Understanding visual attention with RAGNAROC:
A Reflexive Attention Gradient through Neural AttRactOr Competition

Authors:

Brad Wyble

Chloe Callahan-Flintoft

Department of Psychology, Penn State University

Hui Chen

Department of Psychology, Zhejiang University

Toma Marinov

Department of Psychology, Penn State University

Aakash Sarkar

Department of Psychology, Boston University

Howard Bowman

Department of Psychology, University of Birmingham

Computing Department University of Kent

Abstract

A quintessential challenge for any perceptual system is the need to focus on task-relevant information without being blindsided by unexpected, yet important information. The human visual system incorporates several solutions to this challenge, one of which is a reflexive covert attention system that is rapidly responsive to both the physical salience and the task-relevance of new information. Decades of research have outlined the capabilities of such attentional effects using both behavioral and neural correlates. This paper presents a comprehensive model for explaining both kinds of data as the product of brief neural attractor states that are distributed across the visual hierarchy at the moment when attention is engaged at a location in the visual field. Such attractors emerge from a competitive attentional gradient distributed over a population of topographically organized neurons. The model helps to resolve key debates about the nature of reflexive attention, such as whether it is parallel or serial, and whether suppression effects are distributed in a spatial surround, or selectively at the location of distractors. Most importantly, the model develops a framework for understanding the neural mechanisms of visual attention, and links those mechanisms to observable behavioral and neural correlates such as accuracy, reaction time, and ERPs such as the N2pc and Pd. This last contribution is the most crucial for repairing the disconnect that exists between our understanding of behavioral and neural correlates of attention.

1 Introduction

A quintessential challenge for any perceptual system is the need to focus on task-relevant information without being blindsided by unexpected information that is also important. For example, a driver must be able to stop in response to an unexpected obstacle even while searching intensely for a specific landmark. Understanding how perception meets this ubiquitous challenge is crucial for understanding how the brain balances the prioritization of sensory information according to its relevance.

This challenge is matched by a multitude of attentional systems operating across different senses and time scales. For example in vision there are both overt and covert forms of attention and as cognitive scientists we are tasked with understanding the mechanisms of these attentional systems. Decades of research have provided a multitude of data types that define the properties of visual attention, such as accuracy, reaction time and Event Related Potentials (ERPs). These data have driven the development of many theories, but the great majority of those theories are linked to a small number of specific paradigms (e.g. a model of visual search, or a model of the attentional blink). Such models are a useful starting point, but their narrow focus makes it difficult to generalize across experimental paradigms, and also makes it easy to inadvertently overfit a theory to a specific kind of finding. Newell(1973) argued that instead of focusing on individual results as a way to attack or defend a theoretical edifice we can use that collection of results more productively if we build a more comprehensive model that addresses all of them. The approach used here is to build a model that is somewhere close to the algorithmic level of implementation (Marr 1982), which maximizes the number of empirical constraints that can be applied (Love 2015).

The model described here, termed RAGNAROC, which is short for ***R**eflexive **A**ttention **G**radient through **N**eural **A**tt**R**act**O**r **C**ompetition*, is intended as a theoretical framework for understanding the reflexive attentional system, and addresses data in different forms (e.g. accuracy, reaction time and EEG), and from different paradigms. The outcome of this approach is to provide (1) a computational implementation of reflexive attention, (2) an intuition for how different forms of data (e.g. accuracy, reaction time and EEG) emerge from a common set of mechanisms and (3) suggested resolutions for several ongoing debates in the literature by showing how one model is able to account for distinct patterns of data that are often taken as support for mutually incompatible sides of an ongoing debate.

For the reader who is less interested in the model itself, there is a section in the discussion that focuses on the lessons that have been learned through the construction of the model.

1.1 Reflexive Attention

In the broadest of strokes, attention is perhaps best summarized as privileging certain representations at the expense of others and this prioritization takes many forms. In terms of visual attention a distinction is drawn between *voluntary attention*, wherein volitional control mechanisms configure the spatial deployment of attention over a potentially extended period of time and *reflexive attention*, wherein the visual system reacts rapidly to stimulus onsets in order to attend them before the stimulus display changes or the eye moves (Jonides 1981; Muller & Rabbit 1992; Hopfinger & Mangun 1998). The term *reflexive* invokes an analogy with muscle reflexes that are deployed rapidly in response to a stimulus, and without waiting for slower, deliberative processes.

This reflexive form of attention presumably plays a key role in selecting important information for further processing when the eyes are making saccades frequently which means that stimuli are only available at any location on the retina for a brief period of time. It was originally suggested that this system is entirely automatic, and subjects are unable to avoid capture (Remington, Johnston, & Yantis 1992).but the attentional capture debate (Yantis 1996; Theeuwes 1991; Folk, Remington & Johnston, 1992; Woodman & Luck 1999; Egeth, Leonard & Leber 2011) has demonstrated that goals moderate how strongly stimuli can trigger or capture attention. Thus, when looking for a colored letter, any uniquely colored stimulus has an enhanced influence on attention. This task-based configuration is also responsive to categorical target signifiers such as letters among digits (Wyble, Potter, Bowman 2009; Nako, Wu & Eimer 2014), and superordinate concepts (e.g. "marine animal"; Wyble, Folk, Potter 2013). Likewise, neural data from the EEG reflect what are thought to be rapid attentional responses to task relevant colors (Eimer 1996); letters/words (Eimer 1996; Tan & Wyble 2015; Nako, Wu & Eimer 2014) and line drawings (Nako, Wu, Smith & Eimer 2014).

This form of attention is essentially a semi-autonomous decision making system. Goals can influence what kinds of stimuli will be able to trigger attention, but the actual decision to deploy attention is linked to the spatial and temporal onset of a physical stimulus. Therefore, this system provides a tight coupling between bottom-up and top-down determinants of attentional control.

Moreover, this form of covert attention is one of the earliest mechanisms involved in visual processing and presumably plays a key role in most aspects of daily vision, and should manifest in a broad variety of tasks under the umbrella of visual cognition research.

What we do not yet understand is how such attention would function at the level of neural mechanisms. Attention has been theorized to involve a combination of enhancing the processing of target stimuli (Eimer 1996), while also inhibiting distractors (Cepeda Cave & Bichot 1998, Gaspelin Leonard & Luck 2015). However, while it seems straightforward to postulate that attention enhances targets and inhibits distractors, it is non-trivial to implement such mechanisms in a visual system that is distributed across cortical regions. For example, how does the visual system determine which of several stimuli is a distractor when processing is distributed across multiple maps specialized for different kinds of information? Then, once a stimulus has been determined to be a distractor, how is it determined which neurons are currently representing that stimulus in order that they can be suppressed? Furthermore, how does the brain implement such a coordinated modulatory process across the visual system's network of interconnected maps without requiring an exhaustively large number of intra-cortical connections or a considerably prolonged iterative process?

An additional complication arises when we consider that the attentional system cannot afford to implement a crisp categorical distinction between targets and distractors. No matter how strongly a person is engaged on a task, there must always be a possibility for task-irrelevant information to trigger attention so that the system remains responsive to unexpected dangers. This requirement explains why attentional capture is such a robust phenomenon. Participants can be told to ignore certain kinds of stimuli and yet those stimuli, when presented with sufficient salience, consistently produce capture costs (Theeuwes 1991). Thus, it must be the case that all stimuli, whether designated as targets or distractors by the experimental paradigm, are evaluated to some degree. One might be tempted to argue that inattentional blindness experiments (Rock, Linnett, Grant & Mack 1992; Neisser & Becklen 1975; Simons, & Chabris 1999) demonstrate effective suppression of unexpected information. However, many subjects do notice the unexpected stimulus in such experiments. Moreover, the proportion of participants who noticed, for example, the black gorilla in Simons & Chabris (1999), was influenced by the attentional set of the observer. Those monitoring the players in black shirts noticed the black gorilla more often than those monitoring the players in white shirts. Thus it must be the case that specific visual features related to the task-irrelevant gorilla are evaluated.

The goal of this paper is to better understand the architecture and neural mechanisms for implementing reflexive attention in a distributed visual system using computational modeling. This model will describe neural mechanisms that make decisions about where to deploy attention, and implement the consequences of those decisions (i.e. the enhancement and suppression of processing) as an integrated system. This neurocomputationally formalized approach provides the luxury of using both behavioral and EEG data as constraints on model development. Thereby, a diverse set of constraints will be brought into contact with the theory, including neural plausibility at the architectural level, accuracy, and reaction time measures of attentional influences on behavior, and also the N2pc and Pd EEG components which are commonly associated with visual attention deployment. Furthermore, because the model is a neural simulation that can generate simulated EEG components, it provides an account of the underlying cause of those potentials in terms of neural mechanisms.

Scientific Philosophy of this account: This paper is written with the philosophy that no model of a psychological or neural process should be considered as entirely correct, since theories are always incomplete in some respect. This paper presents our current best guess at the mechanisms involved in reflexive attention, according to a list of empirical constraints. The model is used to distill a list of insights about the nature of attentional processes, and the paper concludes with a series of predictions to guide future research. By publishing such predictions in advance of testing them, it is demonstrated that these predictions were generated *a-priori*[1].

Model scope. This model is not to be taken as a complete model of visual attention, which would be beyond the scope of any single paper. The primary goal of the model is to discover the mechanisms by which the reflexive attentional system makes decisions about which stimuli to attend and suppress. RAGNAROC does not address, for example, how attentional control affects eye movements (Rao, Zelinsky Hayhoe, & Ballard 2002; Zelinksy 2008) or slower forms of covert attention that are more firmly under volitional control. The model includes a mechanism for the enhancement of information processing, but this is intended only as a proxy for more comprehensive explanations that would interface more directly with single-unit data (e.g. Reynolds & Heeger 2009; Beuth & Hamker 2015).

---

[1] There are a few cases in which model predictions intersect with existing data and these will be called out in the text.

1.2 Behavioral evidence for covert attentional control mechanisms in vision

1.2.1 Reflexive Attention

Reflexive attention is likely to play a role in many visual tasks, and can be specifically observed in paradigms that produce attentional cueing (Posner 1980; see Chen & Wyble In Press) attentional capture effects (Theeuwes 1991, Folk, Remington & Johnston 1992; Yantis 1996) and the attentional blink (Shapiro Raymond & Arnell 1992; Chun & Potter 1995) (Figure 1). In these paradigms, the effect of attention varies according to the nature of the stimuli and the required response. For example, in cueing paradigms a cue increases the accuracy and decreases reaction times for a subsequent target at the same location, while having the opposite effect for targets at other, uncued locations. In attentional capture paradigms, a highly salient distractor causes slower and/or less accurate report of a target presented at a different location and enhanced report of a target at the same location as the salient singleton (Folk, et al. 1992). In the case of attentional blink paradigms, when two targets are presented sequentially at an SOA of about 100ms or less, the second target is easy to see but only when the two targets are presented at the same location (Visser Bischof & DiLollo 1999; Wyble & Swan 2015).

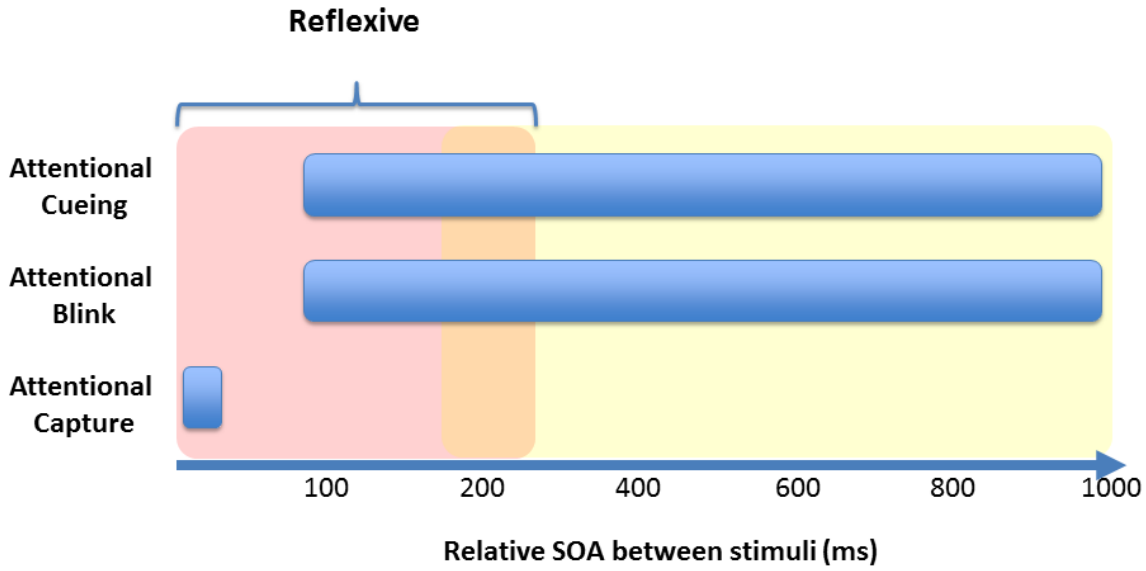

Figure 1. Illustration of the time course during which reflexive attention is likely to play a dominant role in determining performance for three commonly used visual attention paradigms. The blue bars indicate the inter-stimulus SOAs that are most typically used for each.

1.2.2 Semi-Autonomous attentional control

During normal visual function, the eyes move rapidly and abruptly from one location to another which means that information is only available at a given location in the visual field for a few hundred milliseconds. Therefore it seems imperative that some form of attentional control is able to make rapid decisions without waiting for confirmation from slower, more deliberate forms of cognitive control. In this view, reflexive attention is a solution to the demands of the saccadic visual system in that it provides a semi-autonomous decision-making process that is not directly controlled by volitional commands. It is *semi-autonomous* in the sense that it obeys a configuring attentional set that modifies how readily different stimulus attributes will trigger attention; the actual decision to deploy attention, however, is made directly in response to stimulus onset. The autonomy of this system is evident in the phenomenon of attentional capture, wherein the attention system makes inappropriate and sometimes detrimental decisions about where to deploy attention when faced with highly distracting information (Remington, Johnston & Yantis 1992; Theeuwes 1992; Wyble, Folk & Potter 2013; Folk, Leber & Egeth 2002). These studies show that certain kinds of salient stimuli that appear in locations of the visual field that are known to always be task-irrelevant by the experimental subject can nevertheless trigger the deployment of attention. Similarly, salient distractors that are expected and known to be irrelevant to the task will continue to distract attentional control mechanisms over many repeated trials (Remington, Johnston & Yantis 1992; Theeuwes 1992). If reflexive attention were not semi-autonomous, top down control signals would be able to ensure that stimuli presented in locations known to be irrelevant, or stimuli that are themselves known to be task irrelevant would never be able to trigger attention and thus capture effects could be more easily eliminated with foreknowledge or task experience. Note that there are some cases in which top down control settings seem to able to eliminate salience-based capture (Bacon & Egeth 1994)

Another indication of automaticity comes from Krose & Julesz (1989) who demonstrated that cueing effects were localized to the specific location of a cue in a ring of stimuli, even when subjects were informed that the location of the target would typically be opposite to the location of the cue on the ring. Thus expectation induced by both task instructions and experience with the task were unable to alter the immediate, reflexive deployment of attention in response to the cues at SOAs up to 260ms[2]. A finding in electrophysiology by Ansorge, Kiss, Worschech &

[2] CTSOAs were only evaluated in the range of 100-260ms,.

Eimer (2011) showed that spatial cues which are never in the target's position will nevertheless generate an N2pc component, with an amplitude that is weighted by top-down feature settings.

A further line of evidence for automaticity is found in a series of experiments in which the cue (a pair of lines) was much larger than the target, and the subject could, in principle, learn how the cue's properties (e.g. color or shape) determined which part of the cue indicated the likely location of the target (Kristjansson & Nakayama 2003; Kristjansson, Mackeben & Nakayama 2001). It was found that subjects could learn simple relationships, such as that part of a cue (e.g. its left or right half) was more likely to cue a target's location if that relationship remained consistent across trials. However, if the relationship altered in a perfectly predictable fashion (i.e. the left half of the cue indicated the target on trial N and the right half on trial N+1) then there was no learning at all relative to a condition in which the cue-half/target relationship was random. Likewise, if subjects were instructed that the left or right half of the cue would be important on a given trial, they could not force attention to be deployed in accord with the instructions (Kristjansson, Mackeben & Nakayama 2001). These findings dovetail with those of Krose & Julesz (1989) in support of the idea that there are striking limitations on how expectations or instructions can alter the deployment of reflexive attention.

Another limitation on reflexive attention is that its temporal window is limited in duration, even when it would be advantageous for attention to remain for a longer time period. A good example of this is the transient attention demonstration of Nakayama & Mackeben, (1989) in which, a cue appeared, and stayed on the screen to indicate the location of the target. Even though this cue stayed on the screen and was perfectly predictive of the target location, targets that occurred in the 200ms window after cue onset were reported more accurately than targets appearing at later time points. This effect was replicated by Wilschut, Theeuwes & Olivers (2011) though with a smaller magnitude. This transient effect is not merely an alerting effect since it is spatially selective (Muller & Rabbitt 1989).

Therefore, reflexive attention seems to be locked to the temporal and spatial onset of a stimulus. However, one aspect of attention that is amenable to control is what kinds of stimuli can trigger it. Seminal work by Remington, Folk & Johnston (1992) showed that attentional capture effects are moderated by the type of target, such that when subjects are trying to discriminate targets according to their color, distracting stimuli of the target's color are able to cause capture, while stimuli possessing other unique attributes (e.g. motion) have a weaker and sometimes non-

existent ability to capture attention[3]. The functional advantage of attention being directed by top-down factors is clear, since it allows goals to direct attentional deployment to information that is more likely to be useful. This task-defined specificity coupled with the rapidity of reflexive attention provides a potent way for attention to select task-relevant information even when stimuli are changing rapidly (e.g. Potter 1976; Schneider & Shiffrin 1977).

1.2.3 Transient Enhancement at a cued or target location

Several independent lines of research suggest that deploying reflexive attention enhances the processing of targets in the same corresponding location. For example, spatial cueing paradigms find that relative to an uncued condition, a cue will reduce the reaction time to respond to a probe at the same location (Eriksen & Yeh 1985) or increase the accuracy of responding to a masked target at that location (Nakayama & Mackeben 1989; Cheal, Lyon & Gottlob 1994; Wyble, Bowman, Potter 2009). These effects typically onset very soon after a cue (i.e. within about 100ms). This rapid deployment of attention can be contrasted with a slower time course of attention when the cue indicates the to-be-attended location through a symbolic form such as an arrow (Cheal & Lyon 1991; Muller & Rabbit 1989). The key defining characteristic of the rapid onset of attentional enhancement seems to be that the cue and target appear at the same location, which dovetails with the semi-autonomous nature of reflexive attention.

While most studies use cues to trigger the deployment of attention, Wyble Bowman & Potter (2009) tested the hypothesis that targets trigger attention, assuming that the presumed role of reflexive attention is to rapidly enhance target processing. In a design that presented two sequential targets among distractors in a continuously changing array, it was found that report of a second target more accurate, when it occurred immediately after a preceding target, and in the same spatial position.

