Selective inhibition of distracting input

MaryAnn Noonan\textsuperscript{1,}, Ben Crittenden\textsuperscript{2}, Ole Jensen\textsuperscript{3} and Mark Stokes\textsuperscript{1}

\textsuperscript{1}Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, OX1 3UD

\textsuperscript{2}The Oxford Centre for Functional MRI of the Brain, Nuffield Department of Clinical Neurosciences, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DU.

\textsuperscript{3}Centre for Human Brain Health, School of Psychology, University of Birmingham, B15 2TT

\textbf{Address for correspondence:}

Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, OX1 3UD

Tel: +44 (0) 1865 277490

Fax: +44 (0) 1865 310447

E-mail: maryann.noonan@psy.ox.ac.uk

\textbf{Abstract}

We review a series of studies exploring distractor suppression. It is often assumed that preparatory distractor suppression is controlled via top-down mechanisms of attention akin to those that prepare brain areas for target enhancement. Here, we consider two alternative mechanisms: secondary inhibition and expectation suppression within a predictive coding framework. We draw on behavioural studies, evidence from neuroimaging and some animal studies.

We conclude that there is very limited evidence for selective top-down control of preparatory inhibition. By contrast, we argue that distractor suppression often relies secondary inhibition of non-target items (relatively non-selective inhibition) and on statistical regularities of the environment, learned through direct experience.

Keywords: Distractor inhibition, predictive coding, alpha oscillations, attention, frontoparietal network
Introduction

Our everyday view of the world is necessarily biased: we focus our attention on information that is most relevant to our current goals, and ignore behaviourally irrelevant input. Without such bias, we would be lost in a world of information-overload, unable to accomplish even the simplest tasks.

A rich tradition of research in cognitive neuroscience explores the brain systems responsible for biasing perception, memory and action. Much of this research has primarily focused on the mechanisms that facilitate processing of the most relevant sensory input (i.e., target cueing). However, inhibition is an important counterpoint to facilitation: in complex environments it may be equally important to inhibit distracting input as it is to enhance processing of behaviourally relevant information. Moreover, inhibitory neural dynamics are ubiquitous throughout the central nervous system [1], and likely play a key role in shaping information processing [2].

In this review, we focus specifically on the role of preparatory neural states for suppressing task-irrelevant visual input, rather than the reactive inhibitory processes engaged during distractor processing. Where possible we distinguish preparatory control from active online suppression of distractor stimuli to answer the question: can attentional control mechanisms prepare to suppress task-irrelevant information before it is presented. We consider three different models of preparatory inhibition: direct inhibition, secondary inhibition and inhibition by expectation suppression. Direct inhibition is conceived as the selective inhibition of specific representations in lower sensory areas by distractor-specific representations in attentional control areas. We consider this as the literal flipside to target facilitation: top-down control can facilitate or inhibit depending on the valence (positive or negative) of task-relevance. Secondary inhibition, on the other hand, is expressed as a secondary consequence of top-down target facilitation via a complementary balance of excitation and suppression within the visual system. For example, top-down control mechanisms
guided by working memory representations of a target feature could up-regulate activity in target-related neural populations, triggering a complementary spread of inhibition to unattended features via local inhibitory circuits. This form of inhibition, therefore, should be as flexible as target facilitation, but perhaps relatively non-specific (i.e., suppress everything outside the focus of attention). Finally, expectation suppression could provide a more selective form of inhibition governed by principles of predictive coding. Predicted inputs can be selectively suppressed, as is observed in some contexts with repetition suppression. While a general suppression mechanism may inhibit all repeated and predictable inputs, top-down attention could act to release task-relevant stimuli from this inhibition [3], leaving the remaining inhibitory influence focused specifically on expected distractors. Below we review behavioural evidence for distractor inhibition within these three frameworks, and consider putative neural mechanisms for preparatory neural states that underpin distractor inhibition. While we mainly highlight the differences between the models we acknowledge that each might be specialised for different contexts, dimensions or stimuli and how they may interact and complement each other.

**Competition, top-down control and expectation suppression**

Perceptual processing is inherently competitive. For even a simple target detection task, the presence of concurrent task-irrelevant sensory input (i.e., distractors) can drastically impair behaviour [4, 5]. Indeed, perceptual competition is often conceptualised as a race between a number of candidate representations [6]. Stimulus-specific properties endow some representations with an inherent competitive advantage [7], but importantly top-down factors such as task relevance can also bias competition in favour of the most relevant input [6, 8, 9].

Extensive evidence demonstrates how prior knowledge of task-relevant perceptual features (such as the likely location of the target) can bias attention in favour of perceptual targets at the expense of competing distractors [10-12]. Influential models of attentional control, such as the Biased
Competition model, propose that an attentional template is maintained in working memory [e.g. 8], and up-regulation of target-related neurons in turn facilitates perceptual analysis of matching stimuli [13]. However, in many respects the functional consequence of target facilitation is best understood relative to non-target inhibition [14]. For example, the Feature Similarity Gain model argues that this gain of attentional modulation is a function of the similarity between attended feature and a cell’s preferred feature [15]. It is therefore critical to understand how the brain prepares to inhibit distracting input. In this review, we consider three potential modes of preparatory inhibition: direct, secondary inhibition and expectation suppression (see figure 1).

