$(\lambda_1(t), \lambda_2(t), \lambda_1^H(t), \lambda_2^H(t)) = (\lambda_1, \lambda_2, \lambda_1^H, \lambda_2^H) =: \bar{\lambda}. \quad (3.6)$$

We shall address the following problems. Assume, that $\lambda_1 > \lambda_2 > 0$.

Problem I. Are there $0 \leq \lambda_1^H \leq \lambda_2^H$ such that $L_1(t) \leq L_2(t)$ for all large values of t ? This is investigated in Investigation 1.

Problem II. Are there $0 \leq \lambda_1^H \leq \lambda_2^H$ such that $H_1(t) \leq H_2(t)$ for all large values of t ? This is investigated in Investigation 2.

In other words we are looking for the conditions for the parameters of the system which allow the biased competition. Deco and Rolls (2005) found numerically an area of parameters which yields such effect. Here we derive this analytically. This allows us to treat a wide range of parameters and moreover to find out the exact relations between all the parameters of the network when the biased competition takes place.

4. SIMULATIONS OF THE OPERATION OF THE NETWORK WITH PARAMETERS SELECTED BASED ON THE ANALYSIS THAT FOLLOWS

4.1. Investigation 1

The top-down biased competition effect to be considered here is as follows. If λ_1 is greater than λ_2 , then the activity of L_1 will be greater than the activity of L_2 . But if we apply a top-down biased competition input λ_H to H_2 , then at some value of λ_H the top-down bias will result in L_2 having as much activity as L_1 . The aim was to discover the critical value of λ_H by simulation, for comparison with the analytic value.

The system specified by (3.1) and (3.2) was implemented in Matlab. The parameters shown in (3.1) and (3.2) were set as follows:

$J_f = 0.15 / 3$ (the values for these synaptic weights are in the same ratio as in Deco and Rolls, 2005)

$J_b = 0.05 / 3$

$K_f = 0.015 / 3$

$K_b = 0.005 / 3$

$\beta_L = 0.35$ (the decay term for the L nodes)

$c_L = 0.3$ (the competition between the L nodes)

$T_L = 5.0$ (the threshold at which an L node enters its attractor dynamics; in practice this was set to infinity in most of the work described, in order to prevent attractor dynamics operation in the nodes)

$\alpha_L = 0.0$ (the gain factor for self-excitation in an L node for attractor dynamics, α_L can be set as well to 0.1 if recurrent dynamics are required)

$\beta_H = 0.35$ (the decay term for the H nodes)

$c_H = 0.3$ (the competition between the H nodes)

$T_H = 5.0$ (the threshold at which an H node enters its attractor dynamics; in practice this was set to infinity in most of the work described, in order to prevent attractor dynamics operation in the nodes)

$\alpha_H = 0.0$ (the gain factor for self-excitation in an H node for attractor dynamics, α_H can be set to 0.1)

$\lambda_1 = 6.0$

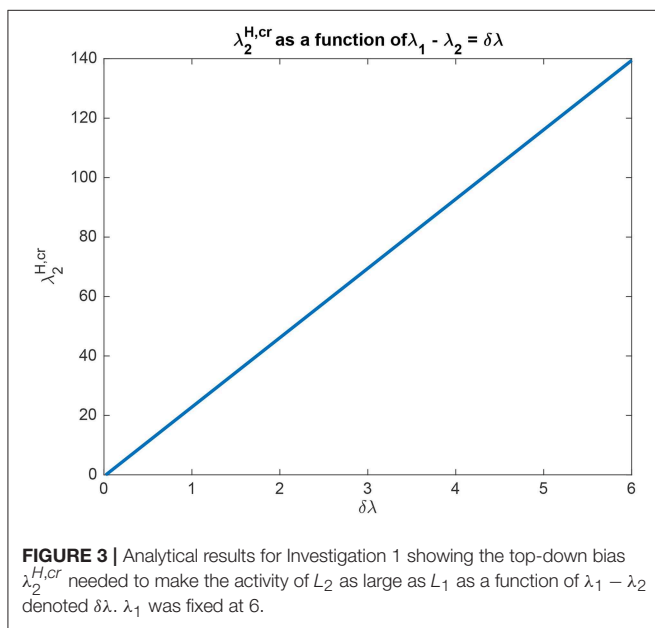
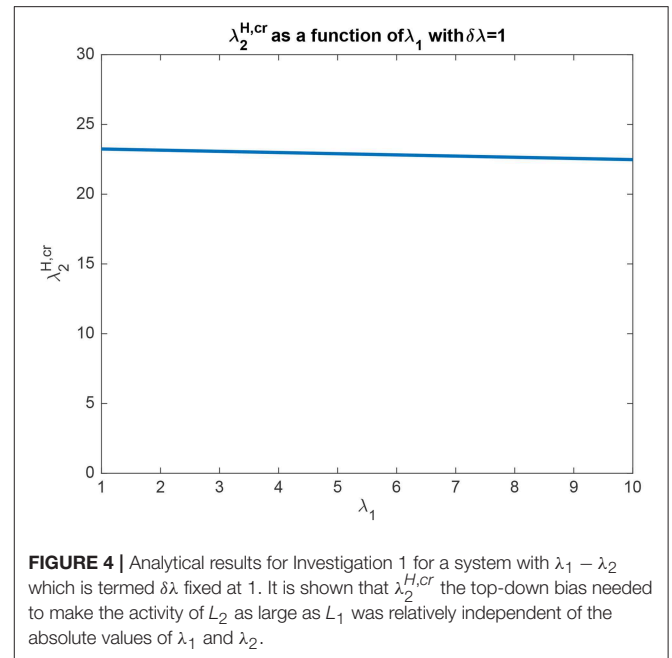
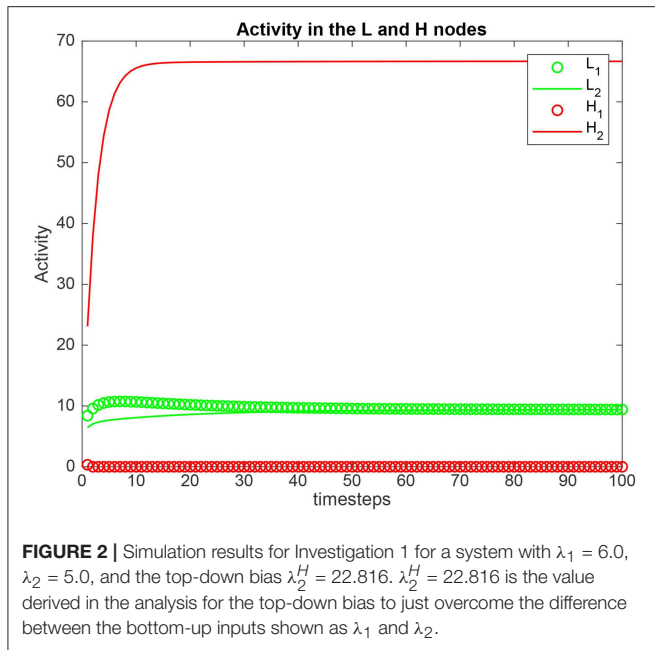
$\lambda_2 = 5.0$

$\lambda_2^H = 0.0$ or the critical value $\lambda_2^{H,cr}$ of λ_2^H derived below in the analysis for the top-down bias to overcome the bottom-up inputs shown as λ_1 and λ_2 .

For all of the simulations illustrated in this paper and for the analytic investigations, the recurrent collateral self-excitatory effects were turned off (i.e., the α values were set to zero, and the parameters T_L and T_H were set to infinity). However, as stated in the Legend to **Figure 1**, the effects obtained when these were enabled were generically the same with respect to the biased competition effects described in this research.