1.2.3 Visual attention suppresses the processing of information at the location of distractors

Because distractors are, by definition, not explicitly reported or responded to, it has been more difficult to understand how they are affected by attention. One source of information has been to

---

[3] There is an ongoing debate concerning the ability of top-down goal settings to mediate attentional capture (recent REFs), with positions ranging from attention being entirely driven by bottom-up factors, to the opposite extreme in which the first stage of attention is entirely driven by bottom-up.

record directly from neurons within the visual system and there are indications in neurophysiology that representations elicited by distractors are suppressed. In single-unit data from monkeys, neurons responsive to a distractor exhibit a sharp reduction in firing rate after about 80 ms (check) when presented alongside a target in the visual field (Chelazzi, Miller, Duncan & Desimone 1993). This finding has been taken as evidence that targets and distractors engage in a competition that is biased towards the target (Desimone & Duncan 1995).

In human behavior, evidence of distractor inhibition in response to a target takes two forms. First, information is suppressed in the surrounding vicinity of a target as demonstrated by methods that ask subjects to report two targets presented in rapid sequence or simultaneously, which reveal an effect termed Localized Attentional Interference (LAI). For example, Mounts (2000) varied the separation of two sequential targets and found a donut-shaped profile of inhibition. The second target was most accurately reported when ut was in the same position as the first target, and least accurate in the area surrounding the first target. Note that this paradigm does not measure distractor suppression directly but it does indicate that processing is suppressed in a region of the visual field near the target. Bahcall & Kowler (1999) presented a similar finding in which targets were presented simultaneously, and their spatial proximity was varied. They found that accuracy of reporting both targets was higher when they were presented farther apart. Cutzu & Tsotsos (2003) reported a similar finding in which a cue induced a spatial inhibitory surround that was measured with target report accuracy. Hickey & Theeuwes (2011) showed that the effect of a distractor that captures attention is greater when spatially proximal to a target which also implicates a proximity based form of inhibition produced by attention elicited by the distractor.

Another form of distractor inhibition concerns effects that are focused at the location of distractors presented concurrently with a target. Cepeda, Cave, Bichot & Kim (1998) found that when distractors were presented concurrently with a to-be-reported target, a subsequent probe would be reported more slowly at the location of that distractor, compared to a blank location. The implication is that the distractors in the display were suppressed and this suppression carried over in time to slow down processing of probes presented at the same location. Similarly, Gaspelin, et al. (2015) found that probe letters in a spatial array following or coincident with a search display were harder to report if there had previously been a salient distractor at the location of that letter. Like Cepeda et al. (1998) these results also indicated inhibition at the location of information that would have potentially interfered with processing of the target, and

was presumably suppressed. Thus there seems to be two different kinds of evidence for distractor inhibition, one which is spatially locked to the region surrounding a target, and the other of which occurs at the location of distractors. The model presented below will attempt to reconcile these two forms of evidence.

1.3 Electrophysiological correlates of visual attention

One of the most influential experimental approaches in the study of visuospatial attention uses Event Related Potential (ERPs) from the EEG. ERPs provide a measure of neural activity that is precisely timed to underlying neural events and therefore can be linked to distinct levels of processing. The electrical potentials are thought to be predominantly caused by current flows induced by excitatory synaptic input, especially from cortical pyramidal neurons that are oriented in parallel to one another and orthogonal to the skull.

1.3.1 ERPs indicating current attentional states

When spatial attention has been directed to a specific location for a sustained period of time, it produces changes in potentials evoked by stimuli that occur in those locations. For example, early components such as the N1/P1 complex, elicited by the onset of a visual stimulus, are larger in amplitude for a stimulus that appears in an attended location (Mangun 1995; Hillyard & Anllo-Vento 1998) and presumably reflects increased neural activity evoked by stimuli at those locations. Likewise, increased amplitude of the Steady State Visual Evoked Potential (SSVEP) for a flickering stimulus has served as a robust indicator of the location of attention and can last multiple seconds (Müller & Hillyard 2000). These indicate that ongoing attention affects the processing of stimuli at the earliest levels of cortical processing.

Note that modulations of early components can indicate short-term changes in the attentional state evoked by an immediately preceding stimulus. For example, in the case of Hopf, Boehler, Luck, Tsotsos, Heinze & Schoenfeld (2006), probe stimuli were presented 250 milliseconds after a target display, and the amplitude of the magnetic impulse induced by the probe in MEG mapped the spatial distribution of attention evoked by the target. The probe evoked a larger response at the same location as the target, the lowest response just next to the target and a more moderate response at more distal locations, which resembles the behavioral measure of the spatial profile of attention described by Mounts (2000).

1.3.2 ERPs indicating changes in the state of attention

Another class of EEG component is thought to indicate the neural mechanisms involved in the deployment of attention. Unlike the N1/P1 and SSVEP, which indicate how ongoing attentional states affect processing of stimuli, this second class of ERPs indicate how attention changes in response to the onset of new stimuli. These potentials, termed the N2pc and the Pd, occur later in time than the modulations of the N1/P1, because they are subsequent to the initial stimulus processing that allows attentional mechanisms to determine which stimuli should be attended.

The N2pc component:

The N2pc is a brief negativity elicited by a laterally presented target stimulus. The potential is typically small in amplitude and occurs on posterior areas of the scalp contralateral to the target, approximately 200-300ms after the onset of the target, and is typically less than 100ms in duration. The original set of theories proposed by the seminal publications on the N2pc suggested that it reflects either the suppression of distractors (Luck & Hillyard, 1994) or the enhancement of the target (Eimer, 1996). The N2pc has also been referred to as a Posterior Contralateral Negativity or PCN (Töllner, Rangelov & Müller 2012).

Newer findings have provided different perspectives. For example, it has been suggested that the N2pc reflects the process of individuating the target from surrounding stimuli, as its amplitude increases with the number of presented targets, but only when target numerosity is task relevant (Pagano & Mazza 2012). Also, Hickey, Di Lollo, & McDonald, (2009) suggested that when a target is paired with a contralateral distractor, the N2pc to the target is actually composed of two dissociable components: a negativity evoked by the target (the Nt) and a positivity evoked by the distractor (the Pd). Since the N2pc is measured as a difference wave between target-contralateral and target-ipsilateral sides of the scalp, the Pd would be measured as a negativity relative to the target, thus contributing to the N2pc amplitude. The Nt and Pd components were isolated by presenting the distractor or the target, respectively in the middle of the display, which eliminates its contribution to the ERP and reveals the neural signature evoked by the other stimulus.

Another perspective on what processes are indicated by the N2pc stems from a finding in which two sequential targets are presented at either the same or different locations on the screen(Tan Wyble 2015). In the same-location condition, subjects could easily see the second target, however it elicited no N2pc. In contrast, when the second target was on the other side of the

display, a second N2pc was evoked by that second target. In terms of behavior, subjects were actually better at reporting the same-location target, which did not evoke an N2pc, compared to the different-location target which did evoke an N2pc. From this work, the conclusion is that the N2pc indicates only the initial localization of a target, rather than enhancement or suppression. This explains the missing-N2pc in the same-location condition, since the second target inherits the attention deployed by the first target, and no additional N2pc is evoked. Moreover, this missing-N2pc phenomenon is only true when T1 and T2 are presented closely in time. At longer temporal separations (e.g. 600 ms), the T2 elicits a second N2pc, even if subjects have a clear expectation that the target will occur in that location (Callahan-Flintoft & Wyble In press). This finding is crucial because it underscores that reflexive attention is driven by a stimulus, and cannot be maintained during a duration.

With these various theories in contention, it is clear that little consensus as to the underlying cause of the N2pc exists. Moreover, a crucial complexity of the N2pc literature is that distractors evoke an N2pc in certain cases (Hickey McDonald & Theeuwes 2006; Burra & Kerzel 2013; Kiss, Grubert, Petersen, & Eimer, 2012; McDonald, Green, Jannati, & Di Lollo, 2012; Liesefeld, Liesefeld, Töllner, & Müller 2017). Such findings highlight the complexity of attentional mechanisms and the difficulty of ascribing unitary functions to neural correlates.

The Pd Component:

Another ERP related to attentional control is the Pd component, a positivity evoked in posterior scalp regions that are contralateral to a distractor (Hickey, McDonald & Theeuwes 2009). For example a target presented in the middle of a display with a flanking distractor on the right or left has been found to elicit a positivity on the posterior scalp contralateral to the distractor. The fact that distractors selectively elicit a Pd has been taken as evidence that inhibition is selectively deployed at the location of distractors. This dovetails with behavioral evidence that distractor inhibition is specifically targeted at the location of distractors.

Another important finding in the Pd literature, as mentioned above, is that a lateralized distractor can produce an N2pc, as mentioned above. As we will argue, such findings are indicative of the competition for attention that occurs when targets and distractors are presented simultaneously. This distractor-induced N2pc also supports the suggestion that distractors have an inherent salience which must be inhibited (Sawaki & Luck 2010). In other words, the aspect of salient distractors which requires suppression (and thus the Pd) is their inherent ability to trigger the

deployment of attention. Such findings complicate the straightforward attribution of the Pd as a correlate of distractor suppression. Deriving a more nuanced understanding of the causes of the N2pc and Pd components is a key contribution of this model.

2. Computational architectures for distractor suppression

Moving to a discussion of how reflexive attentional control might be implemented we begin by considering several architectures that could support the ability to selectivity suppress the processing of distractors.

2.1 Assumptions

This discussion is predicated on several assumptions that are implicit in existing models of attention. An anatomical assumption is that the visual system is hierarchically organized, beginning with low level feature extraction in cortical area V1 that projects to various cortical areas specialized for more specific kinds of information, such as color, and various forms (Van Essen & Maunsell 1983, Kravitz, Saleem, Baker, Ungerleider, & Mishkin, M. 2013; Konkle & Caramazza 2013). These higher-level representations will be assumed to maintain the spatial topography of V1 albeit with larger receptive fields (DiCarlo & Maunsell 2003; Silver, Ress & Heeger 2005). This hierarchy places some important constraints on how different kinds of information flow through the model. Another crucial assumption is that *there is no indicator that definitively determines which stimuli are targets or distractors.* Instead, the attentional system perceives stimuli with varying combinations of intrinsic salience and task relevance, and has to decide which stimuli to attend according to the broader goals of the organism, which may sometimes transcend the specific task imposed by the experimenter. In other words, the notion of target and distractor, while convenient descriptors in designing an experimental task, have no categorical meaning for the attention system. The system must be prepared to deploy attention to any stimulus, regardless of its immediate task relevance so that it can allow threatening or otherwise important stimuli to be perceived. The model must decide what will be attended and what will be suppressed on each trial and this decision is not pre-ordained by a higher level cognitive system.

A final crucial assumption is that there are no indicators as to which neurons are processing which stimuli. When a decision has been made to attend to a stimulus, there must be a way to rapidly distribute the consequence of that decision across a diverse set of cortical areas. The neurons in those areas are not pre-labeled according to what stimuli they represent. One way to

solve this challenge is for visual cortical areas to have shared spatial topography, which allows enhancement and inhibition of processing to be targeted towards the correct neurons.

Given these assumptions, a candidate model of reflexive attention must include mechanisms for making rapid decisions about where to attend, and also mechanisms that implement that decision.

2.2 Four potential architectures:

Local Attentional Control:

The simplest method of reflexive attentional can be implemented at the local circuit level. In such a model, stimuli are processed separately within different maps. Representations of each stimulus compete within these cortical area, and one or more winners of that local competition would be attended, according to internal mechanisms within that area. While simple, this architecture has difficulty explaining how stimuli of different kinds can affect one another. For example attentional capture by a color singleton affects processing of a shape singleton target (e.g. Theeuwes 1991). To explain such data requires an attentional architecture in which stimuli of different types can influence one another.

Peer to Peer attentional control

The simplest architecture that allows attentional effects to interact across stimulus types involves connections between cortical areas (Figure 2a). Any map that resolves its competition between stimuli projects signals to the other maps using intracortical pathways that take advantage of the spatiotopic correspondence between feature maps throughout the visual system so as to ensure that the correct locations are excited or inhibited across maps.

The disadvantage of this architecture is that it requires an enormous number of intracortical projections. Each different feature map within the visual system must send projections to every other map so that targets in one map can inhibit distractors in all other maps. Thus, the number of intra-areal projections grows as N^2, where N is the number of feature maps. Note also that each intra-areal projection would necessarily be a full topographic projection and thus quite dense. Thus, this architecture would be inefficient in terms of the required number of dense fiber tracks for a complex visual system with multiple processing areas. It has been estimated that

only about 30% of the total proportion of possible intra-areal connections exists within the macaque visual system (Felleman & Van Essen 1991).

Feedback Attentional Control

The next architecture is more efficient in terms of intracortical projection (Figure 2b) because it takes advantage of the hierarchical nature of the visual system. Once a stimulus has won a local competition in any map, it projects a combination of enhancement and inhibition back down to the earliest levels of processing in the visual hierarchy (i.e. perhaps V1 or even LGN). These effects then propagate forward to all of the descendent visual processing areas.

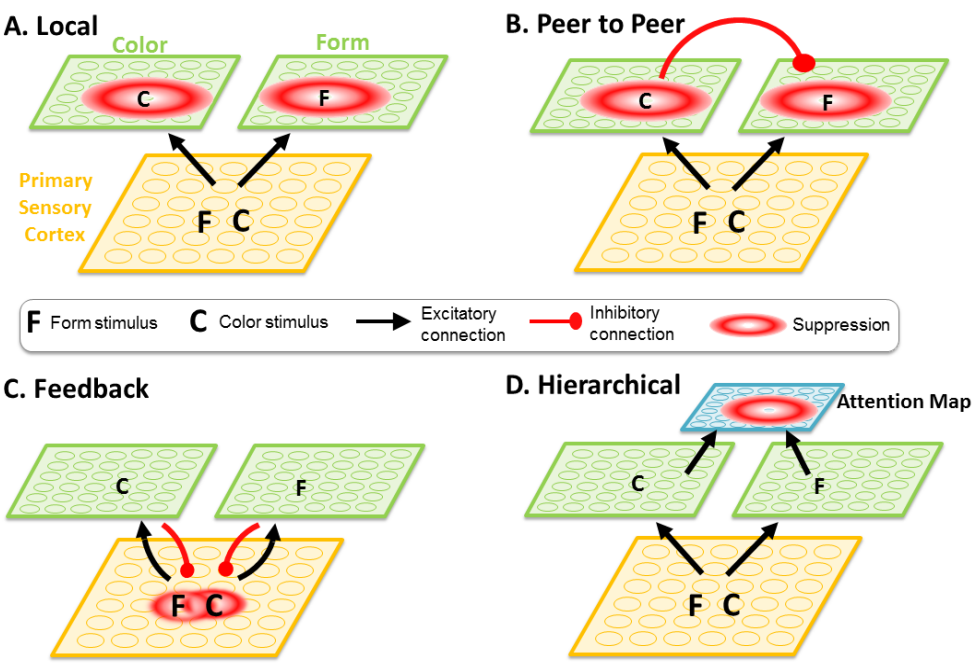

Figure 2. Illustration of four architectures for mediating the competition between two stimuli who's most salient feature are computed in different maps (e.g. color singleton vs a form singleton). See text for an analysis of their relative advantages.

This approach requires dramatically fewer inter-areal connections than the peer-to-peer model, growing linearly with the number of feature maps. The selective tuning model of Tsotsos (2011) et al provides a thorough formalization of such a system.

The disadvantage of this approach is that a competition between two stimuli represented in distinct feature maps would require multiple iterations of feedforward and feedback processing

through the hierarchy. Furthermore the suppressed information is cut off at the earliest level, which precludes it from analysis by higher levels of the visual system. This makes it difficult for deeper levels of meaning to be computed from stimuli that are not attended because that information would be prevented from cascading through the hierarchy.

Inhibition at a superordinate map

The final architecture that we consider, and the one that is used in RAGNAROC, confines the competition to a single cortical area: an attention map that is hierarchically superordinate to all of the spatiotopic maps that comprise the ventral visual system (Figure 2c)[4]. The attention map provides an architecturally compact method to make decisions about where to deploy attention in the visual field, since it accumulates information about the priority of different stimuli from many subordinate topographic maps into a single brain region.

Once a spatial region within the attention map has been sufficiently activated by input from subordinate maps, those neurons enhance the processing of information at the corresponding region of the earliest processing level, and that enhancement then carries forward through the ventral hierarchy. Inhibition of distracting information is accomplished by inhibiting regions of the attention map rather than neurons within the subordinate maps.

An important distinction of this architecture relative to the others is that there is no direct suppression of information in the subordinate layers. Instead, attentional priority is achieved indirectly by reducing the availability of attention at particular locations in the visual field. Thus, attention is represented as a gradient field of activation levels distributed across the spatial extent of the visual field (LaBerge & Brown 1989, Cheal Lyon & Gottlob 1994). Changes in these activation levels provides a convenient way to throttle the processing of information through all of the feature maps that are descended from the early visual area with a relative minimum of intracortical projections. This approach mitigates the disadvantages of the preceding architectures. Attentional decisions can be made rapidly since they occur at a single level within the hierarchy. Also, suppressing attention, rather than representations, preserves the information at the earliest levels of processing which allows a stimulus in an unattended region the chance to make contact with deeper levels of processing.

---

[4] For simplicity we assume that there is only a single cortical area, although the advantages of this architecture are also obtained if there is a small family of interconnected cortical areas instead of just one.

3. RAGNAROC specification

The remainder of this paper describes the architecture and functionality of the RAGNAROC model.

3.1 Inspiration from existing models

There is a substantial literature on computational models of attention that collectively addresses a broad set of mechanisms and processes. The RAGNAROC model is informed by many of these models, theories and ideas. One of the most formative models for our approach is the Theory of Visual Attention (TVA; Bundesen 1990; Bundesen, Habekost, & Kyllingsbæk, 2011) which stipulates a mathematical formulation for how goals adjust the pertinence of certain kinds of information (i.e. when one is looking for red digits, the pertinence of red is upweighted to increase the rate at which stimuli with that color are processed). This pertinence weighting applies across the entire visual field, which explains how stimuli are able to capture attention when they match top-down settings despite being located in a to-be-ignored location of the visual field. TVA also set the stage for understanding stimulus-driven attention as a formalized decision making process. In terms of implementing a hierarchical architecture for attention, the Koch & Ulman (1985) and Itti Koch & Niebur (1998) models of salience were crucial for establishing the utility of a shared salience map which accumulates information from subordinate layers of processing and allows them to compete on an even footing in a compact neural field. Zehetleitner, Koch, Goschy & Müller (2013) elaborated the circuitry of competition at the top of this hierarchy, to provide an illustration of how attention operations can be considered a race between competing selection operations.

Whereas salience map as an output of what amounts to a feed-forward process, the Selective tuning (ST) model of Tsotsos (1995) illustrates how recurrent excitatory signals, propagating backwards through the visual hierarchy could implement the selection process at the earliest levels of the hierarchy. Numerous models by Grossberg also highlight the importance of recurrence in iteratively managing the spatial profile of attention (Fazl, Grossberg, & Mingolla 2009)

The focus of our modeling effort, which incorporates ideas from many preexisting models, is to develop these ideas further, in close contact with a large body of empirical data, and also to extend the implications of such models to understanding the generators of attention-related

ERPs. The RAGNAROC model simulates the consequences of attentional decisions rippling through the visual hierarchy, creating transient attractor dynamics that allow attention to lock-on (Tan & Wyble 2015) to one or more locations that contain a stimulus. The term *lock-on* refers to a state in which top-down enhancement at a given spatiotopic location in the visual field amplifies feed-forward projections from stimuli at that same location, which further strengthens the attentional state and suppresses attentional priority in neighboring regions. Another way to state this is that the decision to attend is simulates as a spatiotemporal impulse response function within a 2-dimensional neural field. This function shapes how information is processed in subordinate visual maps.