Figure 1

Direct inhibition

When a participant is made aware that a particular location (i.e. right of fixation in figure 1) or feature (i.e. blue items) is reliably associated with stimuli that are not relevant to behaviour, i.e. reliably distracting, it would be beneficial to use that knowledge to selectively inhibit that location (via suppression of neural activity of left visual cortex, figure 1A) or feature. Theories of direct distractor inhibition appear to closely resemble models of response inhibition [16, 17]. It has been argued that direct and flexible top-down inhibition is the complementary mechanism to top-down facilitation, both mechanisms mediated by higher-order brain regions such as the dorsolateral prefrontal cortex, inferior frontal gyrus and parietal cortex which may directly boost or suppress processing in the sensory region. Here we use the term direct, however we do not suggest that backward projections from FPC are necessarily mono-synaptic, rather that the regions are effectively inhibited from signals from the FPC without substantial transformation (in contrast to the indirect mechanism described below). While this could depend on the rare and mysterious long range inhibitory (e.g. GABAergic) projections [18], it would be perhaps more likely that long-range excitatory connections directly or indirectly drive the relevant down-stream interneurons to
suppress local activity and/or drive oscillations at the alpha frequency range to exert pulsed inhibition [2].

There is some empirical support for this idea, suggesting that suppression of an irrelevant stimulus [10], colour [19], spatial location [20, 21], or even a memory [22, 23] is achieved via distractor-specific top-down mechanisms. Furthermore, this direct and flexible inhibitory mechanism has been invoked to account for effects seen within a number of cognitive control tasks such as the Stroop, Flanker, task switching and negative priming [see 17 for comprehensive review]. However, closer inspection identifies a number of important conceptual differences between cancelling a pre-programmed behavioural response and preparing to inhibit an expected sensory input. Response inhibition is often conceptualised as a race between a stop signal and the already-initiated go signal [24], whereas preparatory inhibition of a distractor stimulus requires selective modulation of distractor related processing circuits prior to the onset of sensory stimulation. Moreover, evidence from studies on working memory and attention suggest that flexible top-down biasing could depend on WM representation (e.g., [8]), which are inherently facilitative [25-28] as a function of similarity with the working memory template [29], even if the representation is known to match a distractor item [27, 30]. This would imply that flexible top-down control is best suited for boosting target representations, not distractor inhibition. Indeed, if the stimulus template were maintained in working memory it would seem likely that directly cueing a distractor item would actually boost distractor processing, analogous to the paradoxical effects of thought suppression described by Wegner and colleagues [31].

Secondary inhibition

We consider secondary inhibition as an alternative mode of distractor inhibition [32]. It is defined as the secondary spread of neural suppression along inhibitory pathways as a consequence of target facilitation and could be as important functionally for biasing competition as the original
enhancement. Indeed, early studies argued that distracting input is suppressed during perceptual processing via secondary spreading of suppression with homogeneity among non-targets via perceptual grouping of distractors facilitating this spread of inhibition [29, 33-35]. Further, the spread of inhibition during distractor processing has been considered important for efficient processing of multiple distractors that share a similar features [29]. In our example case of preparatory secondary inhibition (illustrated in figure 1C), the ‘ignore right’ goal is converted into a to ‘attend left’ instruction which leads to facilitation of an attended location and triggers neural suppression at non-attended locations. Because secondary inhibition is directly controlled by target facilitation, it should be flexible and rapid, but not necessarily fine-tuned for specific types of distraction (i.e., suppress everything not in the focus of attention [36]). Such an alternative to direct inhibition might sit better with evidence from trial-wise target cueing studies that find flexible allocation of attention depends on working memory representations to bias processing in favour of task-relevant input [e.g. 37]. In such cases attention may be re-directed to the non-cued location(s), facilitating likely target position(s), before intrinsic competition eventually suppresses the originally cued distractor location. Such dynamics are assumed in most models of stimulus-driven competition in winner-take-all networks [7], however it is also possible that similar interactions could be triggered in the absence of a stimulus-drive to prepare for distractor inhibition [14]. This could help establish, and/or maintain a non-target inhibitory state. However, unlike direct suppression, secondary inhibition would be relatively non-specific. The control mechanism is inherently target-focused. Inhibition simply spreads to all competing non-target representation, rather than a specific forewarned distractor item.

*Expectation Suppression*

Finally, we consider a form of inhibition that might not depend on working memory. If task-relevant parameters are relatively stable over time (e.g., blocked over a number of experimental trials), attentional modulation may bypass working memory [38]. Predictive coding models suggest that
each stage of visual processing generates and updates hypotheses about incoming sensory input [3]. Accordingly, expected input, whether task relevant or irrelevant, is attenuated because it provides little new information [39]. Computationally, this provides an elegant strategy for reducing informational redundancy [40]. We suggest that predictive coding could contribute to distractor suppression because predictable input tends to be less behaviourally relevant than new, surprising, events. To consider a real-world example: in many visual scenes, potential distractors are spatially predictable and may repeat over time. Trees, lampposts and post-boxes are rarely our goals as we walk down the street, and so these predictable events can usefully be ignored. However, this is obviously not always the case. If expected stimuli are also task-relevant, in situations in which we wish to post a letter, top-down control systems operating through working memory should be able to break through default inhibition, to release goal relevant stimuli, i.e. the post-box, from expectation suppression [3]. We schematise this model in figure 1D with visual input locally suppressed if it matches expected content. Expected input is released from inhibition via top-down facilitation mechanisms if it is also task-relevant.