Simulation results for a system with $\lambda_1 = 6.0$, $\lambda_2 = 5.0$, and the top-down bias $\lambda_H = \lambda_2^{H,cr}$ are shown in **Figure 2**. $\lambda_2^{H,cr}$ is the value derived in the analysis for the top-down bias to just overcome the bottom-up inputs shown as λ_1 and λ_2 , as shown below in (5.54). $\lambda_2^{H,cr}$ was 22.816. Now the H_2 node has high activity as expected due to the application of λ_2^H , but this has the effect that the network settles into a state where the L_2 node has just higher activity than the L_1 node, despite $\lambda_1 > \lambda_2$. This demonstrates that the analysis described below that derives the critical value $\lambda_2^{H,cr}$ for the biased attention signal to just reverse the difference between the bottom up inputs to make the network respond preferentially to the weaker incoming signal, is accurate.

In further investigations, it was found that the analysis made correct predictions over a wide range of values of the parameters that were confirmed by numerical simulations. For example, the value $\lambda_2^{H,cr}$ was correctly calculated by the analysis over a 10-fold variation of K_f and K_b as shown by the performance of the numerical simulations. It was also found in the simulations that ratios for J_b/J_f and k_b/k_f in the range of 0.1–0.5 produced



top-down biased competition effects with values for $\lambda_2^{H,cr}$ that were in the range of the other activities in the system.

The interpretation of the analytical results shown in for example (5.54) is now facilitated by graphical analysis. **Figure 3** for Investigation 1 shows the top-down bias $\lambda_2^{H,cr}$ needed to make the activity of L_2 as large as L_1 as a function of $\lambda_1 - \lambda_2$ which is termed $\delta\lambda$. λ_1 was fixed at 6, and $\lambda_2^H = 0$. **Figure 3** shows that $\lambda_2^{H,cr}$ is a linear function of $\delta\lambda$. The slope of this function is high (140/6). The implication is that for reasonable values of the top-down bias limited to perhaps 50 in this model, the working range for $\delta\lambda$ is relatively small, approximately 2.

In another example to illustrate the utility of the analysis that produced (5.54), it was found that the top-down bias $\lambda_2^{H,cr}$ needed to make the activity of L_2 as large as L_1 was relatively independent of the absolute values of λ_1 and λ_2 , and so depended on the difference between them, i.e., $\lambda_1 - \lambda_2$ termed $\delta\lambda$. This is illustrated in **Figure 4** in which $\delta\lambda$ was set to 1.0, and the value of λ_1 was increased from 1 to 10. It is clear that the top-down bias required depends very little on the absolute values of λ_1 and λ_2 , but instead on their difference as shown in **Figure 3**.

We can also understand quantitatively the effect of the other parameters using (5.54). For example, if J_b is increased from its default value of 0.05/3 to 0.1/3, then the slope of the function illustrated in **Figure 3** falls to a lower value (66/6). This makes the important point that top-down biased competition can only operate with rates for the top-down bias (in this case λ_2^H) in a reasonable range (not too high) if the backward connection strength (here J_b with a default value of 0.05/3) is not too low compared to the forward connection strength (here J_f with a default value of 0.13/3). This places important constraints on the ratio of the strengths of backprojections relative to forward projections between cortical areas (Rolls, 2016).

In another example to illustrate the utility of the analysis that produced (5.54), it was found that if the crossed connections K_b (default 0.005/3) were increased (for example to 0.01/3), then more top-down bias was needed, with the slope of the function shown in **Figure 3** now 158/6. The reason for this is clear, that some of the top-down bias λ_2^H acts on λ_1 via K_b , but the analytic result in (5.54) quantifies this, as it does the effects of the other parameters involved in this approach to biased competition.

4.2. Investigation 2

The second top-down competition effect considered was as follows. If λ_1 is greater than λ_2 and both are applied

simultaneously, then the activity of H_1 will be greater than the activity of H_2 . However, if we apply top down biased competition λ_H to H_2 , then we can influence the rates H_2 and H_1 (through all the connections in the system) until H_2 and H_1 have the same activity (and at the same time there will be an effect on L_1 and L_2). We wished to quantify these effects analytically in terms of all the parameters of the system.

Simulation results for a system with $\lambda_1 = 6.0$, $\lambda_2 = 5.0$, and the top-down bias $\lambda_2^H = 0$ or $\lambda_2^{H,cr}$ are shown in **Figure 5**. $\lambda_2^{H,cr}$ is the value derived in the analysis for the top-down bias to make the rates in H_1 equal those in H_2 , to just overcome the difference between the bottom-up inputs λ_1 and λ_2 , as shown in (5.60). The upper graph in **Figure 5** shows the results when the top-down bias $\lambda_H = 0$. It can be seen that the rates in H_1 are higher than in H_2 , which is as expected, because $\lambda_1 = 6.0$, and $\lambda_2 = 5.0$. In the lower part of **Figure 5** the top-down bias $\lambda_2^H = 0.775$, the value derived in the analysis for the top-down bias to just produce equal rates in H_1 and H_2 , overcoming the difference between the bottom-up inputs shown as λ_1 and λ_2 . The simulation illustrated in the lower part of the figure thus shows that when the analytically calculated value for $\lambda_2^{H,cr}$ is used, the numerical simulation confirms that this is the correct value. When a different value is used, as shown in the top part of **Figure 5**, then the correct results are not obtained.

The interpretation of the analytical results shown in for example (5.60) for Investigation 2 is now facilitated by graphical analysis. Analytical results for Investigation 2 to show how $\lambda_2^{H,cr}$ is a function of λ_1 are shown in **Figure 6**, for a system with $\lambda_2 = 5.0$. $\lambda_2^{H,cr}$ is the value derived in the analysis for the top-down bias to just produce equal rates in H_1 and H_2 , overcoming the difference between the bottom-up inputs shown as λ_1 and λ_2 . **Figure 6** shows that $\lambda_2^{H,cr}$ is a linear function of λ_1 , when $\lambda_2 = 5.0$. Moreover, **Figure 6** shows that only a small variation of $\lambda_2^{H,cr}$ is sufficient to counteract large changes in λ_1 . Moreover, the implication of 5.60 is that provided that the conditions shown in (5.59) are met, the operation is relatively independent of λ_2 . The understanding to which this leads is that the relative outputs measured at the H nodes are relatively little affected by the values of λ_1 and λ_2 , compared to the effects of the input biases to λ_2^H and λ_1^H . An implication for the operation of the brain is that top-down biased competition can have useful effects on the lower (L) nodes in the system, which could then influence other systems. Another implication is that the output from the higher (H) nodes is relatively strongly affected by any direct inputs to the H nodes, compared to effects mediated by top-down biases acting through the backward connections to the L nodes, and on systems connected to these L nodes.