Model outputs. The model simulates both the behavioral consequences of reflexive attentional deployment, measured in terms of accuracy and reaction time, while also simulating certain ERP correlates of attention, such as the N2pc and the Pd components. As a starting point, the model incorporates the convergent gradient-field (CGF) model of Tan & Wyble (2015), which provided a partial explanation for reflexive attention with a focus on target enhancement and the N2pc component.

Stimulus Processing and Differentiation. In order to provide a generic account of attention that can be applied to many different experimental contexts, we do not commit to the decoding of pixelwise representations. Instead, simulated neurons in each spatiotopic map represent the presence of attributes at locations with a granularity of 0.5 degrees of visual angle. These representations are segregated into distinct maps that are each specialized for particular kinds of stimuli, an assumption that is common to other models (Itti et al. 1998).

Localizing and attending important stimuli. The RAGNAROC model is built with the assumption that an essential function of the visual attention system is to determine a to-be-attended stimulus' precise location from relatively ambiguous input so that its representation can be enhanced. The requirement to localize a target stems from the fact that location information becomes less precise during progression through the visual hierarchy. We assume further that attention enhances the rate of processing of information at that location so as to enable the attended information to be encoded into memory more accurately and more rapidly. Such a model is consistent with findings that detection of a target facilitates the processing of a subsequent target at that same location (Wyble, Potter & Bowman 2009; Mounts 2000). This assumption is also consistent with the theory of biased competition (Chelazzi et al 1993) in

which detection of a target allows neural responses from an attended stimulus to dominate responses from unattended stimuli.

The distinction between Targets and Distractors. Another assumption of RAGNAROC is that all stimuli initially excite activity in the attention map because there is no clear-cut distinction between targets and distractors for processes on the time scale of reflexive attention. This component of the attentional system perceives stimuli only in varying levels of attentional *priority* (defined below) since a decision is made about the relative priority of different stimuli before input from slower, more deliberate stimulus evaluations are completed. In this framework, targets (to the extent that the visual system perceives them as such) are distinguished from distractors by eliciting a more rapid representation at their location in the attention map. The decision process is, in effect, a race between competing representations, and attentional set plays a key role in helping targets to win the race.

However, the outcome of this race based decision process is not preordained and the attention system is prone to deploying attention to highly salient, but task irrelevant distractors on some fraction of trials. Thus, the model explains attentional capture experiments as a mixture of trials, which often exhibit successful deployment of attention to the target, but sometimes allow a distractor to win the competition. The variability in these outcomes is mediated by trial-to-trial variability.

Priority value. Each Stimulus in the visual field receives a priority value, which is a valuation of its ability to trigger the attentional system according to a combination of physical salience, and top-down contributions from attentional set (Figure 3). The priority is defined in part by a stimulus' salience, which reflects the degree to which it stands out from other nearby stimuli in terms of low-level features (e.g. color, orientation, luminance). Priority is also affected by the degree to a stimulus matches the current attentional control settings. Other potential contributions to priority that will not be considered here could involve whether stimulus attributes have been associated with reward (Anderson, Laurent & Yantis 2011), have been recently presented (Awh, Beloposky & Theeuwes 2012), were recently task relevant (Kadel, Feldmann-Wüstefeld & Schubö 2017) or are relatively novel in time (i.e. an oddball). Thus, a strength of the attention-map approach is to allow a broad variety of influences to affect how stimuli are prioritized.

Using Attentional control settings to prioritize targets: To help guide attention towards targets, it is assumed that attentional control settings prioritize certain stimuli as defined by the current task in a spatially nonselective manner (Saenz, Buracas & Boynton 2002; Zhang & Luck 2009). These attentional control settings prioritize simple features like color, or more complex attributes such as conceptual categories (e.g. dinner food, animal, etc) by upweighting feedforward activity from some maps and downweighting feedforward processing from other maps. We assume that the ability to select certain stimulus attributes for weighting is governed by pre-learned stimulus categories (e.g. letters vs digits) but cannot easily be accomplished for arbitrary distinctions (e.g. select letters A, B, C from other letters). This follows from the work of Shiffrin & Schneider (1977) who demonstrated the ability to efficiently attend to previously learned categories, but not to arbitrary subsets of a category. Based on work suggesting that even conceptually defined target sets can be used to select information from RSVP (Potter 1976) or capture attention (Wyble, Folk & Potter 2013), it is assumed that prioritization can happen even at a conceptual level.

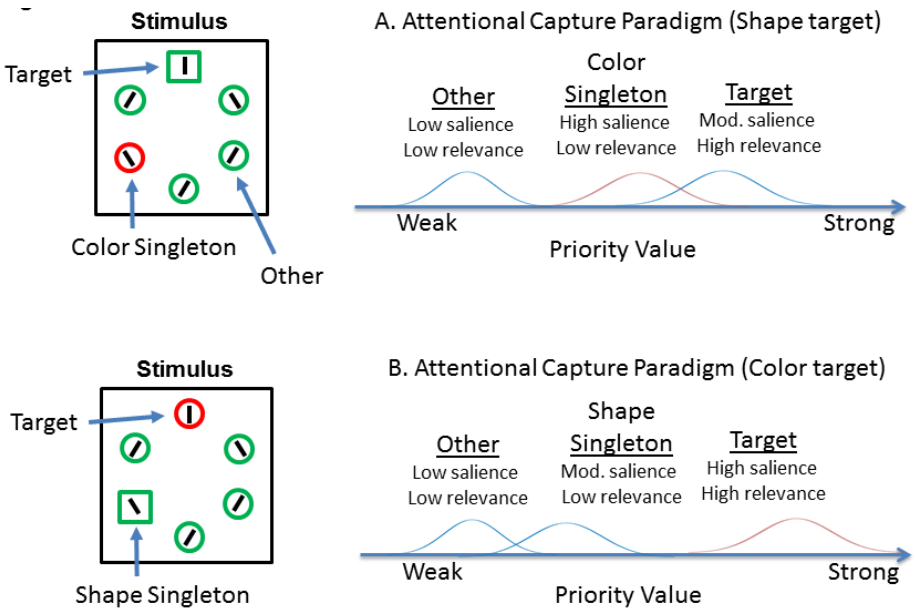

Figure 3. Illustration of how both intrinsic salience and top-down relevance cues can be mapped onto a single priority dimension. The figures illustrate canonical paradigms from attentional capture experiments. The top version of the paradigm uses a highly salient color singleton which results in priority values that are competitive with the less-salient target. When the target and distractor dimensions are switched, the shape singleton is not competitive with the color-singleton target.

3.2 Architecture of the model

3.2.1 A Hierarchy of Visual maps

The visual system is composed of a hierarchy of maps that each represent the visual field and are connected so as to preserve a spatiotopic organization that is rooted in a retinotopic representation at the earliest level (Figure 4). No claims about the number and complexity of this hierarchy are necessary here as this is a model of general principles. Information propagates through the layers via feedforward excitation. The first layer of the model is termed Early Vision (EV) and simulates the earliest cortical regions in the visual hierarchy, which contains neurons with small receptive fields, such as V1. The second layers are collectively termed Late Vision (LV1 and LV2) and contain neurons with larger receptive fields, corresponding to anatomical areas in the ventral visual stream that are thought to be specialized for different kinds of stimuli, such as V4 (color), FFA (faces), the EBA (body parts), the PPA (places), as well as distinctions between animate and inanimate stimuli. In our simulations, EV neurons have a simulated receptive field size of .5 degree while LV neurons have receptive fields of N degrees.

The third layer of the model's hierarchy is the attention map (AM) which receives input from all of the subordinate LV maps[5]. Neurons in the AM have the same broad receptive fields as LV neurons. The primary role of the AM is essentially to implement decision making across the visual field and to enact the consequences of that decision by sending modulatory projections down to the earliest level of the hierarchy. It does this by accumulating spatially imprecise activity from the subordinate LVs and then compute the spatiotopic location of the originating stimulus in the EV. In the AM, convergent input from the LV maps initially forms an activation bump, centered at the location of each stimulus in the visual field. The correct localization of this bump follows naturally from the activation dynamics of the model since the AM neuron that

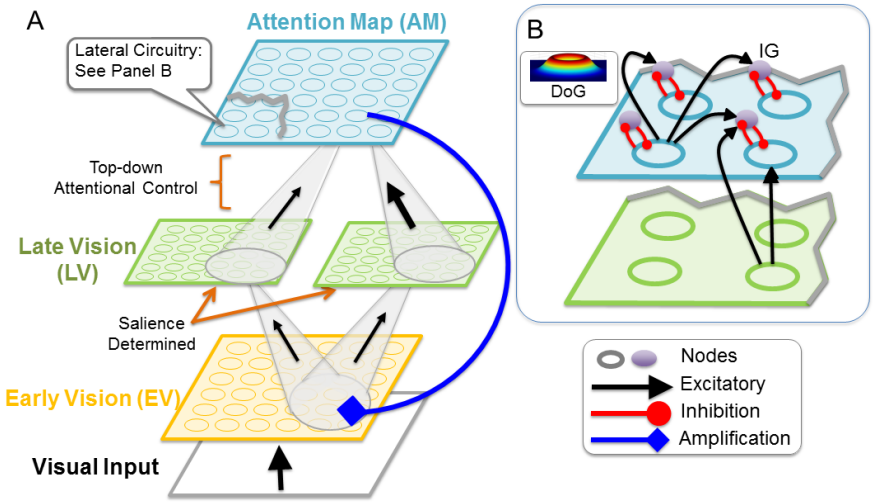

Figure 4. Illustration of the model's macro architecture (A) and the microcircuitry (B) of the attention map. In the hierarchy of visual areas, the cones reflect the set of neurons at a subordinate layer that excite a given neuron in the superordinate layer. Only two layers are shown here, but this architecture would generalize to systems with more levels. The salience of a stimulus is implemented as stimulus-specific differences in feed-forward excitation at early levels. Top-down selection is implemented as feature-specific weights for an entire LV map. The Attention Map returns a location-specific gain enhancement at a given location in the EV. The inset shows how neurons are interconnected within the AM. The small grey circles are Inhibitory Gating (IG) neurons, each of which has a competitive inhibitory relationship with a principle neuron of the AM. The principle neurons excite one another with a spatial distribution that is a Difference of Gaussians (DoG). Not shown in this diagram is the feedback inhibition at the LV layers (see text and Figure 5).

[5] Direct projections from the EV to the AM could exist but they are not represented here for simplicity.

resides at the corresponding topographic locus of the centroid of LV activity will receive the largest amount of input from the LV.

When AM neurons are stimulated above threshold, a multiplicative attentional enhancement is applied to all neurons in the EV at the same spatiotopic location, which increases the strength of the feedforward input from that region of the visual field. Furthermore, the activated AM neuron inhibits other surrounding neurons in the AM. This dynamic creates a brief attractor state, in which the activated peak in the attention map is amplifying its own input and suppressing its neighbors. This condition we term a *lock-on* state and allows a precise, stable localization of attention despite the relatively coarse-grained spatiotopy in the LV representations.

Thus, the attention map integrates information from the subordinate feature maps to localize one or more targets and then projects an enhancement signal back down to earlier areas at the appropriate location(s). There have been a number of proposals for where such an area might exist, including frontal, parietal and portions of the pulvinar nucleus (Shipp 2004). We note that a lateral, parietal location would be broadly consistent with the location of attention-related ERPs, which are typically larger above parietal cortex than directly over occipital, central or frontal areas (e.g. Tan, Wyble 2015). It is also possible that this functionality is distributed over several cortical areas, although that would come at the expense of intracortical white matter to mediate the competition.

3.2.2 Attentional gating circuitry

One of the key innovations in this model is the inhibitory control circuitry within the AM, which has been developed specifically according to certain findings in the literature.

First, to ensure that attention is primarily inhibited at regions of the visual field that contain stimuli (Gaspelin et al 2015; Cepeda et al. 1998), gating circuits ensure that inhibition is only delivered to neurons receiving excitatory input from the visual field (Figure 4b). Each principle neuron in the AM is accompanied by an inhibitory gating neuron (IG) that suppresses its paired AM neuron only when it receives lateral excitation from another AM neuron and excitation from any LV neuron at its own spatial location. As a consequence, only AM neurons that are receiving excitation from the LV can be inhibited by another neuron in the AM. This prevents the AM from inhibiting locations of the visual field that are empty.

Another component of this gating circuit acts to increase the stability of lock on states, such that once a decision to attend to a given location is reached, it is less likely that other stimuli will cause it to disengage. Each strongly active AM neuron inhibits its own IG neuron, a form of competitive inhibition that has been determined to have a similar stabilizing function in well-charted nervous systems such as the drosophila (Jovanic et al. 2016). With this circuit, multiple neurons in the AM can remain active, since they protect themselves from the inhibition generated by the other active AM neurons.

3.2.3 Free and fixed parameters

The model uses predominantly fixed-parameters according to a set of empirical constraints, which are listed below in section 4. These parameter values are invariant for all of the simulations provided below, except for a subset that vary in order to implement the experimental paradigm of each simulation (e.g. Timing and location). There are also two partially-fixed parameters that determine the salience and task-contingent (i.e. bottom-up and top-down ←use this terminology?) weightings of each stimulus type. The term partially-fixed reflects the fact that their relative values are determined by the experimental paradigms. E.g. in simulations of the additional –singleton paradigm (Theeuwes 1991) which has a highly salient color singleton, we allow the specific value of the distractor's salience to vary, but it has to remain higher than the salience of the target. Finally, there is one additional free parameter that defines the accumulator threshold for a behavioral response for each experiment. This parameter is constrained to have a single value for all conditions of a given experiment and prevents behavioral accuracy values from being at ceiling or floor. The specific values will be provided for each simulation in the appendix.

3.3 Mechanisms of the model:

3.3.1 Equations

The model uses rate-coded neurons, governed by the activation equations of O'Reilly & Munakata (2001) as shown in the Appendix. In this equation the activation level of each neuron is governed by three currents: excitatory, inhibitory and leak. This level of abstraction is a compromise that captures the properties of neuronal synaptic interactions in broad strokes while allowing rapid exploration of different model architectures. Moreover the distinction between excitatory and inhibitory currents provides a mapping to current flows underlying EEG components. This set of equations has been used effectively in previous simulations of attentional processes at similar time scales (e.g. Bowman & Wyble 2007)

Each connection within the model is enumerated in Figure 5 and their properties are described below. For simplicity, all maps have the same dimensionality and number of stages. Furthermore only two pathways are simulated here. The mechanisms used her generalize to more complex architectures with more layers and more pathways. All connections between or within layers are assumed to have either an identity projection (i.e. strictly topographic), a Gaussian spread, or a Difference of Gaussians

The following numbers indicate connections specified in Figure 5.

1. Input to EV neurons: The EV represents the earliest stage of cortical visual processing in which neurons have extremely small receptive fields and color/orientation/frequency specific firing preferences. For the sake of simplicity, EV nodes are separated into different areas according to the kinds of stimuli presented, although in the brain these different neurons occupy the same cortical map. Input to a specific EV node is specified as a step function since the simulations are of suprathreshold stimuli (i.e. a value changing from 0 to 1 while the stimulus is visible), which causes the corresponding EV node's membrane potential (MP) to charge according to equation 1.

2. Projection from EV to LV: When an EV node's membrane potential crosses threshold, it sends a feedforward excitation to an array of nodes in the corresponding

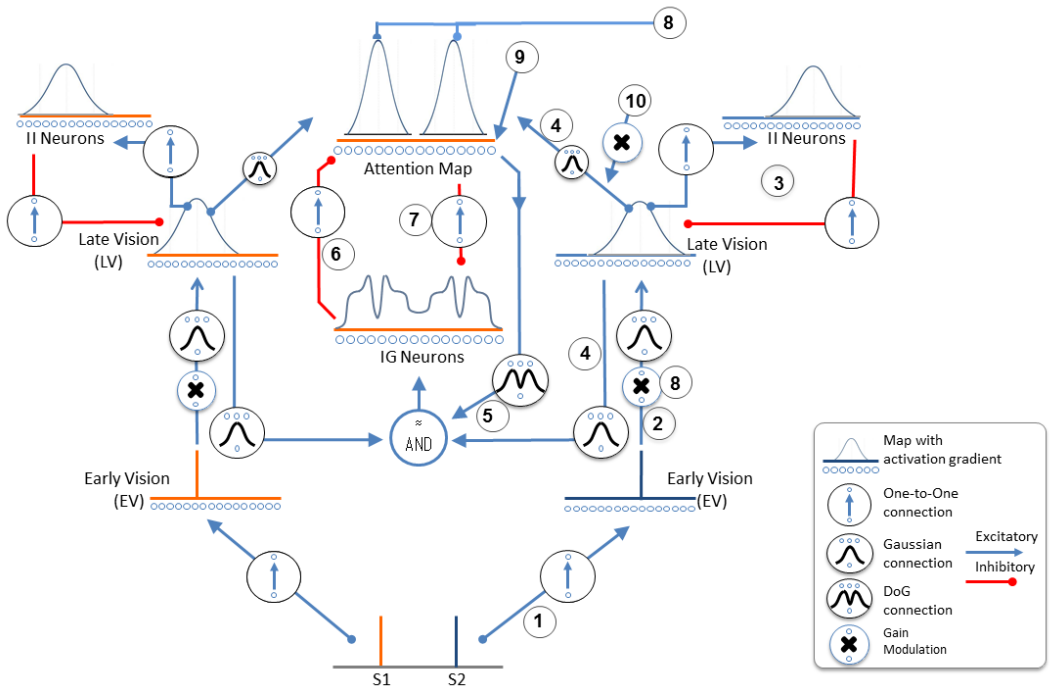

Figure 5. Illustration of the complete architecture for two distinct pathways during perception of two objects with highly distinct dimensions (e.g. form vs color). The left and the right side represent the same connections but for different stimuli. The bubbles indicate the spatial distribution and character of each. The traces above each layer illustrate a typical activation profile for that map in response to a stimulus. Note that the effect of attention affects both pathways, regardless of which stimulus triggered it. The numbers correspond to descriptors in the text.

LV maps. This projection is spatially weighted according to a Gaussian centered at the location of that EV node. The magnitude of this projection is the salience of the stimulus, and indicates its physical dissimilarity to other stimuli in the visual field according to the specific LV it projects to (e.g. a shape singleton would have a high salience in an LV map that is specific for form). Other accounts have shown how to compute salience for some classes of features such as color, orientation and luminance (Zelinsky 2008; Itti et al. 1998;, Bruce & Tsotsos 2006). In RAGNAROC, we abstract over the process of computing salience to accommodate the broad diversity of tasks.

3. Inhibitory feedback nodes in the LV: Each LV node has a dedicated inhibitory interneuron (II) that provides feedback inhibition. This feedback emphasizes the onset of new information and provides LV nodes with the scalloped activity

characteristic of single units in the visual system (e.g. Chelazzi, Duncan, Miller & Desimone 1998).

4. Projection from LV to AM: When an LV node crosses threshold, it projects feedforward excitation to an array of nodes in the AM according to a Gaussian distribution centered at the location of the active LV node. LV nodes also excite IG nodes (see below) with the same Gaussian profile. This projection includes a parameter that represents the task-contingent (i.e. "top-down") weighting of each stimulus type, and is fixed for each LV map.