Expectation suppression could provide an important source of item-specific preparatory inhibition that does not depend on (presumably) facilitative representations in working memory. Instead, the statistical regularities of the environment, learned through direct experience [41], provide a rich source of predictive information to bias perception [42]. As such, expectation suppression could operate in parallel with the control mechanisms that govern target-related facilitation.

*Behavioural evidence for preparatory distractor inhibition*

Having outlined three general models for preparatory inhibition, we now review behavioural studies of preparatory suppression to consider how the extent evidence provides relative support for these defined models.
The contribution of proactive top-down inhibitory control has been examined across a broad range of paradigms and domains [10, 20-23]. One particularly controlled example that explicitly investigates preparatory control via symbolic cueing was provided by Ruff and Driver [43]. Here the authors examined whether foreknowledge of distractor likelihood can be used flexibly to modulate suppression on a trial-wise basis using a modified Posner spatial cuing paradigm (Figure 2A). As standard, a pre-cue directed covert spatial attention to one or other side of space; however, in the critical condition, the format of the cue also indicated whether a distractor in the opposite hemifield would be present or not. Response times (RTs) were faster in blocks where the cue predicted the presence/absence of the distractor compared to when the cue provided no such distractor-predictive information. The authors argued that trial-to-trial advance knowledge about the presence of a distractor in the target-opposite hemifield significantly reduced distractor competition. In principle the distractor cueing benefit could be mediated through a direct top-down suppression of perceptual representations of the expected distractor, analogous to facilitation associated with target cueing (Figure 1B). However, as discussed in more detail below, top-down working memory based control mechanisms may not be able to simply switch from facilitation to inhibition if a cue predicts distractors rather than target relevant information. Within the current framework, inhibition observed in Ruff and Driver [43] may be mediated via a secondary spread of inhibition following additional enhancement in attentional allocation towards the cued target. Although secondary inhibition might be relatively non-specific (i.e., inhibit all uncued locations), it is as flexible as goal directed target facilitation. A cue that predicts the target location as well as the likelihood of a competing distractor could engage a more efficient attentional focus at the expected target location, presumably with a concomitant increase in the spread of inhibition to the task-irrelevant perceptual areas. This would be particularly useful in a binary situation, where target and distractor stimuli can only appear in two mutually exclusive locations.
Many studies have emphasised that information in working memory appears to automatically draw attention to features that match the contents [28, 44] even when detrimental to the task. This would make distractor inhibition via working memory particularly challenging. However, more recent work has suggested the relationship between working memory and distractor processing may not be that simple, instead the degree of distractor interference on performance is dependent on working memory load, content, validity, perceptual difficulty and timing [25, 26, 45-49]. This could suggest that working memory based control mechanisms may be more flexible under some circumstances and perhaps contribute to directed inhibition. Indeed, some evidence seems to suggest that items in working memory can be used to selectively inhibit distracting input. For example, Kim and colleagues have shown that the content of working memory has differential effects on distractor interference. If working memory load contains to-be-target items, in a concurrent task, distractor interference increased. However, when working memory items shared processing with distractors, distractor interference decreased and target selection is facilitated [47, 48]. Cognitive control may therefore play an important mediating factor in these effects [49]. Further research is needed to chart the boundary conditions of these effects.

Analogous to Ruff & Driver’s distractor cueing task, but in the feature domain, Arita and colleagues cued subjects on each trial to the colour of a forthcoming distractor or target set within a visual search task and report reduced reaction times in both contexts relative to a neutral baseline [19]. While their first experiment only used two colours within the search array allowing subjects to potentially use secondary inhibition as we explored in Ruff & Driver’s work, the second experiment used seven colours in each array potentially suggesting a direct inhibition account. Nevertheless, this experiment also used an explicit colour patch to cue distractor inhibition. As discussed below, direct experience of the distractor feature could trigger expectation suppression, therefore it would be important consider whether more abstract cue-types could also be helpful in this particular context.
Other behavioural studies have demonstrated how symbolic cues and expectation interact to modulate distractor inhibition: For example, Awh et al. [21] presented participants with a 5x5 grid within which two target numbers appeared at cued locations (Figure 2B). On ‘clean trials’, only two number stimuli were present, however on ‘noise trials’ distractor letters appeared in the remaining locations. Distractor expectation was manipulated across two critical block types: 80% noise trials, 20% clean trials and; 20% noise trials, 80% clean trials. This allowed the experimenters to compare performance for expected versus unexpected noise displays relative to expected/unexpected clean displays. Unsurprisingly, accuracy was worse in noise trials (standard competition effect), however this distractor interference was reduced when distractors were more likely to occur. Put simply, distractors were less distracting if they were anticipated and could imply that people can exert a certain amount of top-down control to ameliorate the interference caused by expected distractor items. Furthermore, there was no behavioural advantage on distractor-absent trials. A selective improvement on distractor present trials provides important evidence that the benefit of distractor foreknowledge was due to selective distractor inhibition, rather than increased target facilitation or heightened vigilance on high competition trials. Mechanistically, these effects could be achieved by directly inhibiting the expected distractor locations, or by boosting the secondary spread of inhibition associated with facilitative target cueing. If we consider the latter scenario in more detail, the null effect on distractor absent trials further suggests that it is possible to selectively modulate the spread of inhibition triggered by target-relevant facilitation, without necessarily ramping up target facilitation. Finally, these attentional effects could also be understood in terms of expectation suppression as the distractor locations were predictable from trial to trial.