Analytical results for Investigation 2 to show how $\lambda_2^{H,cr}$ depends on λ_1^H are shown in **Figure 7**, for a system with $\lambda_2 = 5.0$. $\lambda_2^{H,cr}$ is the value derived in the analysis for the top-down bias to just produce equal rates in H_1 and H_2 , overcoming the difference between the bottom-up inputs shown as λ_1 and λ_2 . This figure shows that $\lambda_2^{H,cr}$ is a linear function of λ_1^H with a slope of approximately 1. The implication here is that inputs to the H nodes influence each other almost equally, and this will occur primarily through the inhibition between these

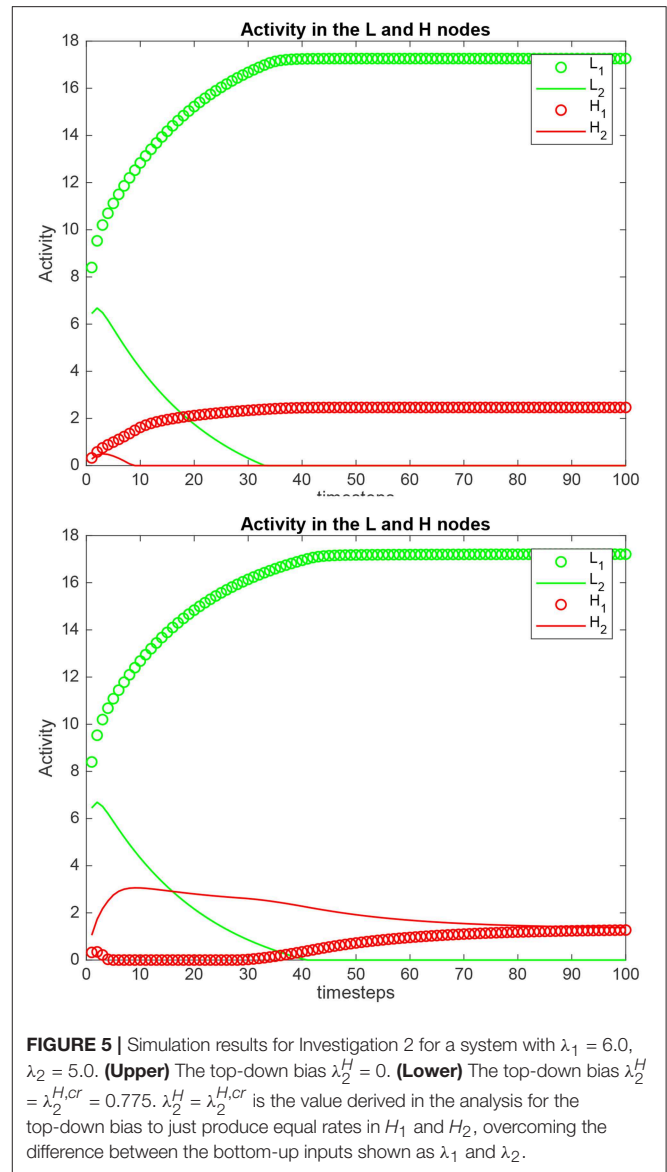


FIGURE 5 | Simulation results for Investigation 2 for a system with $\lambda_1 = 6.0$, $\lambda_2 = 5.0$. **(Upper)** The top-down bias $\lambda_2^H = 0$. **(Lower)** The top-down bias $\lambda_2^H = \lambda_2^{H,cr} = 0.775$. $\lambda_2^H = \lambda_2^{H,cr}$ is the value derived in the analysis for the top-down bias to just produce equal rates in H_1 and H_2 , overcoming the difference between the bottom-up inputs shown as λ_1 and λ_2 .

nodes implemented by c_H , rather than through the top-down connections to the L nodes, and then the return effects from the L to the H nodes.

5. MATHEMATICAL ANALYSIS

To solve the named problems we shall study the system (3.1) and (3.2) which describes the dynamics of $(L_1(t), L_2(t), H_1(t), H_2(t))$ in R_+^4 . Note that the entire state space $R_+^4 = \{(x_1, x_2, y_1, y_2) : x_i \geq 0, y_i \geq 0\}$ is decomposed into $2^4 = 16$ areas depending on whether $x_i \in [0, T_L]$, $y_i \in [0, T_H]$:

$$R_+^4 = \cup_{(e_1, \dots, e_4) : e_i \in \{0,1\}} X_{e_1} \times X_{e_2} \times Y_{e_3} \times Y_{e_4}, \quad (5.7)$$

where

$$X_0 = [0, T_L], X_1 = [T_L, \infty), Y_0 = [0, T_H], Y_1 = [T_H, \infty).$$

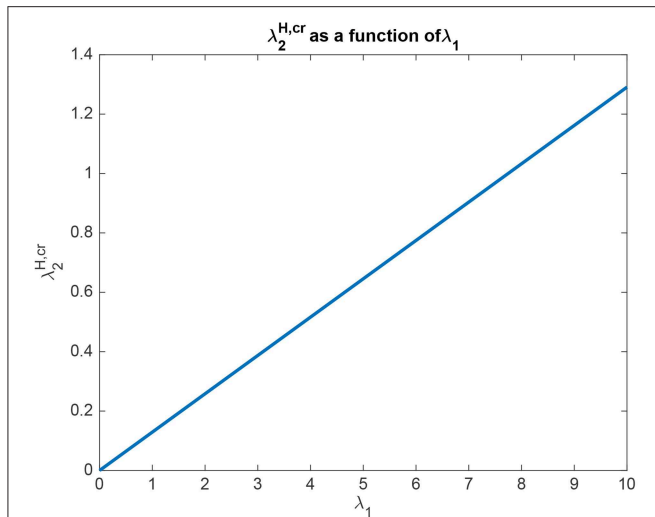


FIGURE 6 | Analytical results for Investigation 2 for a system with $\lambda_2 = 5.0$. $\lambda_2^{H,cr}$ is shown as a function of λ_1 . $\lambda_2^{H,cr}$ is the value derived in the analysis for the top-down bias to just produce equal rates in H_1 and H_2 , overcoming the difference between the bottom-up inputs shown as λ_1 and λ_2 .

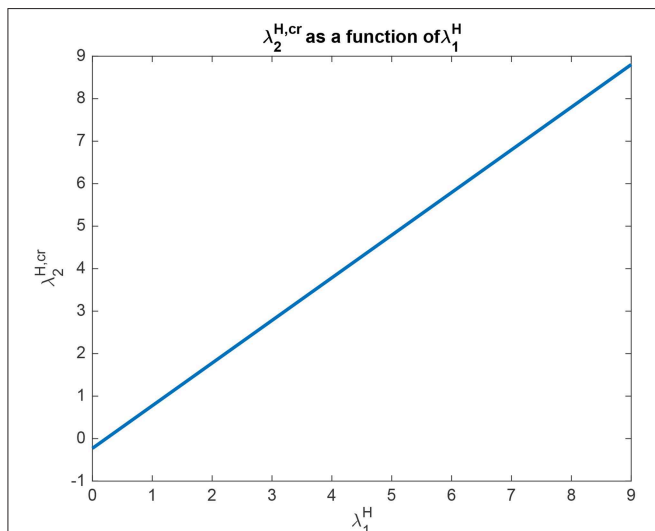


FIGURE 7 | Analytical results for Investigation 2 for a system with $\lambda_1 = 6.0$, $\lambda_2 = 5.0$. $\lambda_2^{H,cr}$ is shown as a function of λ_1^H . $\lambda_2^{H,cr}$ is the value derived in the analysis for the top-down bias to just produce equal rates in H_1 and H_2 , overcoming the difference between the bottom-up inputs shown as λ_1 and λ_2 .

In each area $X_{e_1} \times X_{e_2} \times Y_{e_3} \times Y_{e_4}$ the behavior of a system (3.1), (3.2) is linear; the non-linear nature of this system is seen, roughly speaking, only at the borders of these areas. In particular, there is a “cut” or “threshold” at zero for all involved functions, as their meaning as rates assumes only non-negative values.

We are mostly interested in modeling dynamics of a vector $(L_1(t), L_2(t), H_1(t), H_2(t))$ which does not escape to infinity in any coordinate. Therefore, we shall first study the system in the area

$$S := \{(x_1, x_2, y_1, y_2) : x_i \in [0, T_L], y_i \in [0, T_H], i = 1, 2\} \\ = X_0 \times X_0 \times Y_0 \times Y_0. \tag{5.8}$$

Let us assume first the following simplifying assumption. Set

$$T_L = T_H = \infty, \tag{5.9}$$

which means that there is no facilitation of rates in the network.