5. Inhibitory Gating Nodes (IG): The IG nodes ensure that inhibition with the attention map only occurs at locations receiving input from an LV (see also Beuth & Hamker 2015). Each IG node is paired with a single AM node that it can inhibit. An IG node is excited by neighboring AM nodes according to a Difference Of Gaussians (DoG) activation profile. IG nodes are also excited by LV nodes. Excitation of each IG node from these two sources (AM and LV) is capped such that concurrent AM and LV activity is required to raise an IG node above threshold. Thus, these neurons exhibit the rough equivalent of a logical AND gate in that they require concurrent activation from two pathways in order to fire.

6. IG inhibiting the Attention Map: When activated by convergent AM and LV input, an inhibitory gating neuron inhibits its corresponding AM node. This is the basis of inhibitory suppression.

7. Attention Map inhibiting IG: The inhibition from AM->IG occurs when an AM node's membrane potential is above a higher threshold X. This disinhibitory circuit increases the stability of an AM lock-on state, since AM neurons can protect themselves from inhibition generated by neighboring AM nodes.

8. Attentional Enhancement: Each AM node provides a gain modulation of synaptic transmission from EV to LV for all EV nodes at the same location. Note that enhancement of a given location in the EV occurs for the entire Gaussian spread of all EV neuron's feedforward projection, unselectively across all pathways.

9. AM excitatory Bias: There is a uniform level of bias input to the entire AM, keeping these neurons slightly active in the absence of input.

10. Noise input: Intertrial variability is added to the model as modulations of the weights between the LV and A, which represents fluctuations in attentional control. The variability is constant for a given trial and varies between trials as samples from a Gaussian distribution.

3.4 Example Simulations

A critical question that is addressed by RAGNAROC is how reflexive attentional mechanisms respond to the onset of one or more stimuli within the visual field. The following section illustrates representative simulations.

3.4.1 The simplest case: Single stimulus

When a single stimulus of sufficient salience is presented to the EV, it triggers a *lock-on* of attention at its spatiotopic location, which is a self-excitatory attractor state resonating between EV and the AM through the LV. Figure 6a illustrates the impulse response function in the AM elicited by a single stimulus. Figure 6b illustrates the time course of activation of each of the layers of the model. Information propagates upward through the layers.

3.4.2 Transition to the lock-on state

We demonstrate that the lock-on state has the characteristics of an attractor by illustrating that a broad range of stimulus values evoke a bump in the AM of similar size and duration (Figure 7a). The rapid growth of this neuron's activation is due to the attentional enhancement of feedforward activity from the EV after the corresponding node in the AM crosses the threshold value. The decrease in peak amplitude is a result of the drop in LV activity due to feedback inhibition from II.

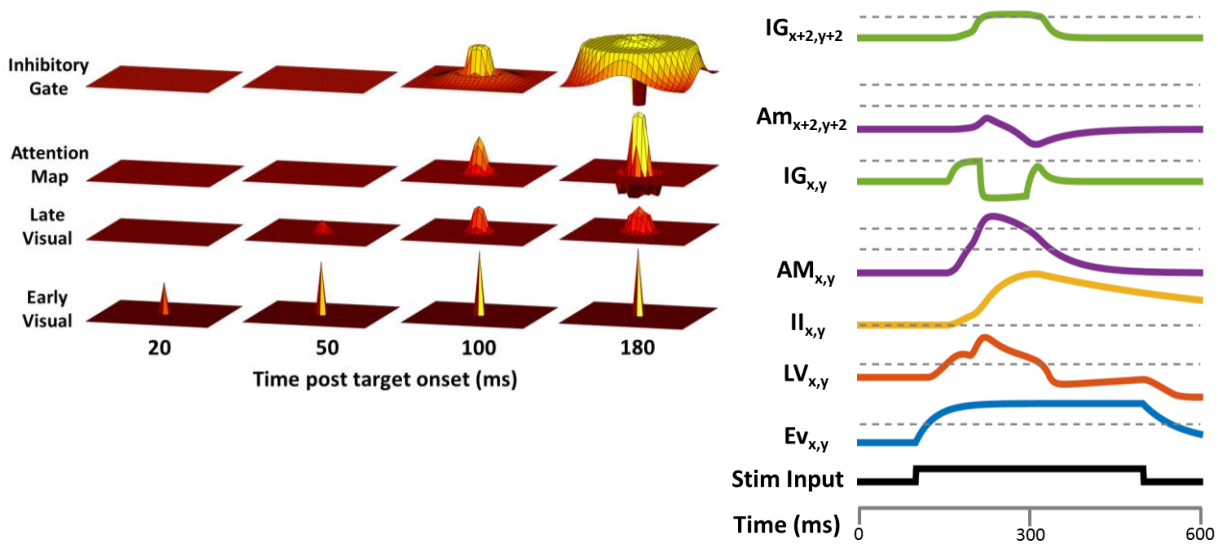

Figure 6. Evolution of a lock-on state across time and layers for a single stimulus. An initial feedforward wave of excitation from a single location in the EV triggers activation in the LV, which carries forward to a peak in the AM. Once the central peak of the AM activation crosses threshold, the surrounding IG neurons are activated producing surrounding inhibition in the AM. However, the suppression of attention in the AM is only expressed at locations containing activation that needs to be inhibited.

3.4.2 Two stimuli

Figure 8 depicts a comparison of AM activity for either one(A) or two stimuli(B,C). In the case of two stimuli, if one of them is of substantially higher priority, it will inhibit the AM at the location of the other (B). However if both stimuli are of similar priority, the two of them will enter lock-on states simultaneously leading to a fully parallel deployment of attention (C). In such a case, each AM inhibits its paired IG (not shown in this figure), and the two lock-on states protects themselves from inhibition by the other.

3.4.3 Sequential stimuli

When stimuli onset sequentially, the first stimulus will typically be able to activate its Lock-on state and suppress activity of the second. In this way, a first target (T1) with a relatively low priority value can suppress attention to a second target (T2) since the temporal advantage of T1 allows it to establish a lock on state before the T2 has a chance.

3.5 Competitive Inhibition helps to stabilize attentional focus

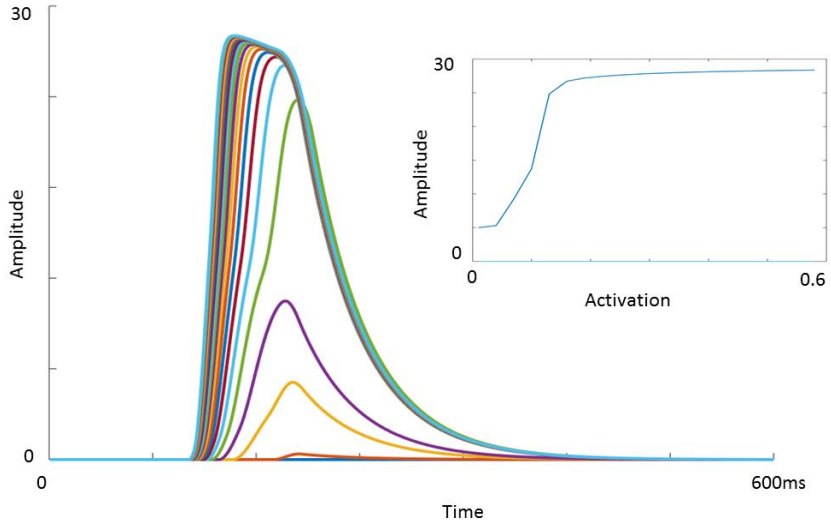

Figure 7: Illustration of the time course of activation of a neuron in the AM centered at the location of a stimulus for varying amounts of activation strength of a stimulus from .01 to .6 in increments of .03. The inset plots peak amplitude as a function of activation strength. The takeaway point is that a lock-on state is an attractor, such that many different values of strength map onto the same amplitude of activation.

The competitive inhibition between AM and IG neurons was developed to meet empirical benchmarks. A virtue of modelling biological systems is the ability to understand the value of particular mechanisms. In this case, RAGNAROC gives us the tools to understand that competitive inhibition improves the functionality of attention by increasing the stability of one or more lock-on states, allowing attention to be simultaneously deployed more easily. To illustrate the effect of competitive inhibition on the attentional state, Figure 9 compares the intact model to one in which the AM->IG inhibition has been removed. This preserves the center-surround inhibition, and the selective inhibition but AM nodes are no longer able to competitively block their own inhibition. The figure illustrates how the duration of lock-on states for two temporal proximal stimuli is compromised without the inhibition. With the inhibition intact, two lock-on states are able to form simultaneously with minimal interference because each protects itself from interference. However within about 50ms, the window of simultaneity has expired because the T1 lock-on starts to inhibit nearby activity in the AM. This makes it difficult for T2 to establish its own lock-on state. The time course of this transition from simultaneous to sequential attention is in agreement with behavioral data showing the onset of attentional inhibition following a T1 onset (Mounts 2000). The advantage of protecting the lock-on state of the T1 at the expense of the T2 is to reduce the volatility of attentional decisions in the case of dynamic or rapidly changing stimuli.

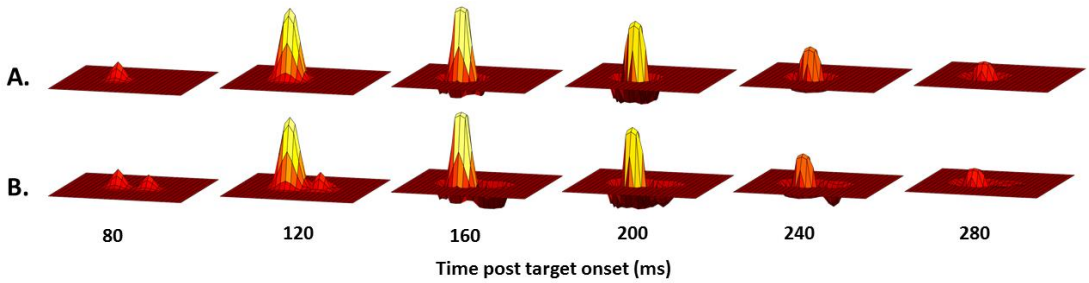

Figure 8. Evolution of activity in the attention map over time for a single stimulus (A), and the same stimulus accompanied by a stimulus of lower priority (B)

3.6 Mapping model activity onto measurable data

In order to compare the model against empirical benchmarks, it is necessary to map model activity to behavioral measures of accuracy and reaction time, as well as EEG correlates of attention such as the N2pc and Pd. Figure 10 illustrates which activity states in the RAGNAROC model are used for generating data.

3.6.1 Model configuration

For each experiment, salience values and attentional weightings are configured for different kinds of stimuli in the task. The attentional weightings are specified as a distinct uniform distribution of possible values for each kind of stimulus. For example in a salient-singleton attentional capture paradigm (Theeuwes 1992), the target stimulus has a bottom-up salience value of .15 and a range of attentional weightings from .17 to .37 in 12 steps of .018 while the salient distractor has a bottom-up salience of .3 and a range of weightings from .07 to .27 in 12 steps of .018. The model is run for all possible combinations of these weightings, (e.g. 144 total simulations in this example) that are then bootstrapped to form the simulated data set of an entire experiment. The bootstrap involves resampling 10,000 times according to a normal distribution (with mean centered at the median attention weight and a standard deviation of .75) under the assumption that there is trial to trial variability in the attentional set of the observers that is normally distributed. For this bootstrapping, the normal distribution is mapped onto the bins of the uniformly distributed priority values over a range of +/- 3 standard deviations. This bootstrapping determines both the simulated behavior (accuracy and RT, as appropriate) and the EEG traces for a given experiment. This bootstrapping is the only source of variability in the model when simulating EEG. For simulations of behavior an additional source of noise is added during the bootstrapping, as described below.

3.6.2 Simulating Behavior

The CGF model simulates the successful detection or response to a target with a thresholded accumulator. The accumulator sums the time course of activation of all LV neurons that are activated over baseline (.5) for the target. For every trial the area under the curve (AUC) is calculated for the entire time course of activation. A trial is considered accurate if this AUC exceeds a threshold value that is calibrated for each task but is not allowed to vary for different conditions within a task. This threshold is a free parameter fit for a given task to achieve a particular accuracy value in one baseline condition chosen for each task. In order to introduce more variability and make simulations less sensitive to the particular threshold value, each simulated trial is jittered with random noise. This is done by adding 15% of the baseline condition's average AUC times a random scalar (ranging from 0 to 1 from a uniform distribution).

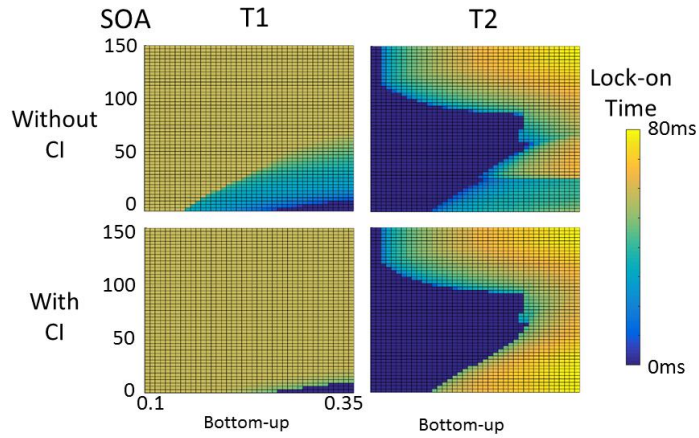

Figure 9 Illustration of the robustness of lock on states to each of two stimuli where the second stimulus varies in the strength of bottom-up strength value (horizontal axis) and temporal latency (vertical axis). Each point represents the total duration for which the AM neuron at the location of the stimulus (T1 or T2) is above threshold. The key takeaway is that without competitive inhibition (CI), T1 and T2 compete destructively at short SOAs such that neither elicits a robust lock-on state. With inhibition intact, both stimuli can achieve a lock-on state if presented with nearly identical onsets. However at longer SOAs T1 suppresses T2 enforcing a serialized deployment of attention. In this simulation T1 and T2 were presented 4 degrees apart and have a duration of 120ms.

Reaction times are simulated similarly for each trial in which the threshold was exceeded but with the AUC calculated at each time step cumulatively in order to get a time estimate for when a stimulus' activation value crossed threshold.

3.6.3 Simulating EEG

To translate from simulated neurons to EEG correlates is a hard problem that, in its most exact form, would require compartmental-level modeling of cortex and thalamic input including all of the synapses in each layer, a fairly complete understanding of the neuroanatomy for each

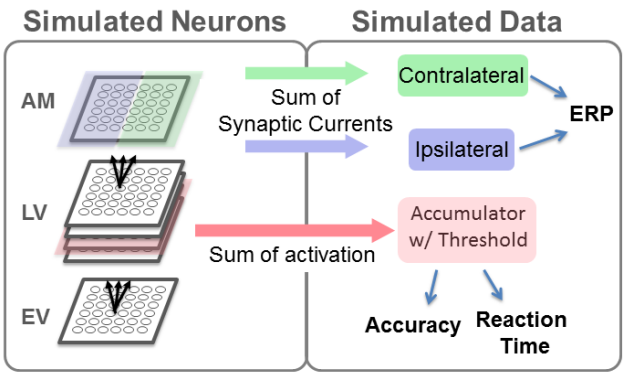

Figure 10 Activity within the CGF model is used to construct simulations of data in the form of behavioral accuracy, behavioral reaction time and EEG components. Behavioral data are extracted from the late vision area, which is assumed to drive the formation of memory representations and response decisions through mechanisms that are outside of the scope of the model. Simultaneously, synaptic currents within the attention map are measured to generate simulated ERPs such as the N2pc and the PD.

individual subject, and a model of the electrical properties of the tissue layers above the cortex (dura, fluid, skull, muscle, skin).

However, it is possible to make effective progress with a simpler model, given some starting assumptions to simplify the forward model of generating scalp potentials. Here, these assumptions are (1) that the attention map exists over a region of cortex situated in posterior-lateral parietal areas of cortex (2) that EEG potentials are largely driven by excitatory synaptic input on pyramidal neurons oriented perpendicular to the scalp (Nunez & Cutillo 1995) (3) that an increase in this synaptic current within the attention map produces (on average) a negative voltage at the scalp and (4) An additional weighting parameter that determines the relative contribution of excitatory and inhibitory synaptic currents. The advantage of such a simple model is that it provides fewer opportunities to overfit the observed EEG.

Given these assumptions, RAGNAROC simulates lateralized EEG components associated with attention by summing synaptic currents across each half of the AM, and taking the difference of those sums relative to the side of the visual field that a particular stimulus was presented on. This is analogous to the measure of potentials such as the N2pc and Pd, which are calculated as the difference in voltage between electrodes contralateral and ipsilateral to the side of the display containing a target (or a distractor in some cases).

The synaptic currents are computed separately for each neuron as its excitatory current, minus its inhibitory current, with a floor of zero. The intuition behind this implementation is that excitatory currents are the primary drivers of the large dipoles that are observable at the scalp,

and inhibitory inputs often shunt those excitatory currents by creating high conductance areas of the cell membrane closer to the soma.

The AM receives a uniform input to elevate all of the neurons above their resting potential. This provides a baseline level of excitatory current that is uniformly distributed across the attention map and therefore drops out during the subtraction of ipsilateral from contralateral. Activation or inhibition of nodes within the AM causes deviation away from this baseline level of current. When this current is summed across the halves of the attention map, laterally asymmetric differences in activation produce changes that are comparable to the N2pc and Pd components. Note that using the sum of currents across a hemifield to simulate voltage means that increases in current for some neurons might be effectively invisible to the simulated EEG signal if there are also corresponding decreases in current for other neurons in the same half of the attention map (Figure 11). Furthermore, any negativity in the simulated voltage difference between the contralateral and ipsilateral sides of the map could be caused either by an increase in activity in the contralateral side, or a decrease in activity on the ipsilateral side. It is important to remember that there are only two possible polarities of a component, positive or negative, but there are (many!) more than two neural processes that could result in a scalp potential at a given latency. Therefore one cannot uniquely ascribe a given functional property to an ERP on the basis of a given polarity/latency.

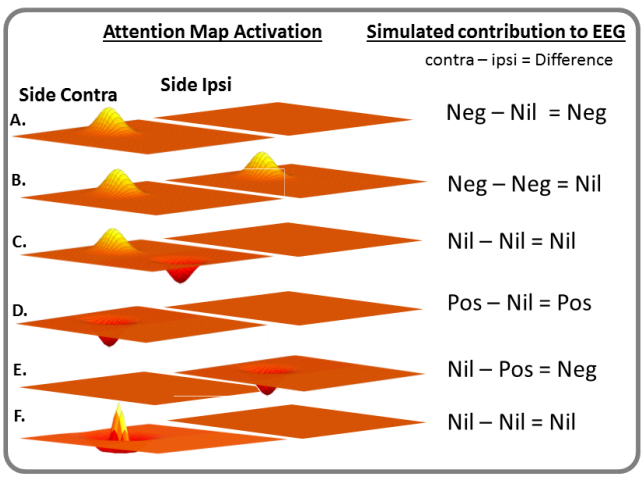

Figure 11 Illustration of how different patterns of activity on ipsi and contralateral sides of simulated cortex can summate to produce either positive, negative or nil voltage differentials. Note that in F, the currents generated by the activity in the peak is effectively cancelled out by the surrounding inhibitory surround producing an effective Nil in the contralateral side. Note that "Nil" in this context doesn't necessarily mean exactly 0, but sufficiently small that it is not detectable at a given level of experimental power. It is unlikely that two current sources would exactly cancel one another, but even if they are approximately equal the small difference may be impossible to detect.