Finally, Watson and Humphreys [50] adapted a standard search task in which a target item was embedded in an array of distractors. Unsurprisingly, the presence of distractors impaired search performance; but critically, they also found that the distractor effect was dramatically reduced if participants were additionally presented with a preview screen containing a subset of distractors
(see Figure 2C). Again, this work could be interpreted as suggesting that participants were able to use the preview frame to prepare to suppress processing of the featured distractors. Indeed, when the number of preview items is below working memory capacity there appears to be little impact of the presence of a distractor [51]. However, according to the predictive coding framework outlined above, processing the distractor stimuli prior to the target presentation removes the need for further processing during the search task. In effect, the task-irrelevant items are suppressed by expectation. This interpretation is conceptually related to inter-trial priming [52] and contextual cueing effects seen in visual search tasks that occur when associations between spatial configurations of distractor contexts and target locations are learned incidentally and facilitate spatial attention towards task-relevant aspects of a scene [53].

Figure 2

Teasing apart the contribution of each putative mechanism to each task is challenging. We therefore recently pit direct inhibition against expectation suppression while minimising the influence of secondary suppression. Specifically, we asked whether top-down control mechanisms are able to flexibly inhibit specific distractor locations [54]. We, like Ruff and Driver, used an adapted Posner design but now targets and distractors could appear in one of four locations (Figure 3A). This non-binary design reduced the strategic advantage of inverting a distractor cue to a target cue and diluted the potential influence of secondary suppression (as in [19]). Specifically, just as foreknowledge of the location of task-relevant stimuli can be used to enhance target processing, we examined whether subjects can use foreknowledge of the location of a task-irrelevant stimulus to suppress distractor processing. Target stimuli were presented with and without a competing distractor, and participants were cued with 100% validity about the location of the target or distractor. RTs in these two conditions were compared to a neutral condition in which the cue was spatially non-informative. Cue condition varied in a block-wise manner while cued location either
varied on a trial-wise basis (Flexible Cueing task), or fixed over a block of trials (Blocked Cueing task). This critical aspect of the experiment tests the flexibility of selective inhibition. Performance on the target discrimination task suggested that subjects could not use prior knowledge about the distractor location in a flexible manner to reduce the behavioural cost associated with its presence. By contrast, when the distractor location was fixed and repeated to the same location over a block of trials subjects were able to suppress the RT cost of the distractor. Interestingly there is no correlation between subjects’ behavioural benefit from a target cue and the benefit of a predictable distractor.

Figure 3

In a closely related study, Leber and colleagues [55] have elegantly shown that cues to a forthcoming distractor can be used to flexibly suppress a distractor if they are tied implicitly to the location of the target. In their study, explicit trial-wise cues always validly predicted target locations, but also implicitly cued distractor location (Figure 3B). Subjects RTs to target identification were faster when distractors were presented at their implicitly cued location. Furthermore if subjects then had to subsequently engage with the implicitly cued distractor location they were slowed relative to non-distractor locations. Further work has even suggested that under some circumstances distractor repetition suppression effects can transfer to other spatial locations [56]. Collectively, these results favour experience-based distractor suppression.

Similar mechanisms have been shown to apply to distractor suppression in feature space. Cunningham & Egeth [57] report that if the cued, to-be-ignored colour varied from trial-to-trial there is no improvement in performance. However, if the distractor colour is same for duration of experiment then initial cost in RT reduces and, with practice, becomes faster compared to trials in which the subject was not forewarned the colour of the subsequent distracting stimuli (neutral). The
build up of a distractor expectation is also evident in neutral trials when targets happen to be presented in the distractor colour. Initially subjects’ RTs are facilitated but over the session RTs slow relative to distractor cued trials or trials in which the target shown in a non-distractor colour. Again, these results can be interpreted in two ways. While subjects could be using the explicit distractor “ignore” colour cue as an attentional suppression set for the duration of the experiment, our own work may suggest that the valid distractor cue may have been superfluous to the implicit development of the suppression of predictable visual input.

Behavioural facilitation via inter-trial priming has been demonstrated for target and distractors [52] and while some parts of the effects described above could be explained by such a mechanism a number of studies have explored the extent suppression relies on implicit expectations gleaned over experience. For example, in Lamy et al., performance was improved when search arrays contained the same distractor arrangement as a previous trial (akin to preview search effects [50]). They interpret this phenomena as a form of inhibition of return [58], that allows old, previously processed material to be ignored with subsequent presentations distractor location. However, the more complex relationship between expectation and past experience shown by Leber suggests predictions can be more abstract than simple repetition.