Consider the linear system associated with the system (3.1), (3.2):

$$\begin{pmatrix} \tilde{L}_1(t+1) \\ \tilde{L}_2(t+1) \\ \tilde{H}_1(t+1) \\ \tilde{H}_2(t+1) \end{pmatrix} = \begin{pmatrix} 1 - \beta_L & -c_L & J_b & K_b \\ -c_L & 1 - \beta_L & K_b & J_b \\ J_f & K_f & 1 - \beta_H & -c_H \\ K_f & J_f & -c_H & 1 - \beta_H \end{pmatrix} \begin{pmatrix} \tilde{L}_1(t) \\ \tilde{L}_2(t) \\ \tilde{H}_1(t) \\ \tilde{H}_2(t) \end{pmatrix} + \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_1^H \\ \lambda_2^H \end{pmatrix}. \tag{5.10}$$

Let us fix

$$\bar{\lambda} \in S. \tag{5.11}$$

Then assuming also as in (3.5) the zero initial conditions

$$(\tilde{L}_1(0), \tilde{L}_2(0), \tilde{H}_1(0), \tilde{H}_2(0)) = (0, 0, 0, 0) = (L_1(0), L_2(0), H_1(0), H_2(0)),$$

we observe that as long as

$$\tilde{L}_i(t) \in [0, T_L], \tilde{H}_i(t) \in [0, T_H], i = 1, 2, \tag{5.12}$$

i.e., $\tilde{L}_i(t) \geq 0$ and $\tilde{H}_i(t) \geq 0$ [recall assumption (5.9)], the system (5.10) describes exactly the same system as in (3.1) and (3.2), i.e.,

$$\tilde{L}_i(t) = L_i(t), \tilde{H}_i(t) = H_i(t). \tag{5.13}$$

Therefore we first derive the conditions for the matrix in (5.10) under which relations (5.12) hold for all $t \geq 0$, i.e., the system remains to be in the bounded area S .

The boundedness of the solution to (5.10) is defined entirely by the eigenvalues of the corresponding matrix. However, deriving the eigenvalues even for a (4×4) -matrix requires already heavy computations. Instead we shall take advantage of some particular properties and symmetries of our model which allow us to reduce the original 4 dimensions of the system.

5.1. Boundedness of the Trajectories

Let us consider the dynamics of two systems related to (5.10):

$$\mathbf{L}(t) = \tilde{L}_1(t) + \tilde{L}_2(t), \quad \mathbf{H}(t) = \tilde{H}_1(t) + \tilde{H}_2(t), \tag{5.14}$$

and

$$\Delta L(t) = \tilde{L}_1(t) - \tilde{L}_2(t), \quad \Delta H(t) = \tilde{H}_1(t) - \tilde{H}_2(t). \tag{5.15}$$

Observe that $\mathbf{H}(t) \rightarrow \infty$ or $\Delta H(t) \rightarrow \infty$ if and only if at least one $\tilde{H}_i(t) \rightarrow \infty$, in which case the original system $(L_1(t), L_2(t), H_1(t), H_2(t))$ cannot remain in the bounded area S .

A similar statement holds about $\mathbf{L}(t)$ and $\Delta L(t)$. Therefore, we begin the analysis by providing the necessary conditions for the stability of systems (5.14) and (5.15), and thus for the stability of (5.10).

Then assuming that the connection parameters provide stability for the sums (5.14) we investigate the behavior of the system of differences (5.15) for different inputs

$$\lambda_1 > \lambda_2 \text{ and } \lambda_1^H < \lambda_2^H.$$

Our goal here is to find parameters such that given asymmetric inputs $\lambda_1 > \lambda_2$ we want to find $\lambda_1^H < \lambda_2^H$ such that $\Delta L(t)$ converges to a negative value for large t or that $\Delta H(t)$ converges to a negative value for large t .

If the equality (5.13) still holds for all t this means that correspondingly, $L_1(t) < L_2(t)$ or $H_1(t) < H_2(t)$ for large t as well. Hence, correspondingly, we have a solution to Problem I or Problem II.

First we derive from (5.10):

$$\begin{pmatrix} \mathbf{L}(t+1) \\ \mathbf{H}(t+1) \end{pmatrix} = W_S \begin{pmatrix} \mathbf{L}(t) \\ \mathbf{H}(t) \end{pmatrix} + \begin{pmatrix} \lambda \\ \lambda^H \end{pmatrix}, \tag{5.16}$$

where

$$W_S = \begin{pmatrix} 1 - \beta_L - c_L & J_b + K_b \\ J_f + K_f & 1 - \beta_H - c_H \end{pmatrix}, \tag{5.17}$$

and

$$\lambda = \lambda_1 + \lambda_2, \quad \lambda^H = \lambda_1^H + \lambda_2^H.$$

Similarly,

$$\begin{pmatrix} \Delta L(t+1) \\ \Delta H(t+1) \end{pmatrix} = W_\Delta \begin{pmatrix} \Delta L(t) \\ \Delta H(t) \end{pmatrix} + \begin{pmatrix} \Delta \lambda \\ \Delta \lambda^H \end{pmatrix}, \tag{5.18}$$

where

$$W_\Delta = \begin{pmatrix} 1 - \beta_L + c_L & J_b - K_b \\ J_f - K_f & 1 - \beta_H + c_H \end{pmatrix}, \tag{5.19}$$

and

$$\Delta \lambda = \lambda_1 - \lambda_2, \quad \Delta \lambda^H = \lambda_1^H - \lambda_2^H.$$

Effectively, using the symmetries we reduced the 4 dimensions of our model to 2 dimensions, as the solution to our original system (5.10) is given by

$$\tilde{L}_1(t) = \frac{1}{2}(\mathbf{L}(t) + \Delta L(t)), \quad \tilde{L}_2(t) = \frac{1}{2}(\mathbf{L}(t) - \Delta L(t)), \tag{5.20}$$

and

$$\tilde{H}_1(t) = \frac{1}{2}(\mathbf{H}(t) + \Delta H(t)), \quad \tilde{H}_2(t) = \frac{1}{2}(\mathbf{H}(t) - \Delta H(t)). \tag{5.21}$$

Recall that for a 2x2 matrix

$$M = \begin{pmatrix} A & C \\ D & B \end{pmatrix}, \tag{5.22}$$

where all the entries are positive (as in our model) the eigenvalues are given by

$$\kappa_1 = \frac{A + B + \sqrt{(A - B)^2 + 4CD}}{2}, \tag{5.23}$$

and

$$\kappa_2 = \frac{A + B - \sqrt{(A - B)^2 + 4CD}}{2}. \tag{5.24}$$

Note that due to the assumption of positivity of entries we have here

$$\kappa_2 < \kappa_1, \quad |\kappa_2| < \kappa_1. \tag{5.25}$$

Hence, we have two (not linearly dependent) eigenvectors

$$E_i = \begin{pmatrix} 1 \\ \frac{\kappa_i - A}{C} \end{pmatrix}. \tag{5.26}$$

Note here for further reference that the inequalities

$$\kappa_1 - A = \frac{B - A + \sqrt{(A - B)^2 + 4CD}}{2} > 0, \tag{5.27}$$

and

$$\kappa_2 - A = \frac{B - A - \sqrt{(A - B)^2 + 4CD}}{2} < 0 \tag{5.28}$$

hold for all positive parameters A, B, C, D .

Let us consider a linear transformation in R^2 :

$$\bar{X}(t+1) = M\bar{X}(t) + \bar{y}, \tag{5.29}$$

where $\bar{X}(0) = (0, 0)$ and $\bar{y} = (y_1, y_2)$ is any vector.