This ambiguity in the interpretation of simulated EEGs

is not a shortcoming of the model, but rather reveals a complication inherent in the interpretation of all ERPs. This complication underscores the importance of understanding EEG signals at the level of their neural sources and emphasizes the role of computational models in understanding those sources.

3.6.4 Simulation of the N2pc

In the RAGNAROC model, presenting a lateralized target to the model produces a simulated ERP that resembles an N2pc (FIGURE 12a). The onset and peak of the N2pc is caused by the initial activation bump in the AM. When the lock-on state is established, the AM activates its neighboring IG neurons, which adds an inhibitory region in the immediate surround. When the central peak is surrounded by inhibition, the sum total of synaptic currents on the contralateral side of the AM nearly cancel out. Thus, while the RAGNAROC model is in general agreement with the theory that the N2pc reflects processes associated with spatial attention, it suggests a more specific temporal relationship, which is that the N2pc reflects, in large part, the processes of localizing a target prior to attentional deployment.

In this case, the end of the N2pc does not indicate the end of attention, but rather the onset of surround suppression. This simulation provides a straightforward explanation for the brief duration of the N2pc, which is typically brief and followed by a strong positive swing (Brisson & Jolicouer 2006)

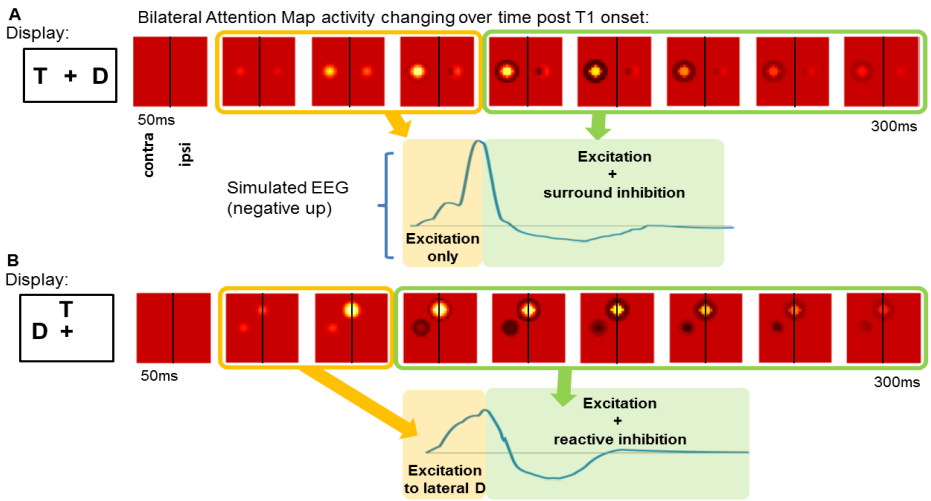

Figure 12 Illustration of how activation levels within the attention map produce simulations of N2pc and Pd components for commonly used experimental paradigms.

3.6.5 Simulation of the Pd

The Pd is an EEG component thought to reflect inhibition of distracting information in the visual field. In RAGNAROC, a Pd can emerge when there is an imbalance in the relative balance of excitatory and inhibitory synaptic activity, and this occurs in at least two ways. First, whenever two stimuli compete for attentional control and one of them loses, the AM is suppressed at the location of the loser (Figure 12b). This suppression reduces synaptic currents in the hemifield containing that stimulus and results in a net positivity in contralateral scalp electrodes. However, a Pd also occurs when the surround suppression encircling an attended stimulus is large enough that it causes a net reduction in current for that half of the visual field. This imbalance would be reflected as a Pd trailing an N2pc, and could occur even in the absence of a suppressed distractor (Figure 12a, Töllner, Zehetleitner, Gramann, & Müller 2011).

4. **Simulations of Empirical Constraints**

4.1. Constraints in model development

As in previous papers (Tan & Wyble 2015; Swan & Wyble 2014) the model is parameterized according to a set of extant findings in the literature. Once the model is able to simultaneously accommodate those findings with one set of fixed parameters, it can be used to generate insights about the underlying system and testable predictions for future work. The philosophy of this approach is to allow a large number of empirical constraints to inform the model's design, with as little parametric flexibility as possible. Here we list a series of behavioral and electrophysiological findings that we consider to be crucial for defining the functionality of reflexive attention. Each of these findings is simulated with the same set of parameters, except for the configural parameters described in 3.2.3. The supplemental describes the exact set of parameters for each simulation.

The fits to the existing data are evaluated for their qualitative similarity to the data given the limited number of parameters and the large diversity of experimental paradigms that provide the constraints.

4.2 Behavioral constraints:

1. Covert spatial attention is triggered rapidly in response to a target or highly salient stimulus. This effect is measurable as an enhancement of accuracy and reduced reaction time for stimuli presented just after a cue, at that same location, and a time course that peaks at about 100ms SOA (Nakayama & Mackeben 1989). Note that this transient form of attention is short lived even when the cue stays on the screen. It is difficult to precisely estimate the duration of this effect because it is followed by slower, more volitional forms of attention that sustains the attentional effect to differing degrees in differing paradigms. However there have been consistent findings of enhanced perception at brief cue-target (Yeshurun & Carrasco 1999; Müller & Rabbitt 1989; Cheal & Lyon 1991) or target-target intervals that attenuate at longer cue-target intervals. Targets elicit such attention as well (Wyble, Potter, Bowman 2009). RAGNAROC simulates the transient attention effect of Nakayama and Mackeben (1989) as a

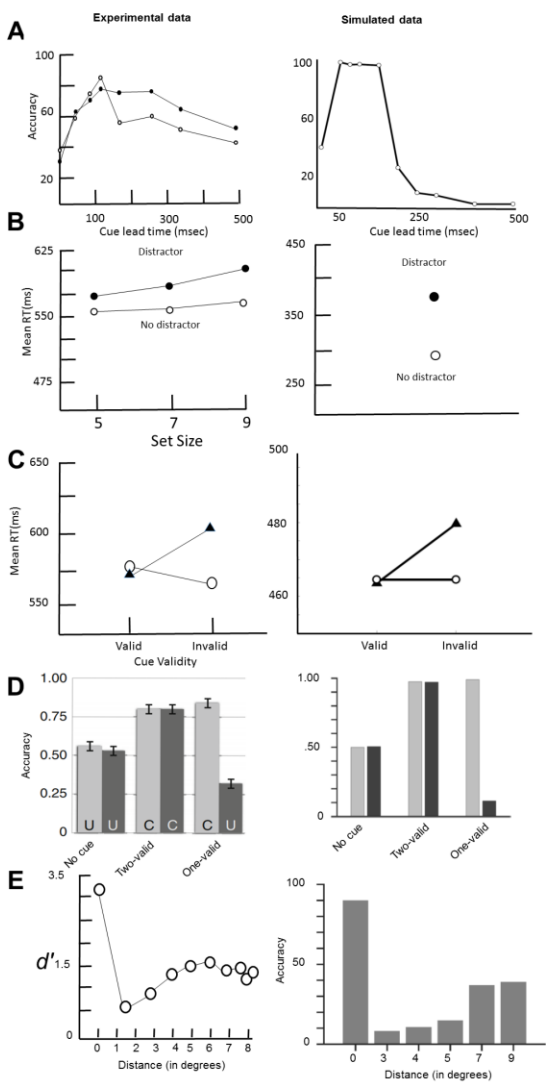

Figure 13 Behavioral constraints and simulations. (a) Accuracy of reporting a target indicates the transient nature of reflexive attention (Nakayama & Mackeben 1987). (b) Reaction time to report a shape singleton target is increased in the presence of a salient color distractor (Experiment 3 from Theeuwes 1992). (c) A singleton only affects the reaction time to report a target only if it matches the type of target, in this case an onset target (Folk Remington & Johnston 1992; Experiment 3). (d) The benefit of a valid cue, relative to a no-cue condition, is not diminished when two cues are used, suggesting simultaneous deployment of attention to two locations with minimal cost (Bay & Wyble 2014). (e) Accuracy of reporting a second target is affected by proximity to a preceding target with a spatial gradient (Mounts 2000). Note that the empirical data is reported as d', but model accuracy is reported as accuracy since it lacks a mechanism for guessing.

brief window of elevated accuracy in reporting a target when it follows anther stimulus at the same location. (Figure 13a)

2. This rapid deployment of attention is reflexive, which means that it is vulnerable to capture by a non-target stimulus that is either highly salient (Theeuwes 1992) or contains a target-defining attribute (Remington, Folk & Johnston 1992). This reflexive form of attention *cannot be prevented* from being deployed to locations that are known to always be task irrelevant (Lamy, Leber & Egeth 2004; Krose & Julesz 1989). Also, highly-salient distractors will trigger this form of attention regardless of instruction, or lengthy practice sessions (Theeuwes 1992). RAGNAROC simulates the attentional capture effect of Theeuwes (1992) as a longer reaction time for a target in the presence of a distractor (Figure 13b)

3. Attention can be biased towards stimuli containing certain features or attributes, provided that there are well-learned, cognitively accessible distinctions between target-defining features and other stimuli (e.g. letters can selected among digits but an arbitrary subset of letters cannot be efficiently selected from other letters without substantial training; Schneider & Shiffrin 1977). This target-defining attentional set is implemented across the entire visual field such that, for example, establishing a control setting for red at one location prioritizes red at all locations (Zhang & Luck 2009). RAGNAROC simulates attentional set as capture costs that are mediated by changes in task-set from Folk Remington Johnston (1992). See Figure 13c.

4. Attention can be deployed to two or more locations at the same time when stimuli are presented in parallel, but behaves more like a spotlight when targets are presented sequentially (Bichot et al. 1999; Dubois Hamker & VanRullen 2009; Bay & Wyble 2014). RAGNAROC simulates divided attention as an attentional benefit that is similar in size regardless of whether one or two locations are cued (Bay & Wyble 2014). See Figure 13d.

5. Presenting a cue or target at one location causes subsequent targets presented at spatially proximal locations to be harder to perceive. This suppression is diminished with increasing spatial distance (Mounts 2000; Dubois et al. 2009). RAGNAROC simulates attentional suppression surrounding an attended region using two sequential targets as in Mounts (2000). See Figure 13e.

6. In the presence of a target, inhibition is localized at the spatiotopic location of non-target stimuli and only weakly affects empty regions of the visual field. Thus, probe stimuli will be harder to perceive when they occur in the locations of former distractors, rather than blank areas (Cepeda et al. 1998, Gaspelin, et al. 2015; Figure 2c). RAGNAROC simulates selective suppression of attention at locations containing stimuli as in Gaspelin, et al. (2015). See Figure 14.

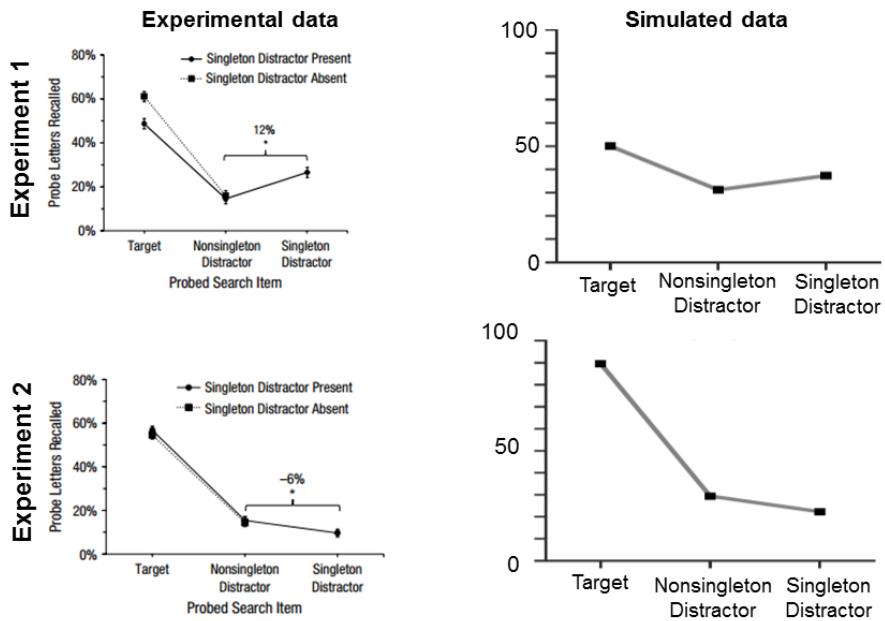

Figure 14 Experimental data from Gaspelin Leonard & Luck (2015) alongside simulations. In the panel (a), the data are from a condition in which the target was a shape singleton, but the subject did not know which of two shapes would be the singleton or the distractor. In this condition, report of the probe at the singleton color distractor was elevated. The model simulates this effect (b) as the result of weaker top-down attention which allows the salient distractor to successfully trigger attention on a large enough proportion of trials that report of the probe letter at that location is elevated. In the panel (c), the participant knows exactly what shape will contain the target. This is simulated (d) by adopting stronger top-down settings which allows the target to inhibit the distractor on nearly every trial, such that the probe letters at that location are reported less often across the entire block of trials.

4.3 EEG Constraints

Data concerning the N2pc and Pd components will be taken as constraints as well.

4.3.3 Specific EEG Constraints:

1. Presenting a target in either hemifield produces a brief negativity in EEG recorded on the contralateral, posterior side of the scalp called the N2pc (Luck & Hillyard 1994; Eimer 1996) or PCN (Töllner, Zehetleitner, Gramann & Müller 2011). This negativity typically peaks at about 250ms after target onset and is observed whether targets are accompanied by distractors or not (Tan & Wyble 2015). RAGNAROC simulates this effect as a contralateral negative voltage for a target on one side of the visual field (Eimer 1996). See Figure 15a.

2. Multiple targets in the same location in immediate succession will produce an N2pc only to the first target in the sequence. That N2pc resembles the N2pc to a single target in amplitude and duration. Note that this is true even for trials in which both targets were reported, so the failure to find a difference between the one and two-target N2pc's does not reflect a failure to attend or encode the second target (Tan & Wyble 2015; Callahan-Flintoft et al. & Wyble 2017; Callahan-Flintoft & Wyble in press). When the two targets are presented in different locations of the visual field, there will be an N2pc to each of them in turn (Tan & Wyble 2015). This effect is only present when the two targets are presented within roughly 150ms. At longer separations a second N2pc is observed, regardless of whether subjects know that the second target will appear in the same location as the first (Callahan-Flintoft JEPG; See Figure 15c).

3. Multiple N2pcs can be evoked in rapid succession (e.g. at 10-100ms intervals), with no delay. When presenting a lateral target at intervals of 10, 20, 50 and 100ms relative to a preceding target, an N2pc is evoked with a target-relative latency that is very similar (i.e. within 10ms) to that evoked by the first target. This finding indicates that deploying attention to one target does not delay the onset of attention to a second target (Grubert, Eimer 2014, Experiment 1). See Figure 15d.

4. The N2pc/PCN is often followed by a positive contralateral potential called the Pd (Hickey et al. 2009; McDonald, Green, Jannati & DiLollo 2012) or Ptc (Hilimire, Mounts, Parks & Corballis 2010). This positivity has been particularly associated with the occurrence of a highly salient lateral distractor, although this positivity can occur without such a distractor (Töllner et al. 2011; Hilimire, Hickey & Corballis 2011). RAGNAROC simulates this effect as a positive voltage after the N2pc for a lateralized target (Töllner, et al. 2011). See Figure 15e.

5. A laterally presented salient distractor can produce an N2pc, and this N2pc will be larger if the distractor is presented without an accompanying target. (Kiss, Grubert, Petersen, & Eimer 2012; Hilimire, Hickey & Corballis 2012; McDonald, et al. 2012). RAGNAROC simulates this effect as a negative contralateral voltage after a lateralized distractor (McDonald et al. 2012). See Figure 15f.

6. Specificity of the attentional set affects the degree to which targets and distractors produce N2pcs. When the task set does not predict a specific stimulus shape (e.g. "find the shape

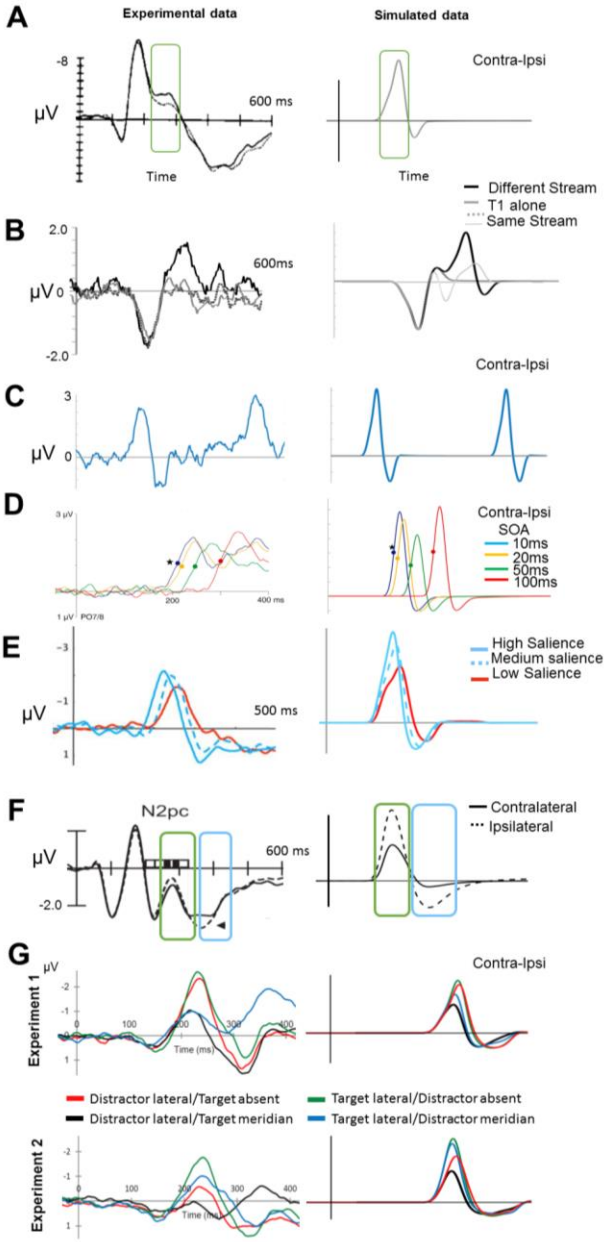

Figure 15 EEG constraints and simulations. (a) A laterally presented target causes a brief contralateral negativity, even if it has a long duration (Eimer 1996; green window added to emphasize the time frame of the N2pc). (b) The N2pc to a second target is muted if it occurs very soon after and in the same location as a preceding target (Tan & Wyble 2014). (c) The second N2pc is of normal size if the two targets are far apart in time (Callahan-Flintoft & Wyble In Press). (d) When two highly-salient, unmasked targets are presented in rapid sequence at different locations, the N2pcs to the second target is not much delayed (Grubert, Eimer 2014). (e) The N2pc is often followed by a positivity, particularly when there are nearby salient distractors, or in this case when the target itself is highly salient (Töllner, et al. 2011). (f) A laterally presented distractor can trigger an N2pc (McDonald et al. 2012). (g) With a highly predictable or salient target class, the distractor produces a minimal N2pc and has little effect on the target's N2pc. When the target set is less certain or salient, the distractor has a greater effect on the target N2pc (Hilimire & Corballis 2014).

singleton"), the distractor induces an N2pc (Hickey, et al. 2006; Burra & Kerzel 2013), and is reduced by the presence of a central target (Hilimire, et al. 2011; Hilimire & Corballis 2014; Figure 3c). Also, *unpredictable task* sets allow the target N2pc to be reduced by a salient distractor (Hilimire & Corballis 2014). When the task set is a *predictable singleton*, then distractors produce a much weaker N2pc, and the target N2pc is unaffected by distractors. RAGNAROC simulates this effect as a specific ordering of N2pc amplitudes for different stimulus configurations across two different specificity manipulations (Hilimire & Corballis 2014). A similar result is obtained when the task is manipulated such that the distractor is of higher or lower salience than the target. For example, when the task is to detect a form singleton, a highly salient color singleton will reduce the target's N2pc, but this is not true when the task is to detect a color singleton, and the distractor is a form singleton (Schubö, 2009). See Figure 15g.