Over extended experience, distractor expectation may build-up over longer periods of time. For example Reder et al [59] showed that RTs in a target localisation task depend on the spatial predictability of the distractor. Moreover, Goschy et al [60]’s work went on to explore the extent the effects depended on distractor repetition. They showed that while the physical repetition of the distractor can result in distractor suppression, even when localised repetitions are removed there is residual, less-specific, expectation-based suppression. This body of work suggests that distractor suppression is a distinct mechanism from target enhancement, with distractor cueing only effective if the predicted feature is held constant over a number of trials. This is consistent with a predictive
coding model of expectation suppression.

Experience-based perceptual expectations might also dovetail with other selection biases for task-relevant information in which lingering effects of past experience guide subsequent attentional control and create integrated priority maps [61]. Previous studies have argued that selection history plays a key role in inter-trial priming [61]. The act of selecting (or rejecting) input from a previous trial effectively carries over to the current trial, akin to a task set inertia. Moreover, experiments like Leber et al [55] clearly demonstrate that trial history effects on distractor processing can extend beyond selection history, with complex distractor contingencies also being learned over time.

In summary, the behavioural evidence clearly demonstrates that distractor inhibition contributes to focused, goal-directed behaviour. However, in many circumstances, the mechanisms for inhibition seem to be distinct from those responsible for control over attentional facilitation. There is limited evidence for direct top-down control over selective preparatory inhibition (but see [19]). In many contexts, distractor inhibition could arise as a secondary consequence of target facilitation, and/or via expectation suppression. Secondary inhibition would be as flexible as target facilitation, but might not be selective for the most likely distractor (i.e., increase the spread of inhibition for all unattended items). In contrast, expectation suppression could be more selective for a specific distractor, but presumably depends on the statistics of the environment rather than flexible working memory based mechanisms more associated with goal-direct attention.

Neural mechanisms of Distractor Inhibition

In this section we review studies that shed some light on the potential neural basis of preparatory states for selective inhibition. To this end, we focus mainly on studies that examine activity prior to the onset of a to-be-inhibited distractor, rather than neural processing during distractor processing. It seems intuitive to look towards inhibitory neurons for clues on behavioural inhibition. For
example, direct inhibition could be mediated by inhibitory, long-range connections, however these are proportionally sparse [62], are under-studied and make up a small proportion of GABA-ergic neurons [63]. An alternative system might rely on long-range, excitatory neural pathways that could be functionally rendered inhibitory if it targeted local inhibitory populations. Around 20-30% of neurons are GABAergic interneurons [1], providing inhibitory synaptic input within local cortical microcircuits that are particularly important for winner(s)-take-all competitive dynamics, for example in the reactive spread of inhibition during perceptual processing [7, 64]. These connections could also be important for preparatory secondary inhibition as described above: top-down activation of a target-relevant representation would trigger a spread of inhibition over uncued representations.

Yet, neural inhibition need not be limited to a basic suppression of neural activity but could be determined by a broad range of possible neurophysiological states. In particular, information processing could be effectively suppressed by specific temporal patterns of activity, such as low frequency oscillations (i.e., pulsed inhibition, below) and/or noise correlations [2]. It is beyond the scope of this article to provide a comprehensive review of these potential mechanisms, but we will instead focus on the neural signatures more commonly associated with preparatory inhibition: reduced neural activity and increased oscillations in the alpha frequency band.

**Reduced activity and neural signatures of inhibition**

Functional magnetic resonance imaging (fMRI) studies provide some evidence for direct preparatory changes in visual activity prior to the appearance of a distractor at expected locations, although the overall picture remains complex and somewhat inconclusive. In the study by Ruff & Driver [43] already discussed, fMRI was used to measure preparatory states in visual cortex that underpin the behavioural benefit of distractor expectation. Curiously, they found that the expectation of a distractor actually increased BOLD activity in occipital cortex contralateral to the distractor, rather
than a BOLD suppression in anticipation of distracting input [also see 65]. The authors offer three potential explanations. First, they suggest that the distractor-related baseline shifts could indicate selective preparatory activation of inhibitory interneurons, which then has a secondary suppressive effect on pyramidal cells. Alternatively they suggest that anticipatory positive BOLD increases may index occipital “predictive coding” of the pattern of expected stimulation. Finally, anticipating a specific stimulus (even task-irrelevant input) could lead to mental imagery of the pattern of expected visual input. These latter two possibilities may not be mutually exclusive. It is possible that cue-related activity of this sort serves as the vehicle for distractor suppression by co-opting repetition suppression. If the cue triggers distractor-related activity prior to the presentation of the actual distractor, the same neurophysiological mechanisms that mediate repetition suppression could also suppress the neural response to the previously accessed distractor stimulus. Neuroimaging evidence suggest that even complex shape representations in visual cortex can be selectivity activated during mental imagery [66], which could presumably result in a repetition-suppression-like effect for subsequent presentation of matching sensory input. Indeed, classic studies in imagery suggest that prior access to a mental image can temporarily lower perceptual sensitivity for the corresponding stimuli (Perky effect) [67, 68], although this is not always the case [69] and effects may sometimes act to shift criterion and not sensitivity [70].