Since vectors E_1 and E_2 make a basis in R^2 , for any vector \bar{y} there are numbers q_1 and q_2 , such that

$$\bar{y} = q_1 E_1 + q_2 E_2 = \begin{pmatrix} q_1 + q_2 \\ \frac{\kappa_1 - A}{C} q_1 + \frac{\kappa_2 - A}{C} q_2 \end{pmatrix}. \tag{5.30}$$

Then the solution to (5.29) is given by

$$\bar{X}(t+1) = \sum_{k=0}^t M^k \bar{y} = \sum_{k=0}^t \kappa_1^k q_1 E_1 + \sum_{k=0}^t \kappa_2^k q_2 E_2. \tag{5.31}$$

Observe that

$$\kappa_1 < 1 \text{ if and only if } \begin{cases} A + B < 2, \\ CD < (1 - A)(1 - B). \end{cases} \tag{5.32}$$

The last system together with the assumption that both C and D are positive yields

$$\kappa_1 < 1 \text{ if and only if } \begin{cases} 0 < A, B < 1, \\ CD < (1 - A)(1 - B). \end{cases} \tag{5.33}$$

Under the last condition equation (5.31) yields the following convergence to the fixed point:

$$\bar{X}(t) \rightarrow \frac{1}{1-\kappa_1}q_1E_1 + \frac{1}{1-\kappa_2}q_2E_2 = \begin{pmatrix} \frac{q_1}{1-\kappa_1} + \frac{q_2}{1-\kappa_2} \\ \frac{\kappa_1-A}{C(1-\kappa_1)}q_1 + \frac{\kappa_2-A}{C(1-\kappa_2)}q_2 \end{pmatrix}, \tag{5.34}$$

as $t \rightarrow \infty$.

Consider W_S defined in (5.17) as an example of the generic matrix M in (5.22). Let us denote its eigenvalues $\kappa_i(W_S)$, $i = 1, 2$, $\kappa_1(W_S) \geq \kappa_2(W_S)$. Under assumption

$$\beta_L + c_L < 1, \quad \beta_H + c_H < 1, \tag{5.35}$$

we have according to (5.33):

$$\kappa_1(W_S) < 1 \text{ if and only if } (J_f + K_f)(J_b + K_b) < (\beta_L + c_L)(\beta_H + c_H). \tag{5.36}$$

Next we study the eigenvalues $\kappa_1(W_\Delta) \geq \kappa_2(W_\Delta)$ of W_Δ . Under assumption

$$\beta_L - c_L < 1, \quad \beta_H - c_H < 1, \tag{5.37}$$

which holds whenever (5.35) holds, we have according to (5.33):

$$\kappa_1(W_\Delta) < 1 \text{ if and only if } (J_f - K_f)(J_b - K_b) < (\beta_L - c_L)(\beta_H - c_H). \tag{5.38}$$

Observe that when $c_L = c_H = 0$, condition (5.38) follows by the above condition (5.36).

Notice that together with the assumption (3.3) when all the parameters except J_b, K_b are fixed, the ratio $q = J_b/J_f$ has to be small to ensure the boundedness of the functions $L_i(t), H_i(t)$.

In words, conditions (5.36) and (5.38) tell us that under the assumption (5.35) the stability of the system (or boundedness of the trajectories) requires that at least some of the connections, forward or backward ones, have to be sufficiently small.

5.2. Dynamics of $(\Delta L(t), \Delta H(t))$ for a Biased Input When All Rates Remain to be Strictly Positive

Here we study the system (3.1)–(3.2) assuming conditions when the functions $L_i(t), H_i(t)$ remain to be strictly positive and bounded. This means that the dynamics is described by linear system.

Assume that initially the input to \tilde{L}_1 is greater than the one to \tilde{L}_2 , i.e., $\Delta\lambda > 0$. We shall find here the sufficient conditions on the parameters of the connections which yield existence of the values

$$\Delta\lambda^H < 0$$

such that eventually, contrary to the initial bias, the state of the system satisfies

$$\tilde{L}_2(t) < \tilde{L}_1(t), \tag{5.39}$$

for all large t .

We shall assume first of all that the above conditions (5.35), (5.36), (5.37), and (5.38) are satisfied.

Consider system (5.19) with input

$$\begin{pmatrix} \Delta\lambda \\ \Delta\lambda^H \end{pmatrix} = \begin{pmatrix} \lambda_1 - \lambda_2 \\ \lambda_1^H - \lambda_2^H \end{pmatrix}.$$

Let us decompose this vector along the eigenvectors of W_Δ as in (5.30):

$$\begin{pmatrix} \Delta\lambda \\ \Delta\lambda^H \end{pmatrix} = xE_1(W_\Delta) + yE_2(W_\Delta) = \begin{pmatrix} x + y \\ \frac{\kappa_1(W_\Delta) - A_\Delta}{C_\Delta}x + \frac{\kappa_2(W_\Delta) - A_\Delta}{C_\Delta}y \end{pmatrix}, \tag{5.40}$$

where $E_i(W_\Delta)$ and $\kappa_i(W_\Delta)$, $i = 1, 2$, denote, correspondingly the eigenvectors and the eigenvalues of W_Δ , and here [compare matrices (5.19) and (5.22)]

$$C_\Delta = J_b - K_b, \quad A_\Delta = 1 - \beta_L + c_L.$$

Then by (5.34) the following convergence takes place when $t \rightarrow \infty$:

$$\begin{pmatrix} \Delta L(t) \\ \Delta H(t) \end{pmatrix} \rightarrow \begin{pmatrix} \frac{x}{1-\kappa_1(W_\Delta)} + \frac{y}{1-\kappa_2(W_\Delta)} \\ \frac{\kappa_1(W_\Delta) - A_\Delta}{C_\Delta(1-\kappa_1(W_\Delta))}x + \frac{\kappa_2(W_\Delta) - A_\Delta}{C_\Delta(1-\kappa_2(W_\Delta))}y \end{pmatrix}. \tag{5.41}$$

Hence, given a positive

$$\Delta\lambda = \lambda_1 - \lambda_2 = x + y$$

we want to find value $x = x(\Delta\lambda)$ which satisfies condition

$$\frac{x}{1-\kappa_1} + \frac{\Delta\lambda - x}{1-\kappa_2} < 0 \tag{5.42}$$

and moreover minimizes function [see (5.40)]

$$F(x) = \left| \frac{\kappa_1 - A_\Delta}{C_\Delta}x + \frac{\kappa_2 - A_\Delta}{C_\Delta}(\Delta\lambda - x) \right|. \tag{5.43}$$

From (5.42) we get

$$x < -\Delta\lambda \frac{1-\kappa_1}{\kappa_1 - \kappa_2}, \tag{5.44}$$

which is negative. Then for all the x which satisfy (5.44) we have

$$\begin{aligned} F(x) &= \frac{1}{C_\Delta} |(\kappa_1 - \kappa_2)x - (A_\Delta - \kappa_2)\Delta\lambda| \\ &= \frac{(\kappa_1 - \kappa_2)|x| + (A_\Delta - \kappa_2)\Delta\lambda}{C_\Delta} > \Delta\lambda \frac{1 + A_\Delta - (\kappa_1 + \kappa_2)}{C_\Delta}. \end{aligned} \tag{5.45}$$

Finally, we can define a negative $\Delta\lambda^H$ such that (5.40) holds, and moreover the limiting state as defined in the first row on

the right in (5.41) satisfies (5.42). Observe that condition (5.42) by (5.41) guarantees condition (5.39). Substituting previously derived formulas into (5.45) we derive that for any given positive $\Delta\lambda$ and any $\lambda_1^H < \lambda_2^H$ such that

$$\lambda_2^H - \lambda_1^H > \frac{\beta_H - c_H}{J_b - K_b} \Delta\lambda \tag{5.46}$$

we have (5.39).