5.0 Discussion: what have we learned?

The RAGNAROC model describes a set of neural mechanisms that explain how attention reflexively responds to new visual input, and makes rapid decisions about which locations to enhance and which to suppress. The decisions are mediated by attractor states and competitive inhibition, that help to ensure that the decisions are stable and accurately targeted at the correct location. It is argued that this reflexive attentional system plays a key role in many experimental paradigms and also provides a generative model of EEG components thought to be related to visual attention.

As a model, RAGNOROC is both an architecture, as well as a specific set of parameters that are calibrated against several decades of data that specify the time course of reflexive attention. Presumably, this time course reflects an adaptation against demands on the attentional system imposed by other constraints of the visual system. For example, the operation of reflexive attention has to occur within the time span of a visual fixation, while the eye's position is relatively stationary, to take advantage of the shared spatiotopy within the ventral stream. With the model developed and parameterized, the next steps are to use it as a tool to learn more about the underlying system, and to assert a series of testable predictions that will measure the validity of the model relative to the human system. We begin with a series of lessons that were learned through the model's development.

Lesson 1. Attention does not draw a clear distinction between targets and distractors.
Experimental paradigms typically designate stimuli as targets or distractors and it is tempting to assume that the mind of the participant adopts the same crisp distinction. However, the visual system is presumably maintaining vigilance for all possible kinds of stimuli (e.g. consider whether the participant would react to an unexpected light flicker in an experimental context). To accomplish this feat of general vigilance, even during a highly explicit visual attention experiment, the visual system must evaluate all stimuli to determine which, if any, should be attended. RAGNAROC assumes that stimuli compete with a common currency of relative priority. Therefore, within the AM, for at least the first two hundred milliseconds of processing, there is no categorical distinction between targets and distractors. Rather, all stimuli compete, and attention is deployed to the winners, and the losers are suppressed. The implications of this become more interesting when we think about tasks with multiple targets of varying priority. In such cases, the reflexive attentional system may allocate attention to the highest priority target and treat the other targets as a distractor.

Lesson 2: What does the N2pc/Pd complex reflect?
It is often tempting to assign concrete roles to ERP components. For example the N2pc is thought to reflect attention evoked by a target while the Pd reflects inhibition evoked by a distractor. However, as we note above, there are cases in which targets produce Pd components and distractors produce an N2pc, and it is difficult to interpret such findings when components are considered as isomorphic correlates of cognitive processes.

This modeling approach illustrates why it is important to consider that there is a many-to-one mapping between current sources and ERPs. The neutrality of a scalp potential at a given latency could indicate a period of neural inactivity, but it could also be the case that there are strong underlying dipoles which happen to cancel one another out at that particular moment in time. It is therefore crucial to ultimately understand ERPs at their source level if we want to fully incorporate them into the inference process.

Models such as RAGNAROC make it easier to understand that there are several different ways in which neural activity evoked by a stimulus could lead to a negativity or positivity. This perspective is generally consistent with the typical interpretation of these components, because, for example the N2pc is more often associated with targets than distractors. However the model provides additional insights that are unavailable from the isomorphic view. For example

RAGNAROC illustrates why the N2pc is often followed by positive rebound after about 100ms, even though the stimulus stays on the screen (Brisson & Jolicouer 2007). Furthermore, the model explains why this rebound can even produce a trailing positivity as target salience is increased (e.g. Tollner et al. 2011).

Lesson 3. Experiment outcomes are an admixture of different trial outcomes

In RAGNAROC trial-to-trial variability in the simulations (see section X) accounts for uncontrolled sources of variability and allows us to generate simulations of accuracy from dichotomous outcomes. The model clarifies that differences in the magnitude of an effect could reflect proportions of different outcomes, a point that was also emphasized by Zehetleitner et al. (2013). For example, a given experiment which exhibits a weak attentional capture effect by a salient distractor, may in fact have a very strong capture effect on a minority of trials and no effect on the other trials. Likewise, a manipulation that produces a stronger N2pc in one condition than another may not actually produce stronger N2pcs in individual trials, but rather a larger proportion of moderate N2pcs.

Lesson 4. Understanding reaction time costs in attentional capture

The term attentional capture typically refers to a behavioral phenomenon of slowed responses to a target due to the presence of a distractor, but what exactly causes the reduced performance? In the RAGNAROC model there are three possible outcomes when a target and at least one distractor are presented together. First, the target might trigger attention and suppress attention to the distractor(s); second, the target and distractor might trigger attention together; and third a distractor might trigger attention and suppress attention to the target.

RAGNAROC predicts that RTs would be fastest when the target is attended and the distractor is suppressed because this reduces interference caused by distractor processing. When both the target and the distractor are attended (i.e. simultaneous attention), RTs to the target would be slightly slowed because simultaneous lock-on states, while stable, are often slightly smaller compared to a case in which the target is clearly dominant. The final case produces the slowest RTs because the target is not enhanced by attention which reduces the strength of evidence for that target.

The RAGNAROC model predicts that any given experimental block of an attentional capture experiment is composed of a combination of these three outcomes with

proportions determined by the relative priority of the targets and distractors. Thus, even in a paradigm that has minimal evidence of attentional capture at the group level, the distractor may nevertheless trigger the deployment of attention on a subset of trials.

Lesson 5. Architectural answers to the bottom-up/top-down attentional capture debate

One of the most enduring discussions in the attentional literature is whether bottom-up stimuli are always able to capture attention, or are top-down attentional control signals able to override bottom-up salience. Driving this debate are classic findings that some kinds of distractors elicit capture costs consistently, even though they are never task relevant (Theeuwes 1991). In others studies, capture effects seem to be entirely contingent on top-down settings (Folk, Remington & Johnston 1992). This debate has continued without a clear resolution.

In the model, there is a sense to which bottom-up selection occurs prior to top-down guidance. Differences in salience are represented at the junction between EV and LV, and differences in task-related attentional set are represented between the LV and AM. This means that a difference in physical salience will often manifest in the AM prior to a difference in task relevance. Figure 16 compares the time course of activation bumps generated by highly-salient, irrelevant stimuli, to less-salient but task relevant stimuli. Thus, the model exhibits a form of precedence that is in general agreement with Theeuwes Atchley & Kramer (2000). Moreover, this result is not due to specific parameter values, but rather is an outcome of the model's feedforward architecture. Since salience differences are thought to be processed earlier in the hierarchy (Zhaoping 2002), highly salient stimuli will tend to activate their corresponding LV nodes earlier than less salient stimuli. However, this temporal advantage does not mandate that salient stimuli will always be attended first, since a strong top-down weighting can ensure that a task-relevant, but low-salience stimulus will cross threshold first within the AM.

Lesson 6. Architectural answers to the singleton-detection mode debate

Another crucial issue in the attentional capture debate has been the idea that singleton detection mode allows the system to select unique information for any attribute dimension (e.g. the red item among green items). The advantage of such a mode is that it does not need to be configured in advance for a specific value, preferring equally a red among green item or a green among red items. It has been suggested that subjects use singleton detection when looking for a target that has a unique property, such as a color or form singleton(Bacon & Egeth 1994). However, a

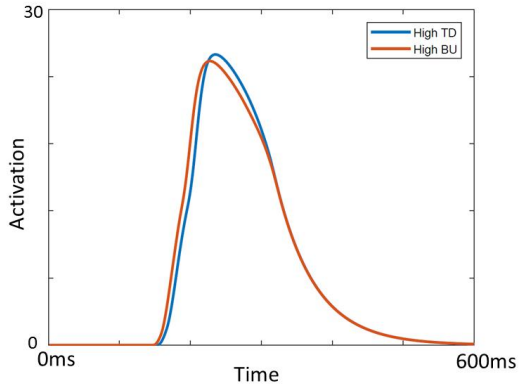

Figure 16. Simulation of the time course of attention map activation for two stimuli that have similar attentional priority, except that one has high salience and a low bottom-up weighting (BU: .2, TD: .15) while the other has the reverse (BU: .15, TD: .2). Of note, despite the higher peak of the high-TD stimulus, the high-BU stimulus has an earlier peak. This is an overlay of two traces; the two stimuli were simulated separated and had the same onset.

limitation of singleton detection mode is that it cannot be directed towards a specific dimension. Thus, using singleton mode to detect an oddball shape will also prioritize an oddball color.

Models like RAGNAROC make singleton-detection mode straightforward; it is simply the lack of strong top-down set, that allows stimuli which stand out in any feature dimension to dominate the computation of attentional priority. This explains the observation that singleton detection mode cannot be specific for a given dimension. Moreover, since singleton mode is effectively the absence of a top-down set, it is the default search policy (Bacon & Egeth 1994; Lamy & Egeth 2003).

Lesson 7: Architectural answers to the distractor suppression debate

Competing accounts of inhibitory control in reflexive attention pit the notion of a suppressive surround (Mounts 2000; Cutzu & Tsotsos 2003; Tsotsos 2011) against accounts in which inhibition is selectively deployed to distractor locations (Cepeda et al. 1998; Gaspelin et al. 2015). RAGNAROC illustrates how readily a single model can exhibit both behaviors depending on the paradigm that is being used. A spatial gradient in AM->IG connectivity simulates the surround inhibition effect of Mounts (2000). However, within that surround field, inhibition is selectively applied to the locations of stimuli as a function of their spatiotopic distance to the lock-on state.

The RAGNAROC model thus explains why the Mounts paradigm and other paradigms which also surround the initial target with distractors such as Cutzu & Tsotsos (2003) were so successful in eliciting the inhibitory surround, while other paradigms have no clear pattern of inhibitory surround (Wyble & Swan 2015). In the Mounts paradigm the first target is surrounded by a large number of simultaneously presented distractors. This display is followed immediately

by a second display that is used to probe the state of attention. According to the RAGNAROC, the dense field of distractors in the first display plays a key role in revealing the shape and size of the inhibitory gradient, since each of those distractors will elicit inhibition in their location, and this inhibition will affect the following target. For paradigms in which the initial target is not surrounded by a dense field of distractors (e.g. Wyble & Swan 2015), the IG neurons in the large area surrounding the target are not stimulated by input from the LV and therefore the only inhibition that is actually expressed in the AM is that immediately surrounding the target's lock-on state.

Lesson 8. The Competition for attention can result in a tie.

The conventional notion of spatial attention is that it behaves like a spotlight, focusing on only one location at a time. This provides a ready explanation for cueing costs and attentional capture effects, since attention directed at one location can therefore not be at another. However there is also mounting evidence that attention can be deployed simultaneously at two distinct locations (Bay & Wyble 2014; Bichot Cave & Pashler 1999; Kyllingsbaek & Bundesen 2007;Kawahara & Yamada 2012). The RAGNAROC model provides an explanation for these seemingly incompatible sets of findings. The circuitry in the attention map is designed to encourage a competition between competing items, however it is a competition in which there can be multiple winners, which allows simultaneous attention for two stimuli of approximately equal priority.

Lesson 9. Reflexive Attention may have almost unlimited capacity.

A common assumption of cognitive theories is that attentional limitations play a key role in determining performance in complex tasks. However, attention is a broad concept and it is often difficult to understand exactly what forms such limits take. In many cases, attention is equated with the ability to "process" information, which includes some mixture of identification, decision making, response generation, and memory encoding.

In RAGNAROC, a very specific definition of attention is employed, which is the reflexive enhancement of feedforward excitation at a given location in the visual field deployed reflexively in response to a stimulus. In the model, this form of attention has no clearly defined limit in terms of the number of attended locations since the model only simulates an increase in gain at a given location and that increase could occur at an almost arbitrary number of locations, especially for simultaneous stimuli. The variability in the representations of multiple items

makes it unlikely that more than a few locations would be attended in this way, but it is possible in principle. Thus the model proposes that at least the initial stage of attentional selection may operate without strict limits. Of course subsequent stages of processing are surely limited. For example, even if four stimuli produced simultaneous lock-on states, this does not necessarily mean that all four of those stimuli could be rapidly encoded into memory or responded to in a speeded paradigm. Parallel selection at early stages does not necessarily entail parallel processing at later stages.

Lesson 10. Attention can be suppressed without suppressing the representations

It is often suggested that distractors are inhibited during the selection of target information. Models like the RAGNAROC elucidate an important distinction between suppressing the representation of a stimulus, and suppressing attention towards a stimulus. Suppressing a stimulus representation entails direct inhibition of the neurons that represent the attributes and features activated by that particular stimulus (e.g. Reynolds & Heeger 2009; Beuth & Hamker 2011). On the other hand, suppressing attention at the location of a stimulus, as in RAGNAROC does not directly affect the representation of that stimulus, though it may do so indirectly by preventing that item triggering attentional enhancement.

It is, unfortunately, difficult to clearly distinguish between the two implementations of suppression at a behavioral level since they can both have a similar outcome. However this difficulty illustrates an advantage of the modeling approach, since models are able to clarify distinctions of implementation that are not obvious when considering only the functional implications through observations of behavioral data. Moreover, the model illustrates why it would be advantageous to suppress attention, rather than the representation. Suppressing the representation of a stimulus would require an enormous number of long-range connections to deliver inhibition to the appropriate neurons throughout the set of LV maps. Suppression of attention is much simpler to implement, since the inhibitory circuitry is entirely self-contained within the AM.

6: Review of other theories

There is an enormous literature of theories and models of visual attention. With a term as broad as attention, it is impossible to identify a canonical set of effects that all models should strive to explain and therefore a point-by-point comparison with other models is impossible. Speaking broadly, the RAGNAROC model occupies a relatively unique position in the modeling

landscape by focusing on the neural mechanisms associated with reflexive attention at a short time scale. The model is also unique in explaining the N2pc/Pd complex. Other models have touched on the idea of simulating the N2pc (Fragopanagos, Kockelkoren, & Taylor 2005) but have not provided a complete picture of neural firing that would translate into a lateralized potential, nor linked to a diversity of findings.

This review will cover a subset of models that focus on the mechanisms of spatial attentional selection. This excludes, for example, models of the attentional blink, which typically focus on how processing of information affects attentional selection.

Models inspired by neurophysiology. There is a family of models and theories of visual attention inspired by single unit neurophysiology in monkeys. Some of the research in this domain explores the properties of attention in spatial and feature domains. For example the normalization model of Reynolds & Heeger (2009) proposes that the neural response to any given stimulus is downweighted by the activity of nearby stimuli. Thus, when one stimulus is attended, other stimuli in the vicinity will evoke less activity, all else being equal. The normalization model provides a very simple and neurally plausible mechanism for the effects of attention at the level of single-unit data. Beuth & Hamker (2015) provide an even more detailed account of how attention can be mediated at the level of cortical representations. Such models interface directly with single-unit data from a variety of cortical areas, although they do not explain the decision-making aspect of attention that is the focus of RAGNAROC. Instead, attention is directed by other mechanisms that are outside the scope of those models and those approaches are therefore complementary to this model. However we consider it an open question as to whether the suppression effects described above are best explained at the circuit level within subordinate cortical areas, as in Reynolds & Heeger (2009) and Beuth & Hamker (2015) or at a superordinate level.

These findings and theories support the assumption that attention acts in part by enhancing neural responses to information at early levels of visual processing. The most well-known of such models is the idea of biased competition, in which stimuli compete with one another for representation in a neural system with overlapping receptive fields and the competition is biased in favor of neurons that respond preferentially to task-relevant information (Desimone & Duncan 1995). There is an important point of correspondence between the biased competition (BC) and the RAGNAROC accounts, which is that both incorporate an initial period of non-

selective processing before the deployment of attention begins. In the RAGNAROC, this is the time during which information feeds forward through the hierarchy to reach the AM. The time at which attention begins is reflected by the inflection in the LV trace of Figure 6b. The RAGNAROC and BC models differ in the mechanism of enhancement, since the BC implements attention as a contraction of receptive fields around the target stimulus. The difference is key because at the core of BC is the idea that representational space is a limited resource, which strongly limits the ability to attend to multiple locations at once. In RAGNAROC, this first stage of attention has fewer limits and thus can be deployed to multiple locations in certain situations.

Theory of Visual Attention. The Theory of Visual attention (Bundesen 1990) is a mathematical abstraction of the process of attending to and perceiving a stimulus as a function of the attributes of multiple stimuli.. In TVA, there are two ways to prioritize certain kinds of information selection: *filtering*, and *pigeonholing*. *Filtering* involves upweighting the priority for certain features, which increases the rate at which stimuli possessing those features attract attention. This is generally equivalent to attentional control settings in RAGNAROC. The *pigeonholing* mechanism relates to how efficiently certain kinds of information are categorized, which allows them to be reported and remembered. The TVA model thus represents two distinct types of attentional control setting, which might also be described as key feature and response feature (Botella, Barriopedro, & Suero 2001). The RAGNAROC model differs from the TVA model in that it provides a more complete model of the neural mechanisms associated with the computation and use of priority to direct spatial attention. The TVA model, on the other hand, provides a concise mathematical formulation of how two different kinds of filters interact to facilitate perception. A Neural implementation of TVA has been proposed (Bundesen et al. 2011), however it is less clear how such a model would scale up to a full working specification, since it requires a large scale winner-take-all implementation to complete attentional selection. Given the similarity of the role of priority within the two models, it could be fruitful to consider the RAGNAROC as a neurophysiologically plausible front-end for computing the priority-based competitive selection process and the TVA as a clearer specification of subsequent processing.

Guided Search. To better understand a number of complex patterns in the visual search literature, the Guided search model (Wolfe 1994) simulates how top down goals interact with bottom up salience signals to determine likely target locations. Like RAGNAROC, this model attempts to explain how the visual system mediates the balance between salience and top-down

relevance. Its focus is on a longer time scale than reflexive attention, and incorporates both overt and covert forms of attention. RAGNAROC is complementary to this model, by explaining attentional dynamics at a short time scale, and with a greater emphasis on inhibitory processes.

Feature Map Models. Another class of models simulates spatial attentional effects across sheets of neurons corresponding to different visual features. Perhaps the most canonical of such models is the salience model of Koch & Ulman (1985), which is architecturally quite similar to RAGNAROC. A descendent of this model is often invoked as a benchmark in computer vision algorithms (Itti Koch & Niebur 1998).

In such neural sheet models, feature detectors in multiple channels (i.e. luminance, color, motion flicker) converge at a master salience map that ultimately makes decisions about where attention will be deployed using a simple winner take-all mechanism, coupled with a form of memory that erases salience values at recently-visited locations. Like RANGAROC, this model uses salience as a common currency across all stimuli in the visual field, and would be able to simulate capture effects. The Itti Koch & Niebur (1998) model has been foundational in understanding how simple mechanisms can reproduce complex gaze behavior when iterated over many distinct feature dimensions and levels of scale. Also, because the Itti et al. model simulates responses to pixelwise visual data, and can be compared against visual fixation data from human subjects, it set the stage for a generation of further modeling efforts.