The difficulty interpreting the results from Ruff and Driver highlight some difficulties interpreting the BOLD signal, which reflects afferent input (postsynaptic effects) more than to the efferent output (i.e., spiking activity of projection neurons from that brain region to another) [71, 72]. With “apparent neural inhibition”, related to a relative reduction of fMRI signal potentially reflecting paradoxical ‘increased efficiency’ of information processing [73] or confounds such as differential blinking rates [74]. This makes it difficult to determine the net neural effect on the target region from the BOLD signal alone [75].
In general the BOLD response is likely to be dominated by activation of the more numerous pyramidal cells as recent methodological studies suggest that the net effect of inhibition is a reduction in BOLD response [76]. Indeed, negative BOLD response in the pulvinar [74] and early visual cortex, [77] is evident contralateral to an unattended hemifield. Interestingly, this BOLD suppression of visual cortex is reduced when subjects simultaneously perform a task that places a high load on working memory resources. St. John-Saaltink and colleagues have suggested that expectation and working memory rely on common processes to reinstate stimulus templates in visual cortex, with expectation mechanisms using these stimulus templates to compare subsequent bottom-up input. Indeed, a distinct set of posterior parietal regions are engaged when working memory contents is predicted to match targets and when working memory is predicted to match subsequent distractor items [78, 79]. Again, these effects could potentially suggest a degree of flexible control over working memory based control, or at least an interaction between expectation and working memory.

Less prone to some, but not all of these issues, are neurophysiological recordings made with EEG or MEG. Here then, direct distractor suppression has been associated with changes to particular components of the EEG waveform. Hickey and colleagues [80] have decomposed the N2pc, a lateralised component thought to reflect attentional selectivity, into the two separable components: Pd and NT. The Pd reflects active distractor suppression, while the NT is associated with target processing. Their evidence suggests working memory capacity modulates the former providing an argument for active suppression [81]. Increased activity in frontal electrodes have also been observed in a visual search task when high-value distractors, cued on a trial-wise basis, are behaviourally suppressed relative to low value and control conditions [82]. Again, invoking an interpretation based on enhanced working memory of distractor features to mediate direct distractor suppression. It is important to note, however, in both these experiments that the distractor colour feature was actually held constant across experiments, therefore leaving room for
an alternative explanation based on expectation suppression. Indeed, modulation of the Pd is seen after more predictive targets, potentially allowing more efficient suppression of salient distractors [83]. Although, intriguingly, salient angry faces, which capture attention, as indexed by amplified N2pc [84], also trigger a Pd suggesting that compared to happy faces they naturally induce more distractor suppression [85]. This effect appears fairly flexible in time and specific to the angry face and so whether it invokes specialised top-down directed suppression should be more explicitly explored.

In contrast to the studies reviewed above, others have argued that inhibitory-like effects could be explained more parsimoniously in terms of excitation alone. For example, some researchers contend that the onset of any stimulus will attract attention, regardless of top-down processes [86]. Tsal and Makovski [87] concluded both distractor and target stimuli capture attention as subjects were able to perceive a probe stimuli presented subsequent to both these events. Indeed others have argued that all cognitive inhibition, including preparatory inhibition, but notably not motor inhibition, can be accounted for solely by enhancing relevant information and that failure in this type of control is related to poor or disrupted working memory and not a failure to inhibit [88-90]. Egner & Hirsch [90] observed no suppression of fusiform face area (FFA) BOLD activity during a stroop-like task which showed subjects faces of actors or politicians with career congruent names (actor face + actor name) or incongruent names (actor face + politician name) written over the top: On some trials the subjects had to classify the face as either actor or politician and ignore the name (see figure 4). In incongruent trials the FFA recognises the face but the automatic processing of the name causes a response conflict. The authors found that the BOLD response in the FFA only increased when faces are relevant and attention is at a heightened state as a result of the previous trial being incongruent, while there was no change in activity when faces were irrelevant, perhaps suggesting that only excitatory processes were involved. However, it is worth reiterating that the difficulties interpreting the BOLD signal outlined above apply to this fMRI study as well.
The evidence for direct preparatory neural suppression is limited, but secondary inhibition (often referred to as lateral inhibition in the neural literature) is a well-studied phenomena [32]. In the context of distractor suppression, Sylvester et al. [20] found that anticipation of low contrast stimuli in the attended hemifield was associated with a relative reduction in visual cortical activity in the contralateral hemisphere corresponding to the unattended hemifield and a complementary increase in BOLD activity in the frontal eye fields and inferior frontal sulcus. Interestingly, the secondary suppression was predictive of task performance, and therefore unlikely to be epiphenomenal. These results suggest that when a stimulus will be difficult to distinguish from the background, top-down signals from frontal eye fields (FEF) and inferior frontal sulcus facilitate perception by increasing activity at the attended location in visual cortex, which results in a suppressive effect at unattended locations. On a more fine-grained scale, single unit recordings in macaques show lateral inhibition in dorsolateral prefrontal cortex (dLPFC) of distractor areas that were adjacent, in time and space, to target locations [91]. This secondary inhibition also correlated with error performance and dLPFC deactivation caused more distractor related errors. The authors argued distractor suppression arises from enhancement of target location and local competition. It would be important to investigate how these effects in prefrontal cortex manifest in early visual cortex, as recordings in lateral intraparietal area did not show the same pattern of distractor suppression and, compared to dLPFC, temporary cortical inactivation produced much smaller increases in distractibility.