We conclude that in the case when (5.13) still holds for all large t , i.e., if the system remains to be in the positive area, we have $L_2(t) \geq L_1(t)$ for all large t if conditions (5.35), (5.36), and (5.38) are fulfilled and λ_2^H is greater than the following critical value:

$$\lambda_2^{H,cr} = \lambda_1^H + \frac{\beta_H - c_H}{J_b - K_b} (\lambda_1 - \lambda_2). \tag{5.47}$$

5.3. Operation With a Threshold at a Rate of Zero

Here we explore the non-linear effects of the system (3.1)–(3.2) considering the case when some of the rates $L_i(t)$ or $H_i(t)$ become zero and may stay at zero due to the non-linear threshold function that does not allow negative rates $(\cdot)^+$.

5.3.1. Equality in the L-Nodes

We shall find here the conditions when a state

$$L_1(t) = L_2(t) = L, H_1(t) = 0, H_2(t) = H \tag{5.48}$$

with strictly positive L and H can be a fixed point for the dynamical system (3.1)–(3.2) under assumption that

$$\lambda_1 > \lambda_2, \lambda_2^H > \lambda_1^H$$

and also (5.9) holds, i.e., when a facilitation of rate above a certain threshold is not applied.

Assuming that the trajectories of $L_1(t)$, $L_2(t)$, and $H_2(t)$ remain to be strictly positive and do not hit zero, while $H_1(t)$ does decay to zero, the constants in (5.48) should satisfy the following system derived from (5.10) [see also 3.1)–(3.2)]:

$$\begin{cases} -\beta_L L - c_L L + K_b H + \lambda_1 = 0, \\ -c_L L - \beta_L L + J_b H + \lambda_2 = 0, \\ J_f L + K_f L - c_H H + \lambda_1^H \leq 0, \\ K_f L + J_f L - \beta_H H + \lambda_2^H = 0. \end{cases} \tag{5.49}$$

Observe that the inequality in the third line in (5.49) after the threshold at a rate of zero as in the original system (3.1)–(3.2) yields limiting state $H_1(t) = 0$.

The system (5.49) is equivalent to

$$\begin{cases} 0 < H = \frac{\lambda_1 - \lambda_2}{J_b - K_b} \leq \frac{\lambda_2^H - \lambda_1^H}{\beta_H - c_H}, \\ L = \frac{J_b H + \lambda_2}{\beta_L + c_L} = \frac{\beta_H H - \lambda_2^H}{K_f + J_f}. \end{cases} \tag{5.50}$$

This requires the following conditions for the parameters in order for the last system to have a solution:

$$\frac{\lambda_2(K_f + J_f) + \left(\frac{\lambda_1 - \lambda_2}{J_b - K_b}(\beta_H - c_H) + \lambda_1^H\right)(\beta_L + c_L)}{\beta_H(\beta_L + c_L) - J_b(K_f + J_f)} \leq H \tag{5.51}$$

$$= \frac{\lambda_2(K_f + J_f) + \lambda_2^H(\beta_L + c_L)}{\beta_H(\beta_L + c_L) - J_b(K_f + J_f)} = \frac{\lambda_1 - \lambda_2}{J_b - K_b},$$

which in turn requires

$$\frac{\lambda_2(K_f + J_f)(J_b - K_b) + ((\lambda_1 - \lambda_2)(\beta_H - c_H) + \lambda_1^H(J_b - K_b))(\beta_L + c_L)}{\beta_H(\beta_L + c_L) - J_b(K_f + J_f)} \leq \lambda_1 - \lambda_2. \tag{5.52}$$

Observe that when $\lambda_1^H = 0$ the last condition is equivalent to

$$(\lambda_1 J_b - \lambda_2 K_b)(K_f + J_f) \leq (\lambda_1 - \lambda_2)c_H(\beta_L + c_L). \tag{5.53}$$

Assuming that (5.52) holds we derive from (5.51) the following critical value

$$\lambda_2^{H,cr} = \frac{\lambda_1 - \lambda_2}{J_b - K_b} \left(\beta_H - J_b \frac{K_f + J_f}{\beta_L + c_L} \right) - \lambda_2 \frac{K_f + J_f}{\beta_L + c_L}. \tag{5.54}$$

The above analysis yields the following statement.

Proposition 5.1. *Let conditions (5.36) and (5.38) hold, and let*

$$\lambda_2^H \geq \lambda_2^{H,cr}.$$

Assume also (5.52) (or (5.53) if $\lambda_1^H = 0$) and

$$\frac{\lambda_2(K_f + J_f) + \lambda_2^H(\beta_L + c_L)}{\beta_H(\beta_L + c_L) - J_b(K_f + J_f)} = \frac{\lambda_1 - \lambda_2}{J_b - K_b}. \tag{5.55}$$

Then the system (3.1)–(3.2) converges to a state where

$$H_1(t) = 0, H_2(t) = H = \frac{\lambda_1 - \lambda_2}{J_b - K_b}, L_1(t) = L_2(t) = \frac{J_b H + \lambda_2}{\beta_L + c_L}.$$

Notice that the limiting state described in the last Proposition satisfies (5.48).

Observe that the formula (5.54) for the critical value $\lambda_2^{H,cr}$ is in a good agreement with the previous case (5.47); in fact the same condition as in (5.47) reads directly from the inequality in (5.50). However, this is precisely the non-linearity of the system that we use here to derive the exact formula (5.54) for the critical value.

5.3.2. Equality in the H-Nodes

We shall find here the conditions when a state

$$L_1(t) = L, L_2(t) = 0, H_1(t) = H_2(t) = H \tag{5.56}$$

with strictly positive L and H can be a fixed point for the dynamical system (3.1)–(3.2) under the assumption that

$$\lambda_1 > \lambda_2, \lambda_2^H > \lambda_1^H$$

and also (5.9) holds, i.e., when a facilitation of rate above a certain threshold is not applied.

Assuming that the trajectories of $L_1(t)$, $H_1(t)$, and $H_2(t)$ remain to be strictly positive and do not hit zero, while $L_2(t)$ does

decay to zero, the constants in (5.56) should satisfy the following system derived from (5.10) [see also 3.1)–(3.2)]:

$$\begin{cases} -\beta_L L + (J_b + K_b)H + \lambda_1 = 0, \\ -c_L L + (J_b + K_b)H + \lambda_2 \leq 0, \\ J_f L - (\beta_H + c_H)H + \lambda_1^H = 0, \\ K_f L - (\beta_H + c_H)H + \lambda_2^H = 0. \end{cases} \quad (5.57)$$

Observe that the inequality in the second line in (5.57) after the threshold at zero as in the original system (3.1)–(3.2) yields limiting state $L_2(t) = 0$. The system (5.57) is equivalent to

$$\begin{cases} 0 < L = \frac{\lambda_2^H - \lambda_1^H}{J_f - K_f} \leq \frac{\lambda_1 - \lambda_2}{\beta_L - c_L}, \\ H = \frac{K_f L + \lambda_2^H}{\beta_H + c_H} = \frac{\beta_L L - \lambda_1}{K_b + J_b}. \end{cases} \quad (5.58)$$

First we derive the conditions for the parameters in order for the last system to have a solution:

$$\lambda_1 \left(c_L - J_f \frac{K_b + J_b}{\beta_H + c_H} \right) \geq \lambda_1^H \frac{K_b + J_b}{\beta_H + c_H} (\beta_L - c_L) + \lambda_2 \left(\beta_L - J_f \frac{K_b + J_b}{\beta_H + c_H} \right). \quad (5.59)$$

Assuming that the latter holds we derive from (5.57) the following critical value

$$\lambda_2^{H,cr} = \frac{\lambda_1 (J_f - K_f) (\beta_H + c_H) + \lambda_1^H ((\beta_H + c_H)\beta_L - K_f(K_b + J_b))}{(\beta_H + c_H)\beta_L - (J_b + K_b)J_f}. \quad (5.60)$$

If $\lambda_2^H \geq \lambda_2^{H,cr}$ then the system ends up in a state where $L_1(t) = L > L_2(t) = 0$ and $H_2(t) = H_1(t) = H$.