Itti et al. and the RAGNAROC model are primarily concerned with phenomena at different time scales. The former is intended as a model of gaze behavior on time scales of a second or more involving multiple fixations. RAGNAROC is developed to understand how covert attention deployment is computed anew with each visual fixation or significant change to the visual display. Moreover, the salience map in Itti et al. lacks the inhibitory mechanisms to suppress distractors without first attending to them. The two models are thus complementary; they operating at distinct time scales, emphasize different kinds of processes and simulate fundamentally different kinds of data.

Other models provide more direct simulations of neural processes of enhancement in neural sheets. The Selective Tuning model by Tostsos (1995; 2011) implements a form of inhibition in which detection at an upper level of the hierarchy produces surround inhibition at earlier layers of the hierarchy. This model is perhaps the most well-formulated attention model that has ever

been proposed since it proposes a gating control circuitry that allows information to be effectively linked across differences in spatial invariance. Selective Tuning would hit several of the benchmarks described here, but does not propose a reactive control circuit. It applies inhibition in a region surrounding a target, irrespective of the presence of distractors. Moreover, in the full architecture, decisions to deploy attention are made independently for different stimulus dimensions and it is not precisely formulated how cross-dimensional competition between stimuli would be implemented at the time scale of reflexive attention (see p121, Tsotsos 2011).

There is a variety of models by Grossberg and colleagues based on the overarching principle of adaptive resonance. One of the more recent of this series have proposed neural architectures for attentional control using an attentional shroud (Fazl, Grossberg & Mingolla: cognitive psychology 2009), which is essentially a means to delineate the boundaries of an object and then ensure consistent focus on that object during learning. The notion of a shroud is not entirely unlike the AM described here, which also has the ability to deploy attention in an arbitrary shape determined by the boundaries of the stimulus. The shroud model does not contain an explicit mechanism for inhibiting distractors, and its success in attentional focus in very complex displays could be offered as an example that distractor inhibition is not necessary. However in terms of hitting the empirical benchmarks described here, it is not clear how well inhibitory effects would be simulated.

The class of interacting neural sheet models includes a large number of others that make important contributions to understanding particular empirical phenomena. For example, Zirnsak, Beuth & Hamker (2011) simulate the temporal dynamics of attentional competition in response to one or more stimuli; Lanyon & Denham (2004) simulate eyegaze during visual search as a product of interacting attentional systems. Another class of models simulates how neglect symptoms arise from damage to the attentional system (Heinke & Humphreys 2003) at longer time scales than are considered here.

7. Predictions:

RAGNAROC is part of an ongoing investigation that involves a cyclic iteration between theory and experiment. Driving this cycle are *a-priori* predictions, which provide a roadmap for future experimental work to diagnose the model's validity. By publishing these predictions in advance of testing them we minimize the file drawer problem, which occurs when model tests are

selected for publication after the results are known. Furthermore, our goal here is to specify an ambitious set of predictions, with the goal that some of them should be inaccurate. Since all models, being abstractions of the real system, are wrong by definition, the prediction/testing cycle should be most efficient when there is a mix of true and false predictions. True predictions give evidence that the model has at least some resemblance to the underlying system. The true predictions are helpful in building confidence but it is the false predictions that are truly valuable, for they indicate where the model is inaccurate, and thereby guide development of the theory.

7.1 Competition within the attention map

These predictions concern the essential architecture of the model. Failure to validate them would require at a minimum, significant parameter or architectural changes. In RAGNAROC, the competition for attention exists between all stimuli, and the priority values of the stimuli are the common currency with which they compete. Since the attention map does not represent the distinction between targets and distractors, the following predictions should obtain:

Prediction 1. Lower priority targets will elicit AM suppression

In the RAGNAROC model, input to the attention map does not distinguish between targets and distractors. A counterintuitive prediction of this assumption is that when a display contains two targets with sufficiently different priority values, the lower priority target will lose the completion and be treated as a distractor. This would mean that it elicits a weak N2pc when presented laterally, and a clear Pd component. In terms of behavior, the location of the low-priority target should exhibit the same lower probability of probe letter reporting as the salient distractors of Gaspelin et al. (2015). Target priority could be manipulated either by varying the salience of targets or their proximity to the task-defined attentional set in some feature dimension, such as color (Becker Folk & Remington 2013).

Prediction 2. Higher priority distractors will more often elicit a lock-on state

One of the most fundamental predictions of the RAGNAROC is that all stimuli are evaluated by the attention system prior to the deployment of attention, which occurs after a competition is resolved within the attention map. Therefore, if a display consists of only distractors of three or more clearly discernable levels of salience (e.g. by adjusting their relative luminance), the distractors will elicit N2pc and Pd components as if the most salient distractor were a target and the next most salient distractor were the key distractor in the additional singleton paradigm. The

most salient distractor will also capture attention resulting in improved accuracy and reduced reaction times for probes (e.g. Gaspelin et al. 2015) at its location. Conversely, probes at the second-most salient distractor location will be less well perceived than distractors at the location of the least salient distractor. This prediction stems from the fact that the amount of inhibition delivered to the location of a lower-priority stimulus in the AM is proportional to its priority. Testing this prediction would require embedding distractor-only trials within a larger set of trials that contain targets as well. Some of these trials would contain probe letters as in Gaspelin et al.

Note that that there is conflicting evidence about the ability of distractors to elicit an N2pc. Distractors that are highly salient on a different dimension than the target (e.g. color singleton distractors with shape singleton targets elicit an N2pc, while difference in salience along the same dimension (color) does not; Gaspar & McDonald 2014).

Prediction 3. Salient Distractors can sustain an existing lock on state

One of the most counterintuitive predictions of the RAGNAROC model is that once an attentional lock-on state has been established by a target, it can be sustained by a distractor because the attention map is agnostic about target/distractor categories. Attentional control settings bias attention towards the target, but distractors also have excitatory connections to the AM; they just have reduced priority. Thus in a similar manner as two sequential targets can maintain a lock-on state (Tan & Wyble 2015), a target followed by a distractor should also maintain the lock-on state. The prediction can be tested by presenting either three targets in sequence (i.e. letters among digits), at an SOA of about 120ms, or two targets separated by a single distractor that is similar to other distractors (i.e. a black digit), or two targets separated by a highly salient distractor (i.e. a red digit). It should be observed that for three targets in a row, the second and third targets elicit small-amplitude N2pcs that peak early (roughly 30ms earlier than the relative latency of the T1's N2pc). If the middle of the three targets is replaced by a highly salient distractor, the last target's N2pc should still be early and small in amplitude. However in the case of two targets separated by a non salient distractor, that last target should evoke an N2pc of normal amplitude and latency since the lock-on state will have partially dissipated during the 300ms lag between the onset of the two targets. In terms of behavior, the salient intervening distractor should result in more accurate report of the following target relative to the non-salient distractor condition, since the highly salient distractor sustains the lock-on state across the temporal gap between the targets.

7.2 Unified Attentional Map:

A central theme of the RAGNAROC architecture is that the competition for reflexive attention should be confined to one cortical map. This allows the entirety of the visual system to participate in scene analysis, and project the results of these computations to a single comparator map. This map allows the priority signals generated by different stimuli to compete, taking into account their salience, task relevance, emotional/reward history, or any other potential factor that influences how a given stimulus should be prioritized. The advantage of this centralized approach is to provide a cohesive decision-making process that computes a single spatial attention profile. This idea of a single, superordinate attention map is also shared by many some models of visual attention (Itti Koch et al 1998; .. Zelinsky2008) but not others (Tsotsos 2011).

Prediction 4: EEG correlates of lock-on occur regardless of stimulus type

A core finding of the lock-on state presents a straightforward means to test this architectural prediction. In Tan & Wyble, it was found that two targets in the same location produced only an N2pc to the first target, which the RAGNAROC explains as a carryover of the attentional lock-on state from one target to the next in the attention map.

However, in that study, both targets were of the same kind (letters among digit distractors). If there is a single architectural map, the same effect should occur even when T1 and T2 are of different types. Callahan-Flintoft & Wyble (In Press) have provided support for this prediction already, with a study in which targets could be defined by either shape or color, and the results were that there was only a single N2pc to the T1 regardless of whether the two targets were of the same or different type. It is nevertheless possible that the prediction may be falsified if the two targets even more distinct. For example RAGNAROC predicts that even if subjects are simultaneously looking for letters and faces of a particular gender, then two sequential targets (either letter-face or face-letter) should produce a clear N2pc only for the first of the two targets, regardless of order. Digits and faces should provide a strong test for the hypothesis since previous work has suggested that they are processed through sufficiently distinct channels in the visual system that the attentional blink evoked by a digit T1 has little effect on a face T2 (Awh et al. 2004).

7.3 Lock-on states in visual cueing:

The RAGNAROC model implements a reflexive form of attention that should be common across many visual attention paradigms, including visual cueing. Thus, we should be able to

predict behavioral and ERP effects for cueing experiments as well. Ansorge, Kiss, Worschech & Eimer (2011) have demonstrated that cues evoke clear N2pcs at moderate cue-target SOAs (200ms), as we would expect.

Prediction 5. Lock-on states in visual cueing, valid trials

The RAGNAROC model predicts that a lock-on state is sustained from one stimulus to the next. Thus, from a behavioral perspective, RAGNAROC provides a natural explanation for cueing benefits at short SOAs if one assumes that the cue initiates a lock-on state that carries forward in time to enhance the target, but only on valid trials. The model also generates EEG predictions for cueing experiments. Since the N2pc is caused by the formation of a new lock on state, then a validly cued target should result in an N2pc only for the cue, and not the subsequent target. This would only be true for cases in which the cue-target SOA is short enough that the lock-on state can be maintained, which is on the order of 100ms.

At longer SOAs between the cue and target, the lock-on state elicited by the cue would have disintegrated before the target appeared, with the result being that both the cue and the target would produce a complete N2pc. Note that expectation of the target occurring at the cued location cannot prevent the lock-on state from disintegrating in the RAGNAROC as suggested by a finding that two targets 600ms apart produce two N2pcs (Callahan-Flintoft & Wyble In Press). Thus, even in cases where the cue is 100% valid, a second N2pc would be produced at longer SOAs.

Prediction 6. Lock-on states in visual cueing, invalid trials

When a cue and target are not in the same location, then the cue and target will each evoke an N2pc at all SOAs, since the lock-on state elicited by a cue is spatially specific. Thus a target in an invalid trial needs to build a new lock on state which elicits a new N2pc.

Conclusions and Future Directions

Reflexive visual attention is a cornerstone of our visual system's ability to meet the challenge of rapidly choosing which information to selectively process, which pits stimuli that are inherently salient against those that are relevant for an ongoing task. A variety of experimental paradigms have provided a wealth of data that we have distilled into a common architecture for controlling the selection and suppression of information. The goal of the RAGNAROC model is to build a theoretical bridge between different paradigms (e.g. visual cueing and capture), and also

between different kinds of data (e.g. behavior and EEG). While designing the model to hit its empirical benchmarks, we have developed circuits that implement competing attractor states to stabilize the deployment of attention in both time and space.

Moving forward, the model's predictions are intended as a roadmap for further empirical investigation of reflexive attention. Testing these predictions, either within our lab, or any other, will provide diagnostic data regarding the model's validity, but more importantly, will help to inform us of the model's inaccuracies.

Acknowledgements:
This work was performed with the support of NSF grant 1734220
to B. W.

References

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences*, *108*(25), 10367-10371.

Ansorge, U., Kiss, M., Worschech, F., & Eimer, M. (2011). The initial stage of visual selection is controlled by top-down task set: new ERP evidence. *Attention, Perception, & Psychophysics*, *73*(1), 113-122.

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in cognitive sciences*, *16*(8), 437-443.

Awh, E., Serences, J., Laurey, P., Dhaliwal, H., van der Jagt, T., & Dassonville, P. (2004). Evidence against a central bottleneck during the attentional blink: Multiple channels for configural and featural processing. *Cognitive psychology*, *48*(1), 95-126.

Bahcall, D. O., & Kowler, E. (1999). Attentional interference at small spatial separations. *Vision research*, *39*(1), 71-86.

Bay, M., & Wyble, B. (2014). The benefit of attention is not diminished when distributed over two simultaneous cues. *Attention, Perception, & Psychophysics*, *76*(5), 1287-1297.

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & psychophysics*, *55*(5), 485-496.

Bichot, N. R., Cave, K. R., & Pashler, H. (1999). Visual selection mediated by location: Feature-based selection of noncontiguous locations. *Perception & Psychophysics*, *61*(3), 403-423.

Becker, S. I., Folk, C. L., & Remington, R. W. (2013). Attentional capture does not depend on feature similarity, but on target-nontarget relations. *Psychological Science*, *24*(5), 634-647.

Beuth, F., & Hamker, F. H. (2015). A mechanistic cortical microcircuit of attention for amplification, normalization and suppression. *Vision research*, *116*, 241-257.

Bowman, H., & Wyble, B. (2007). The simultaneous type, serial token model of temporal attention and working memory. *Psychological review*, *114*(1), 38.

Botella, J., Barriopedro, M., & Suero, M. (2001). A model of the formation of illusory conjunctions in the time domain. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(6), 1452.

Brisson, B., & Jolicoeur, P. (2007). The N2pc component and stimulus duration. *Neuroreport*, *18*(11), 1163-1166.

Bruce, N., & Tsotsos, J. (2006). Saliency based on information maximization. In *Advances in neural information processing systems* (pp. 155-162).

Bundesen, C. (1990). A theory of visual attention. *Psychological review*, *97*(4), 523.

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2011). A neural theory of visual attention and short-term memory (NTVA). *Neuropsychologia*, *49*(6), 1446-1457.

Burra, N., & Kerzel, D. (2013). Attentional capture during visual search is attenuated by target predictability: evidence from the N2pc, Pd, and topographic segmentation. *Psychophysiology*, *50*(5), 422-430.

Callahan-Flintoft & Wyble (In press) A hierarchical model of visual processing simulates neural mechanisms underlying reflexive attention, *Journal of Experimental Psychology: General*.

Cepeda, N. J., Cave, K. R., Bichot, N. P., & Kim, M. S. (1998). Spatial selection via feature-driven inhibition of distractor locations. *Perception & Psychophysics*, *60*(5), 727-746.

Chen & Wyble (In Press) The Neglected Contribution of Memory Encoding in Spatial Cueing: A New Theory of Costs and Benefits. *Psychological Review*

Cheal, M., & Lyon, D. R. (1991). Central and peripheral precuing of forced-choice discrimination. *The Quarterly Journal of Experimental Psychology*, *43*(4), 859-880.

Cheal, M. L., Lyon, D. R., & Gottlob, L. R. (1994). A framework for understanding the allocation of attention in location-precued discrimination. *The Quarterly Journal of Experimental Psychology*, *47*(3), 699-739.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*(6427), 345.

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of neurophysiology*, *80*(6), 2918-2940.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental psychology: Human perception and performance*, *21*(1), 109.

Cutzu, F., & Tsotsos, J. K. (2003). The selective tuning model of attention: psychophysical evidence for a suppressive annulus around an attended item. *Vision research*, *43*(2), 205-219.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, *18*(1), 193-222.

DiCarlo, J. J., & Maunsell, J. H. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *Journal of neurophysiology*, *89*(6), 3264-3278.

Dubois, J., Hamker, F. H., & VanRullen, R. (2009). Attentional selection of noncontiguous locations: The spotlight is only transiently "split". *Journal of Vision*, *9*(5), 3-3.

Egeth, H. E., Leonard, C. J., & Leber, A. B. (2010). Why salience is not enough: Reflections on top-down selection in vision. *Acta psychologica*, *135*(2), 130.

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and clinical neurophysiology*, *99*(3), 225-234.

Eriksen, C. W., & Yeh, Y. Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(5), 583.

Fazl, A., Grossberg, S., & Mingolla, E. (2009). View-invariant object category learning, recognition, and search: how spatial and object attention are coordinated using surface-based attentional shrouds. *Cognitive psychology*, *58*(1), 1-48.

Felleman, D. J., & Van, D. E. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, *1*(1), 1-47.

Folk, C. L., Remington, R. W., & Wright, J. H. (1994). The structure of attentional control: contingent attentional capture by apparent motion, abrupt onset, and color. *Journal of Experimental Psychology: Human perception and performance*, *20*(2), 317.

Folk CL, Leber AB, Egeth HE. Made you blink! Contingent attentional capture produces a spatial blink. Perception & Psychophysics. 2002;64:741–753

Fragopanagos, N., Kockelkoren, S., & Taylor, J. G. (2005). A neurodynamic model of the attentional blink. *Cognitive Brain Research*, *24*(3), 568-586.

Gaspar, J. M., & McDonald, J. J. (2014). Suppression of salient objects prevents distraction in visual search. *Journal of neuroscience*, *34*(16), 5658-5666.

Gaspelin, N., Leonard, C. J., & Luck, S. J. (2015). Direct evidence for active suppression of salient-but-irrelevant sensory inputs. *Psychological Science*, *26*(11), 1740-1750.

Heinke, D., & Humphreys, G. W. (2003). Attention, spatial representation, and visual neglect: simulating emergent attention and spatial memory in the selective attention for identification model (SAIM). *Psychological review*, *110*(1), 29.

Hickey, C., McDonald, J. J., & Theeuwes, J. (2006). Electrophysiological evidence of the capture of visual attention. *Journal of cognitive neuroscience*, *18*(4), 604-613.

Hickey, C., Di Lollo, V., & McDonald, J. J. (2009). Electrophysiological indices of target and distractor processing in visual search. *Journal of cognitive neuroscience*, *21*(4), 760-775.

Hickey, C., & Theeuwes, J. (2011). Context and competition in the capture of visual attention. *Attention, Perception, & Psychophysics*, *73*(7), 2053.

Hilimire, M. R., Mounts, J. R., Parks, N. A., & Corballis, P. M. (2010). Event-related potentials dissociate effects of salience and space in biased competition for visual representation. *PloS one*, *5*(9), e12677.

Hilimire, M. R., Hickey, C., & Corballis, P. M. (2012). Target resolution in visual search involves the direct suppression of distractors: evidence from electrophysiology. *Psychophysiology*, *49*(4), 504-509.

Hilimire, M. R., & Corballis, P. M. (2014). Event-related potentials reveal the effect of prior knowledge on competition for representation and attentional capture. *Psychophysiology*, *51*(1), 22-35.

Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences*, *95*(3), 781-787.

Hopf, J. M., Boehler, C. N., Luck, S. J., Tsotsos, J. K., Heinze, H. J., & Schoenfeld, M. A. (2006). Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. *Proceedings of the National Academy of Sciences*, *103*(4), 1053-1058.

Hopfinger, J. B., & Mangun, G. R. (1998). Reflexive attention modulates processing of visual stimuli in human extrastriate cortex. *Psychological science*, *9*(6), 441-447.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, *20*(11), 1254-1259.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. *Attention and performance*, 187-203.

Jovanic, T., Schneider-Mizell, C. M., Shao, M., Masson, J. B., Denisov, G., Fetter, R. D., ... & Zlatic, M. (2016). Competitive disinhibition mediates behavioral choice and sequences in Drosophila. Cell, 167(3), 858-870.