Finally, it is well known that repetitions of a stimulus reduce neural activity [92]. It is therefore possible that the behavioural improvements in our own study described above were underpinned by reduced neural activity associated with the repeated distractor [54] with effects therefore interpretable within an expectation suppression framework. Analysis of the EEG data collected while
subjects performed the blocked version of the 4-location target discrimination task showed that distractor suppression in blocks when the distractor repeated to the same location was associated with a diminished P1, contralateral to distractor locations, and a loss of the N2pc. The P1 is an occipitoparietal component thought to reflect the first sweep of visual input and bottom up processing. Classically this component is positively modulated towards stimuli at attended locations and is thought to reflect sensory gain [93]. Suppression of the P1 is also associated with inhibition of distractor features [94]. Peripheral irrelevant probes shown in the same colour as task-irrelevant distractor (the distractor colour was fixed for the duration of the experiment) also produce a smaller contralateral P1 relative to non-distractor coloured probes.

Repetition suppression clearly offers one route for distractor inhibition yet stimulus repetition it is often confounded by increasing ‘higher order’ expectations [see 95 for extensive discussion]. Yet in some studies these factors have been shown to interact [96, 97] in an attention- [98] and experience-dependent manner [99]. Feldmann-Wustefeld and Schubo [100] have attempted to tease apart the contribution of these two factors on distractor induced ERP components. In heterogenous visual search tasks where singleton-distractor feature colour repeated for three trials, RTs were faster in predictive repetition compared to random trials. With increasing repetitions, this RT benefit was complemented by an increase in amplitude of the ND component, thought to reflect attentional capture by the distractor (1st>2nd>3rd). Repetition (even unexpected) had no impact on the Pd or NT (reflective of active distractor suppression and target-related activity respectively). The authors interpreted the findings by arguing that increased distractor expectation reduces attentional distractor capture, while the statistical or physical repetition of the distractor does not affect target processing or direct distractor suppression.

These findings, and others, may suggest a reinterpretation of a number of visual search tasks that keep the distractor colour feature fixed for the duration of the experiment (such as Gaspar’s [81]
task). It may be possible that attentional suppression templates build up over the course of the session, and trigger changes in neural signatures. Indeed, Seidl et al [101] has shown that pattern classification accuracy of BOLD signal in object-selective cortex of a currently irrelevant distractor stimulus (which had been a previous target stimuli) was suppressed relative to target and a never-relevant neutral stimulus. While the effects could depend on direct inhibition mediated by prefrontal regions such as FEF, the attentional sets were constant for the duration of an entire block and not newly instantiated on each trial potentially giving expectation-based mechanisms time to build up distractor predictions. However, such reversals between target and distractor categories, and the degree of dimensional similarity, can push the limits of inhibitory mechanisms. For example, if a past learning experience has defined target-relevance as one dimension (eg colour), then subsequent distractors defined by the same dimension capture attention to a greater extent and require more subsequent suppression, as indexed by greater ND and Pd components respectively, than if the target and distractor dimensions do not match [102].

Inhibition related to neural oscillations

Oscillations in the alpha band (8-12Hz) might play a special role in setting the preparatory state for selective inhibition. At first it might appear counterintuitive that the strongest electrophysiological signal produced by the brain reflects functional inhibition. Nevertheless, there is converging evidence supporting the inhibitory role of alpha oscillations at the physiological level. For instance, an increase in alpha band power has been demonstrated to be inversely correlated with firing rate in monkey sensorimotor regions [103, figure 5A]. Furthermore, an increase in alpha power predicts a decrease in gamma band activity [104, 105]. Consistently the BOLD activity is inversely correlated with alpha band power [106, 107].

Figure 5
Human studies have supported this notion for visual and somatosensory regions making use of the fact that alpha activity is strongly hemispherically lateralized with respect to covert attention. In an early EEG study, Worden et al [108] demonstrated a clear link between alpha band oscillations and neural inhibition. In anticipation of a cued target the authors observed a focal and sustained decrease of alpha power in electrodes contralateral to the target. Importantly this was complemented by a relative increase in alpha band power at electrodes over ipsilateral visual cortex, i.e. over visual cortex where the target was not expected. Such a pattern has been argued to reflect an active attentional suppression mechanism [109]. Sauseng et al. [110] report similar findings but also found evidence that alpha band activity in frontal and occipital electrodes contralateral to the cued visual hemifield were phase-locked, consistent with frequency-specific functional coupling between these areas. Consistently, a recent MEG study relying on Granger causality suggested that FEF drives the alpha oscillation during spatial cueing [111]. Furthermore, alpha band oscillations are relatively feature specific and cross-modal in nature [112-114, 115, figure 5B], with the same pattern for the somatosensory system [103]. Mechanistically, while direct inhibition may account for these findings, feature-specific inhibitory alpha band oscillations may also fit within the secondary inhibition framework, with all non-target features being subjected to preparatory suppression via intrinsic local competition [112].