6. DISCUSSION

6.1. The Analysis

We consider here a 4-dimensional system of linear equations with thresholds for the rates at zero. Although it is “almost” a linear system, which admits rather straightforward analysis, the focus is on the relations between the numerous parameters. Notably, the latter relations are mostly non-linear.

Our approach takes advantage of the authentic symmetries in the system which allowed us to reduce the original 4 dimensions to 2 dimensions. This method may have some interest on its own as it can be used in other similar situations.

The derived conditions for the parameters which yield certain desired properties (specified as Problems I and II) disclose non-trivial relations between the parameters.

We considered several cases: (i) when the system keeps a positive rate at each node (section 5.2), and (ii) when the rate at one node, namely $L_2(t)$ (section 5.3.2), or $H_2(t)$ (section 5.3.1) is suppressed to zero after a long enough time.

Consider first our solution to Problem I. Remarkably the formulas for the critical value of the bias λ_2^H to yield the success of competition are different for the above two cases, namely when

all the rates are strictly positive or when one rate is zero. These formulas are given by (5.47) and (5.54). This level of accuracy would not be possible without analytic formulas. Observe that (5.47) is the first condition in system (5.50), which also has to be fulfilled for the formula (5.54) to work.

As we mentioned above, formula (5.47) holds only under assumption that all the rates remain to be strictly positive, that can be achieved, for example, by choosing λ_1^H sufficiently large.

Further, we notice that the solution to Problem II provided by formula (5.60) reveals an interesting relation: as long as λ_2 satisfies condition (5.59) it does not enter directly formula (5.60).

Deco and Rolls (2005) inferred from their particular simulations that the ratio of 2.5 between J_f and J_b provides a good working point for the biased competition. As we do not find any universal ratio between J_f and J_b in our analysis we conclude that the ratio 2.5 reflects particular scaling when the remaining parameters are fixed at certain values. On the other hand, our analysis tells us that the product $J_f J_b$ has to be sufficiently small for the boundedness of the trajectories. More precisely, (5.36) and (5.38) under assumption (3.3) require the following sufficient conditions for our analysis

$$q = \frac{J_b}{J_f} < \min \left\{ \frac{(\beta_L + c_L)(\beta_H + c_H)}{(J_f + K_f)^2}, \frac{(\beta_L - c_L)(\beta_H - c_H)}{(J_f - K_f)^2} \right\} \quad (6.61)$$

Furthermore, each of formulas (5.54) or (5.47) works under additional conditions as specified in the text. In particular, formula (5.54) requires (5.53), which is

$$q = \frac{J_b}{J_f} < c_H \frac{\lambda_1 - \lambda_2}{\lambda_1 J_f - \lambda_2 K_f} \frac{\beta_L + c_L}{K_f + J_f}. \quad (6.62)$$

A reasonably large set of parameters satisfies the above conditions, as shown by the computational results.

6.2. Implications for Understanding Biased Competition and the Interaction Between Neural Systems

The analysis elucidates some interesting properties of the biased competition system described. For example, the system is sensitive to the difference between λ_1 and λ_2 , with the amount of biased competition required to produce the biased competition effects described related to this difference, as shown by the analytical results leading to (5.54), and the results shown in **Figures 3, 4**. These analyses and results show that while the system tolerates a wide range for the absolute values of λ_1 and λ_2 , the difference between them $\delta\lambda$ must be relatively small for the values of the top-down bias $\lambda_2^{H,cr}$ to be within a reasonable range of activity values, which in the context of the simulations described here might be up to 50.

The mean value of the λ inputs on the other hand influences how high the rates are of the output neurons. Another feature revealed by the analysis is how the parameters can be set to achieve asymptotically stable performance.

The analysis has interesting implications for understanding the operation of the backprojections that are important in top-down biased competition mechanisms of attention. Equation

(5.54) and **Figure 3**, the associated results show that in order for the top-down critical value $\lambda_2^{H,cr}$ not to have to be too large, the backprojections J_b must not be too weak. At the same time, J_b must be less than J_f , so that perceptual bottom-up inputs can dominate neural processing, which must not be dominated by internally generated top-down signals. This leaves a relatively small region for J_b/J_f between perhaps < 1.0 and 0.3 . However, this ratio must be kept fairly low so that the two systems being coupled in this way can operate with separate attractors at the bottom (L) and top (H) ends (Renart et al., 1999a,b).

Another interesting property of this top-down biased competition system elucidated by the analysis is that the operation of the system, including the effects of the top-down bias λ_2^H , was influenced especially by the difference $\delta\lambda$ between λ_1 and λ_2 , rather than by their absolute value, as shown in (5.54) and illustrated in **Figures 3, 4**. This is similar to the operation of integrate-and-fire decision-making networks (Rolls and Deco, 2010; Deco et al., 2013; Rolls, 2016), with the similarity reflecting the way in which the competition between the nodes or attractors operates.

The key correspondence between the mathematical analysis and the numerical simulations is that the simulations show that the mathematical analysis very accurately specifies the exact value of the top-down bias that is needed. That is useful confirmation that the analysis accurately specifies the interactions between the parameters in the biased competition system. The simulations are additionally useful in illustrating the operation of the biased competition system investigated analytically.

In conclusion, the major advantage of the analytical approach brought to bear here for the first time on biased competition between cortical areas is that it discloses relations between all the parameters of the model, and helps to identify those values that yield the desired effect of biased competition. This task cannot be fulfilled purely by numerical simulations.

DATA AVAILABILITY

All datasets analyzed for this study are included in the manuscript and the supplementary files.

AUTHOR CONTRIBUTIONS

TT performed the mathematical analyses. ER performed the numerical simulations and assessments. ER and TT wrote the paper jointly.

ACKNOWLEDGMENTS

TT acknowledges support of a COSTNET EU grant for supporting the collaboration with Oxford. ER acknowledges support for the collaboration with Lund from a grant of the Swedish Academy of Science (VR). The authors thank Y. Kazanovich (Institute of Applied Mathematics, Russian Academy of Sciences) for helpful discussion and comments.