Kadel, H., Feldmann-Wüstefeld, T., & Schubö, A. (2017). Selection history alters attentional filter settings persistently and beyond top-down control. *Psychophysiology*, *54*(5), 736-754.

Kawahara, J., & Yamada, Y. (2006). Two noncontiguous locations can be attended concurrently: Evidence from attentional blink. Psychonomic Bulletin & Review, 13, 594–599

Kiss, M., Grubert, A., Petersen, A., & Eimer, M. (2012). Attentional capture by salient distractors during visual search is determined by temporal task demands. *Journal of cognitive neuroscience*, *24*(3), 749-759.

Koch, C, & Ullman, S. 1985. Shifts in selective visual attention: towards the underlying neural circuitry. Hum Neurobiol, 4(4), 219–27

Kristjánsson, Á., & Nakayama, K. (2003). A primitive memory system for the deployment of transient attention. *Perception & Psychophysics*, *65*(5), 711-724.

Kristjánsson, Á., Mackeben, M., & Nakayama, K. (2001). Rapid, object-based learning in the deployment of transient attention. *Perception*, *30*(11), 1375-1387.

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in cognitive sciences*, *17*(1), 26-49.

Kröse, B. J., & Julesz, B. (1989). The control and speed of shifts of attention. *Vision research*, *29*(11), 1607-1619.

Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and object size. *Journal of Neuroscience*, *33*(25), 10235-10242.

Kyllingsbæk, S., & Bundesen, C. (2007). Parallel processing in a multifeature whole-report paradigm. Journal of Experimental Psychology: Human Perception and Performance, 33, 64–82

Lanyon, L. J., & Denham, S. L. (2004). A model of active visual search with object-based attention guiding scan paths. *Neural Networks*, *17*(5-6), 873-897.

Lamy, D., & Egeth, H. E. (2003). Attentional capture in singleton-detection and feature-search modes. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(5), 1003.

Lamy, D., Leber, A., & Egeth, H. E. (2004). Effects of task relevance and stimulus-driven salience in feature-search mode. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(6), 1019.

LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological review*, *96*(1), 101.

Liesefeld, H. R., Liesefeld, A. M., Töllner, T., & Müller, H. J. (2017). Attentional capture in visual search: Capture and post-capture dynamics revealed by EEG. *NeuroImage*, *156*, 166-173.

Love, B. C. (2015). The algorithmic level is the bridge between computation and brain. *Topics in cognitive science*, *7*(2), 230-242.

Luck, S. J., & Hillyard, S. A. (1994). Spatial filtering during visual search: evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(5), 1000.

Marr, D. (1982). Vision. San Francisco: W. H. Freeman

Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology*, *32*(1), 4-18.

McDonald, J. J., Green, J. J., Jannati, A., & Di Lollo, V. (2012). On the electrophysiological evidence for the capture of visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 849.

Mounts, J. R. (2000). Attentional capture by abrupt onsets and feature singletons produces inhibitory surrounds. *Perception & Psychophysics*, *62*(7), 1485-1493.

Müller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *Journal of Experimental psychology: Human perception and performance*, *15*(2), 315.

Müller, M. M., & Hillyard, S. (2000). Concurrent recording of steady-state and transient event-related potentials as indices of visual-spatial selective attention. *Clinical Neurophysiology*, *111*(9), 1544-1552.

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision research*, *29*(11), 1631-1647.

Nako, R., Wu, R., & Eimer, M. (2014). Rapid guidance of visual search by object categories. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 50.

Nako, R., Wu, R., Smith, T. J., & Eimer, M. (2014). Item and category-based attentional control during search for real-world objects: Can you find the pants among the pans?. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1283.

Neisser U, Becklen R, 1975 ``Selective looking: Attending to visually specified events'' Cognitive Psychology 7 480 ^ 494

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium.

Nunez, P. L., & Cutillo, B. A. (Eds.). (1995). *Neocortical dynamics and human EEG rhythms*. Oxford University Press, USA.

O'Reilly, R. C., & Munakata, Y. (2000). Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain. MIT press.

Pagano, S., & Mazza, V. (2012). Individuation of multiple targets during visual enumeration: New insights from electrophysiology. *Neuropsychologia*, *50*(5), 754-761.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of experimental psychology: human learning and memory*, *2*(5), 509.

Rao, R. P., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision research*, *42*(11), 1447-1463.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink?. *Journal of experimental psychology: Human perception and performance*, *18*(3), 849.

Remington, R. W., Johnston, J. C., & Yantis, S. (1992). Involuntary attentional capture by abrupt onsets. *Perception & Psychophysics*, *51*(3), 279-290.

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron*, *61*(2), 168-185.

Rock, I., Linnett, C. M., Grant, P., & Mack, A. (1992). Perception without attention: Results of a new method. *Cognitive psychology*, *24*(4), 502-534.

Schubö, A. (2009). Salience detection and attentional capture. *Psychological Research*, *73*(2), 233.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological review*, *84*(2), 127.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature neuroscience*, *5*(7), 631.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *perception*, *28*(9), 1059-1074.

Sawaki, R., & Luck, S. J. (2010). Capture versus suppression of attention by salient singletons: Electrophysiological evidence for an automatic attend-to-me signal. *Attention, Perception, & Psychophysics*, *72*(6), 1455-1470.

Sawaki, R., Geng, J. J., & Luck, S. J. (2012). A common neural mechanism for preventing and terminating the allocation of attention. *Journal of Neuroscience*, *32*(31), 10725-10736.

Silver, M. A., Ress, D., & Heeger, D. J. (2005). Topographic maps of visual spatial attention in human parietal cortex. *Journal of neurophysiology*, *94*(2), 1358-1371.

Shipp, S. (2004). The brain circuitry of attention. *Trends in cognitive sciences*, *8*(5), 223-230.

Swan, G., & Wyble, B. (2014). The binding pool: A model of shared neural resources for distinct items in visual working memory. *Attention, Perception, & Psychophysics*, *76*(7), 2136-2157.

Tan, M., & Wyble, B. (2015). Understanding how visual attention locks on to a location: Toward a computational model of the N 2pc component. *Psychophysiology*, *52*(2), 199-213.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & psychophysics*, *51*(6), 599-606.

Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. *Perception & psychophysics*, *50*(2), 184-193.

Theeuwes, J., Atchley, P., & Kramer, A. F. (2000). On the time course of top-down and bottom-up control of visual attention. *Control of cognitive processes: Attention and performance XVIII*, 105-124.

Töllner, T., Zehetleitner, M., Gramann, K., & Müller, H. J. (2011). Stimulus saliency modulates pre-attentive processing speed in human visual cortex. *PLoS One*, *6*(1), e16276.

Töllner, T., Rangelov, D., & Müller, H. J. (2012). How the speed of motor-response decisions, but not focal-attentional selection, differs as a function of task set and target prevalence. *Proceedings of the National Academy of Sciences*, *109*(28), E1990-E1999.

Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial intelligence*, *78*(1-2), 507-545.

Tsotsos, J. K. (2011). *A computational perspective on visual attention*. MIT Press.

Van Essen, D. C., & Maunsell, J. H. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in neurosciences*, *6*, 370-375.

Visser, T. A., Bischof, W. F., & Di Lollo, V. (1999). Attentional switching in spatial and nonspatial domains: Evidence from the attentional blink. *Psychological Bulletin*, *125*(4), 458.

White, A. L., Lunau, R., & Carrasco, M. (2014). The attentional effects of single cues and color singletons on visual sensitivity. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 639.

Wilschut, A., Theeuwes, J., & Olivers, C. N. (2011). The time course of attention: selection is transient. *PLoS one*, *6*(11), e27661.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review*, *1*(2), 202-238.

Woodman, G. F., & Luck, S. J. (1999). Electrophysiological measurement of rapid shifts of attention during visual search. *Nature*, *400*(6747), 867.

Wyble, B., Bowman, H., & Potter, M. C. (2009). Categorically defined targets trigger spatiotemporal visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 324.

Wyble, B., Folk, C., & Potter, M. C. (2013). Contingent attentional capture by conceptually relevant images. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 861.

Wyble, B., & Swan, G. (2015). Mapping the spatiotemporal dynamics of interference between two visual targets. *Attention, Perception, & Psychophysics*, *77*(7), 2331-2343.

Yantis, S. (1996). Attentional capture in vision. In A. F. Kramer, M. G. H. Coles, & G. D. Logan (Eds.), *Converging operations in the study of visual selective attention* (pp. 45-76). Washington, DC, US: American Psychological Association. http://dx.doi.org/10.1037/10187-002

Yeshurun, Y., & Carrasco, M. (1999). Spatial attention improves performance in spatial resolution tasks1. *Vision research*, *39*(2), 293-306.

Zehetleitner, M., Koch, A. I., Goschy, H., & Müller, H. J. (2013). Salience-based selection: Attentional capture by distractors less salient than the target. *PLoS One*, *8*(1), e52595.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological review*, *115*(4), 787.

Zirnsak, M., Beuth, F., & Hamker, F. H. (2011). Split of spatial attention as predicted by a systems-level model of visual attention. *European Journal of Neuroscience*, *33*(11), 2035-2045.

Zhang, W., & Luck, S. J. (2009). Feature-based attention modulates feedforward visual processing. *Nature neuroscience*,

Li, Z. (2002). A saliency map in primary visual cortex. *Trends in cognitive sciences*, *6*(1), 9-16.

Appendix

MATLAB Code for running the simulations is available on the OSF at https://osf.io/rwynp/

Appendix 1, Equations

Early Visual layer:

These are the activation equations for each neuron in the EV, and note we represent each possible stimulus (T1 and T2) as having distinct EV neurons to reflect the fact that two distinct stimuli will activate distinct groups of neurons in V1 even if presented at the same location. These equations match those specified by O'Reilly & Munakata (2001) where *RInput*, represents the presence of an input stimulus at a given location and time point (either 1 or 0) and EV represents the activation level of that neuron. dtVM is a time constant that dictates the rate of change of a neuron by scaling the excitatory, and leak currents. EE and EL are the reversal potentials for excitatory and leak currents.

$$EV_{excite}(x,y,t) = dt_{VM} * \left(EE - EV(x,y,t)\right) * Input(x,y,t) \tag{1.1}$$

$$EV_{leak}(x,y,t) = dt_{VM} * \left(EL - EV(x,y,t)\right) \tag{1.2}$$

$$EV(x,y,t+1) = EV(x,y,t) + EV_{excite}(x,y,t) + EV_{leak}(x,y,t) \tag{1.3}$$

Late Visual layer:

The LV neurons have essentially the same dynamics except that they receive input from a region of of EV neurons and the value of the input is scaled by a square-masked Gaussian distribution (GRF), for computational efficiency.

The parameter Attn is the value of attention as determined by activity at the corresponding location in the AM. EI is the reversal potential of the inhibitory current. IItoIT is a parameter that determines the strength of the feedback inhibition interneurons for each neuron.

The max function is used to clip the value of certain parameters.

$$LV_{excite}(x,y,t) = dt_{VM} * (EE - LV(x,y,t)) * BU_{type} * \sum_{x'=-mask}^{mask} \sum_{y'=-mask}^{mask} GRF(x',y') * Attn(x+x', y+y', t) * (EV(x+x', y+y', t) - Thresh_{EV})_0$$

(1.4)

$$LV_{inhib}(x,y,t) = dt_{VM} * \left(EI - LV(x,y,t)\right) * IItoIT * \left(II(x,y,t) - II_{Thresh}\right)_0 \quad (1.5)$$

$$LV_{leak}(x,y,t) = dt_{VM} * \left(EL - LV(x,y,t)\right) \quad (1.6)$$

$$LV(x,y,t+1) = max\left\{EI, LV(x,y,t) + LV_{excite}(x,y,t) + LV_{inhib}(x,y,t) + LV_{leak}(x,y,t)\right\}$$

(1.7)

The II neurons govern the feedback inhibition of the LV neurons following a similar dynamic as the EV.

$$II_{excite}(x,y,t) = dt_{VM_II} * \left(LV(x,y,t) - LV_{thresh}\right)_0 * ITtoII \quad (1.8)$$

$$II_{leak}(x,y,t) = dt_{VM_II} * \left(EL - II(x,y,t)\right) \quad (1.9)$$

$$II(x,y,t+1) = II(x,y,t) + II_{Excite}(x,y,t) + II_{Leak}(x,y,t) \quad (1.10)$$

Attention Map:

The AM neurons receive input from all LV maps (1-*n*) scaled by the same masked Gaussian function. LAI is a parameter that controls the magnitude of inhibitory suppression from the IG to the AM neurons.

$$AM_{excite}(x,y,t) = dt_{VM} * (EE - AM(x,y,t)) * \sum_{type=1}^{n} \sum_{x'=-mask}^{mask} \sum_{y'=-mask}^{mask} GRF(x',y') * (LV_{type}(x+x', y+y', t) - Thresh_{LV})_0 * TD_{type}$$

(1.11)

$$AM_{inhib}(x,y,t) = dt_{VM} * \left(EI - AM(x,y,t)\right) * \left(IG(x,y,t) - IG_{thresh}\right)_0 * LAI$$

(1.12)

$$AM_{leak}(x,y,t) = dt_{VM} * \left(EL - AM(x,y,t)\right) \quad (1.13)$$

$$AM(x,y,t+1) = max\left\{EI, AM(x,y,t) + AM_{excite}(x,y,t) + AM_{inhib}(x,y,t) + AM_{leak}(x,y,t)\right\}$$

(1.14)

$$Attn(x,y,t) = max\left(1, log\left(AM(x,y,t) - AM_{threshLow} + 1\right)_0 * Attnweight\right) \quad (1.15)$$

The IG neurons within the Attention Map receive joint input from the LV and the AM. For each IG neuron, the input from each of those two sources has a ceiling value (*MaxInputtoIG*). Thus, the input from the LV and the AM to each IG neuron is computed separately.

$$IG_{exciteLV}(x,y,t) = max(MaxInputtoIG, \sum_{type=1}^{n} \sum_{x'=-mask}^{mask} \sum_{y'=-mask}^{mask} GRF(x',y') * (LV_{type}(x+x', y+y', t) - Thresh_{LV})_0 * TD_{type}$$

(1.16)

$$IG_{exciteAM}(x,y,t) = max(MaxInputtoIG * \sum_{x'=1}^{\dim x} \sum_{y'=1}^{\dim y} (AM(x',y',t) - Thresh_{AMLow})_0 * DoG(x',y') * AMtoIG$$

(1.17)

DoG is the difference of two Gaussians as specified below.

$$IG_{excite}(x,y,t) = (IG_{ExciteLV}(x,y,t) + IG_{ExciteAM}(x,y,t)) * \ dtVM_{IG} * \left(EE - IG(x,y,t)\right)_0$$

(1.18)

$$IG_{inhib}(x,y,t) = \ dt_{VM_IG} * \left(AM(x,y,t) - Thresh_{AMHigh}\right)_0 * AMtoIGinhib$$

(1.19)

$$IG_{leak}(x,y,t) = \ dt_{VM_IG} * \left(EL - IG(x,y,t)\right) \qquad (1.20)$$

$$IG(x,y,t+1) = \ max\left\{EI, IG(x,y,t) + IG_{excite}(x,y,t) + IG_{inhib}(x,y,t) + IG_{leak}(x,y,t)\right\}$$

(1.21)

Gaussian Functions

$$GRF(x,y) \ = e^{-0.5\left(x^2+y^2\right)} \qquad (1.22)$$

$$DoG(x,y) \ = -1 * \left[2e^{-0.5\left(x^2+y^2\right)outerGaussian} - e^{-0.5\left(x^2+y^2\right)innerGaussian}\right]_0 \qquad (1.23)$$

EEG Scalp Voltage

$$EEGVoltage(t) = \left[\sum_{x=1}^{x\dim} \sum_{y=1}^{y\dim} \left(AM_{excite_EEG}(x,y,t) + AM_{inhib}(x,y,t) \right)_0 \right] \tag{1.24}$$

$$AM_{excite_EEG}(x,y,t) = dt_{VM} * (EE_{EEG} - AM(x,y,t)) * \sum_{type=1}^{n} \sum_{x'=-mask}^{mask} \sum_{y'=-mask}^{mask} GRF(x',y') * (LV_{type}(x+x', y+y', t) - Thresh_{LV})_0 * TD_{type} \tag{1.25}$$

Fixed Parameters:
dt_vm = .015
dt_vm_II = .0025
dt_vm_IG = .04
EE = 30
EL = 0;
EI = -10;
EE_{EEG} *= 65;*

Weights
TD_{type} *= Free Parameter*
BU_{type} *= Free Parameter*

ITtoII = .02
IItoIT = 6.5
AMtoIG = .4
AMtoIGinhib = .25
LAI = .45
Attnweight = 2
MaxInputtoIG = .35

Thresholds
$Thresh_{EV} = 7$
$Thresh_{IT} = 5$
$Thresh_{II} = 0$
$Thresh_{IG} = 8$
$Thresh_{AMLow} = 14$
$Thresh_{AMHight} = 22$

Misc
outerGaussian = .07
innerGaussian = .2

Appendix 2, Fitted Parameters

Sim Name	Type1BU	Type2BU	Type3BU	Type1TD	Type2TD	Type3TD
Bay & Wyble	0.3	0.3	N/A	0.17	0.12	N/A
Nakayama	0.15	0.15	N/A	0.18	0.18	N/A
Theeuwes (with salient D)	0.15	0.3	N/A	0.27	0.17	N/A
Theeuwes (without salient D)	0.15	0.05	N/A	0.27	0.17	N/A
Mounts	0.3	0.3	N/A	0.24	0.24	N/A
Gaspelin exp 1 (singleton search)	0.15	0.19	0.15	0.2	0.15	0.2
Gaspelin exp 2 (feature search)	0.15	0.19	0.15	0.4	0.15	0.2
LatDMidlineT	0.15	0.17	N/A	0.5	0.2	N/A
Tan & Wyble	0.15	0.15	N/A	0.2	0.2	N/A
Tollner (low)	0.17	0	N/A	0.15	0	N/A
Tollner (med)	0.2	0	N/A	0.15	0	N/A
Tollner (high)	0.23	0	N/A	0.15	0	N/A
Hillimire (unpredictable)	0.15	0.15	N/A	0.22	0.22	N/A
Hillimire (predictable)	0.15	0.15	N/A	0.4	0.25	N/A
Eimer Grubert	0.6	0.6	N/A	0.7	0.7	N/A
High TD	0.15	0	N/A	0.2	0	N/A
High BU	0.2	0	N/A	0.15	0	N/A

Sim Name	Mean	SD	Threshold
Bay & Wyble	3	0.75	10,891
Nakayama	3	0.75	6,997
Theeuwes (with salient D)	3	0.75	19,522
Theeuwes (without salient D)	3	0.75	19,522
Mounts	3	0.75	11,493
Gaspelin exp 1 (singleton search)	3	0.75	8,667
Gaspelin exp 2 (feature search)	3	0.75	8,667
LatDMidlineT	3	0.75	N/A
Tan & Wyble	3	0.75	N/A
Tollner (low)	3	0.75	N/A
Tollner (med)	3	0.75	N/A
Tollner (high)	3	0.75	N/A
Hillimire (unpredictable)	3	0.75	N/A
Hillimire (predictable)	3	0.75	N/A
Eimer Grubert	3	0.75	N/A
High TD	N/A	N/A	N/A
High BU	N/A	N/A	N/A