REFERENCES

- Battaglia, F., and Treves, A. (1998). Stable and rapid recurrent processing in realistic autoassociative memories. *Neural Comput.* 10, 431–450. doi: 10.1162/089976698300017827
- Chelazzi, L. (1998). Serial attention mechanisms in visual search: a critical look at the evidence. *Psychol. Res.* 62, 195–219. doi: 10.1007/s004260050051
- Chelazzi, L., Miller, E., Duncan, J. and Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature* 363, 345–347. doi: 10.1038/363345a0
- Corchs, S., and Deco, G. (2002). Large-scale neural model for visual attention: integration of experimental single cell and fMRI data. *Cereb. Cortex* 12, 339–348. doi: 10.1093/cercor/12.4.339
- Corchs, S., and Deco, G. (2004). Feature-based attention in human visual cortex: simulation of fMRI data. *Neuroimage* 21, 36–45. doi: 10.1016/j.neuroimage.2003.08.045
- Deco, G., and Lee, T. S. (2002). A unified model of spatial and object attention based on inter-cortical biased competition. *Neurocomputing* 44–46, 775–781. doi: 10.1016/S0925-2312(02)00471-X
- Deco, G., Pollatos, O., and Zihl, J. (2002). The time course of selective visual attention: theory and experiments. *Vision Res.* 42, 2925–2945. doi: 10.1016/S0042-6989(02)00358-9
- Deco, G., and Rolls, E. T. (2002). Object-based visual neglect: a computational hypothesis. *Eur. J. Neurosci.* 16, 1994–2000. doi: 10.1046/j.1460-9568.2002.02279.x
- Deco, G., and Rolls, E. T. (2003). Attention and working memory: a dynamical model of neuronal activity in the prefrontal cortex. *Eur. J. Neurosci.* 18, 2374–2390. doi: 10.1046/j.1460-9568.2003.02956.x
- Deco, G., and Rolls, E. T. (2004). A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Res.* 44, 621–644. doi: 10.1016/j.visres.2003.09.037
- Deco, G., and Rolls, E. T. (2005). Neurodynamics of biased competition and cooperation for attention: a model with spiking neurons. *J. Neurophysiol.* 94, 295–313. doi: 10.1152/jn.01095.2004
- Deco, G., Rolls, E. T., Albantakis, L. and Romo, R. (2013). Brain mechanisms for perceptual and reward-related decision-making. *Prog. Neurobiol.* 103, 194–213. doi: 10.1016/j.pneurobio.2012.01.010
- Deco, G., Rolls, E. T., and Horwitz, B. (2004). ‘What’ and ‘where’ in visual working memory: a computational neurodynamical perspective for integrating fMRI and single-neuron data. *J. Cogn. Neurosci.* 16, 683–701. doi: 10.1162/089892904323057380
- Deco, G., and Zihl, J. (2001). Top-down selective visual attention: a neurodynamical approach. *Vis. Cogn.* 8, 119–140. doi: 10.1080/13506280042000054
- Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222. doi: 10.1146/annurev.ne.18.030195.001205
- Duncan, J. (1996). “Cooperating brain systems in selective perception and action,” in *Attention and Performance XVI*, eds T. Inui and J. L. McClelland (Cambridge, MA: MIT Press), 549–578.
- Duncan, J., and Humphreys, G. (1989). Visual search and stimulus similarity. *Psychol. Rev.* 96, 433–458. doi: 10.1037//0033-295X.96.3.433
- Elston, G. N. (2002). Cortical heterogeneity: implications for visual processing and polysensory integration. *J. Neurocytol.* 31, 317–335. doi: 10.1023/A:1024182228103
- Elston, G. N., and Fujita, I. (2014). Pyramidal cell development: postnatal spinogenesis, dendritic growth, axon growth, and electrophysiology. *Front. Neuroanat.* 8:78. doi: 10.3389/fnana.2014.00078
- Grossberg, S. (1988). Non-linear neural networks: principles, mechanisms, and architectures. *Neural Netw.* 1, 17–61. doi: 10.1016/0893-6080(88)90021-4

- Heinke, D., Deco, G., Zihl, J., and Humphreys, G. (2002). A computational neuroscience account of visual neglect. *Neurocomputing* 44–46, 811–816. doi: 10.1016/S0925-2312(02)00477-0
- Kastner, S., De Weerd, P., Desimone, R., and Ungerleider, L. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science* 282, 108–111. doi: 10.1126/science.282.5386.108
- Kastner, S., Pinsk, M., De Weerd, P., Desimone, R., and Ungerleider, L. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron* 22, 751–761. doi: 10.1016/S0896-6273(00)80734-5
- Kesner, R. P., and Rolls, E. T. (2015). A computational theory of hippocampal function, and tests of the theory: new developments. *Neurosci. Biobehav. Rev.* 48, 92–147. doi: 10.1016/j.neubiorev.2014.11.009
- Miller, E. K., Gochin, P. M., and Gross, C. G. (1993). Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. *Brain Res.* 616, 25–29. doi: 10.1016/0006-8993(93)90187-R
- Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J. Neurophysiol.* 70, 909–919. doi: 10.1152/jn.1993.70.3.909
- Panzeri, S., Rolls, E. T., Battaglia, F., and Lavis, R. (2001). Speed of feedforward and recurrent processing in multilayer networks of integrate-and-fire neurons. *Netw. Comput. Neural Syst.* 12, 423–440. doi: 10.1088/0954-898X/12/4/302
- Renart, A., Moreno, R., Rocha, J., Parga, N., and Rolls, E. T. (2001). A model of the IT–PF network in object working memory which includes balanced persistent activity and tuned inhibition. *Neurocomputing* 38–40, 1525–1531. doi: 10.1016/S0925-2312(01)00548-3
- Renart, A., Parga, N., and Rolls, E. T. (1999a). Associative memory properties of multiple cortical modules. *Network* 10, 237–255. doi: 10.1088/0954-898X_10_3_303
- Renart, A., Parga, N., and Rolls, E. T. (1999b). Backprojections in the cerebral cortex: implications for memory storage. *Neural Comput.* 11, 1349–1388. doi: 10.1162/089976699300016278
- Renart, A., Parga, N., and Rolls, E. T. (2000). “A recurrent model of the interaction between the prefrontal cortex and inferior temporal cortex in delay memory tasks,” in *Advances in Neural Information Processing Systems*, Vol. 12, eds S. Solla, T. Leen, and K.-R. Mueller (Cambridge, MA: MIT Press), 171–177.
- Reynolds, J., and Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron* 24, 19–29. doi: 10.1016/S0896-6273(00)80819-3
- Reynolds, J. H., Chelazzi, L., and Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.* 19, 1736–1753. doi: 10.1523/JNEUROSCI.19-05-01736.1999
- Rolls, E. T. (2012). Invariant visual object and face recognition: neural and computational bases, and a model, VisNet. *Front. Comput. Neurosci.* 6:35. doi: 10.3389/fncom.2012.00035
- Rolls, E. T. (2016). *Cerebral Cortex: Principles of Operation*. Oxford: Oxford University Press.
- Rolls, E. T. (2018). The storage and recall of memories in the hippocampo-cortical system. *Cell Tissue Res.* 373, 577–604. doi: 10.1007/s00441-017-2744-3
- Rolls, E. T., Cheng, W., Gilson, M., Gong, W., Deco, G., Zac Lo, C.-Y., et al. (2019). Beyond the disconnectivity hypothesis of schizophrenia. *Cereb. Cortex.*
- Rolls, E. T., and Deco, G. (2002). *Computational Neuroscience of Vision*. Oxford: Oxford University Press.
- Rolls, E. T., and Deco, G. (2010). *The Noisy Brain: Stochastic Dynamics as a Principle of Brain Function*. Oxford: Oxford University Press.
- Szabo, M., Almeida, R., Deco, G., and Stetter, M. (2004). Cooperation and biased competition model can explain attentional filtering in the prefrontal cortex. *Eur. J. Neurosci.* 19, 1969–1977. doi: 10.1111/j.1460-9568.2004.03211.x
- Treves, A. (1993). Mean-field analysis of neuronal spike dynamics. *Network* 4, 259–284. doi: 10.1088/0954-898X/4/3/002
- Usher, M., and Niebur, E. (1996). Modelling the temporal dynamics of IT neurons in visual search: a mechanism for top-down selective attention. *J. Cogn. Neurosci.* 8, 311–327. doi: 10.1162/jocn.1996.8.4.311

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Turova and Rolls. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.