More recently, a number of studies describe associations between alpha oscillations and behavioural performance. For example, Payne et al [116] serially presented subjects two Gabor patches, having been pre-cued as to which was to be remembered and which was to be ignored. Subjects were then asked to report the spatial frequency of the remembered Gabor. Payne and colleagues demonstrated that greater alpha power prior to the onset of a distracting stimulus was associated with less intrusion of that Gabor’s spatial frequency into the reported frequency of the target. Consistently, alpha power prior to the onset of anticipated distractors has been suggested to be modulated to protect against expected inference during a working memory delay [117, see figure
Moreover, pre-stimulus alpha power appears to help gating task-relevant information in working memory [118]. It has been demonstrated that when subjects have to perform a working memory or attention task involving somatosensory stimuli presented to one hand, the alpha power in the task-irrelevant ipsilateral regions was predictive of task performance [103]. Theses links between alpha, inhibition and working memory could again imply a role for alpha in direct or secondary forms of suppression. In sum, this set of studies suggests that alpha activity serves to inhibit specific regions not required for a given task. If this is the case then increasing alpha power in these regions should causally impact behaviour. Indeed this has been demonstrated in a number of studies that used TMS to directly entrain alpha band activity [e.g. 119]. For example, performance is improved after repeated stimulation of parieto-occipital regions at 10 Hz ipsi-lateral to attended objects [120, 121].

Given its involvement in suppressing information processing, selective decrease of alpha oscillations could bias processing by releasing local neuronal populations coding the attended location from inhibition. Indeed, there is evidence to suggest that alpha oscillations operate at this more fundamental level. For example, not only does the spontaneous fluctuations in alpha synchronisation correlate with perceptual sensitivity [122-124], but also the likelihood a TMS-induced phosphene will be elicited [125]. In a particularly elegant study, Romei and colleagues [126] unilaterally stimulated occipital and parietal cortex with 5 TMS pulses at a frequency of either 5, 10 or 20 Hz prior to the onset of a target stimulus. They found that 10 Hz stimulation over the region of cortex ipsilateral to the target led to an enhancement of target detection, whereas contralateral stimulation led to a decrease in target detection. This could be interpreted as increased inhibition of non-target visual areas, thus improving performance during ipsilateral stimulation; whereas contralateral stimulation presumably impaired performance by increasing inhibition of target regions. Furthermore, the ipsilateral enhancement effect was significantly greater at 10 Hz
compared to 5 or 20 Hz stimulation, i.e. showing the specificity of the effect within the alpha frequency range.

Identifying the source of alpha modulation could provide additional clues as to its mechanistic role. While clearly manifesting locally, there is evidence that top-down control may provide an additional route for organising the alpha state to suppress expected distractors (selective increase in alpha power, and phase alignment) whilst optimising processing for task-relevant input (selective decrease in alpha power, and/or phase alignment). For example, interference TMS delivered to FEF perturbs the modulation of posterior alpha activity by attention [127, 128]. At this level, top-down alpha modulation could bias processing by directly suppressing task-irrelevant neural populations, and/or selectively releasing from inhibition neuronal populations coding task-relevant features. This top-down mechanism might rely on a phasic control of posterior alpha oscillations in posterior regions [129], potentially driven by FEF [111] via the dorsal part of the superior longitudinal fasciculus [130].

When considering subcortical pathways modulating the posterior alpha band activity, a strong case has been made for a role of the pulvinar. Recordings in dogs and monkeys have demonstrated that intracranially recorded field potentials in the pulvinar are phase-locked to field potentials in visual neocortical regions [131]. Analysis of multi-electrode recordings suggested that the pulvinar synchronized ventral visual regions in the alpha band; this synchronization allowed the neocortical regions to communicate directly. Thus while some insight has been gained on how posterior brain regions are modulated by top-down control, a lot more work is required in order to identify the networks exercising the control and by which mechanisms.

Collectively this work has led Jensen and Mazaheri [2] to describe a pulsed inhibition hypothesis which emphasizes the importance of alpha band oscillations for inhibition. Figure 5B illustrates how phasic inhibition in the alpha band serves to increase oscillatory modulation at the population level.
This is based on the assumption that measured EEG and MEG signals are a consequence of population activity generating dendritic currents in pyramidal cells in deeper cortical layers. These dendrites’ currents are most likely a consequence of synaptic distal inputs and after-hyperpolarizing currents [132, 133]. Consider pulses of inhibition in the alpha band. When the pulsed inhibition is relatively low, firing is hardly modulated (Figure 5B). As a result, there is only a modest oscillatory modulation at the population level. However, when the inhibitory pulses are strong, the blocking of neuronal spiking in each alpha cycle produces a signal strongly modulated by the alpha band albeit the total firing rate is decreased. The gamma band activity is modulated by the alpha oscillations in a phasic, task-dependent manner as well [134-137]. The lower the inhibitory pulses, the more gamma cycles are assumed to occur in each alpha cycle [104]. In short, the proposed framework suggests that as the amplitude of oscillations in the alpha band increase, neural assemblies become more restricted in when they can fire, resulting in a net increase in inhibition.

While the role of alpha in distractor inhibition has been relatively well explored and appears particularly suited to direct and secondary frameworks of inhibition, neural oscillations in other frequency bands may act in complementary ways and support inhibition through other mechanisms. For example, while we observed that expectations learned from past-experience can be used to coordinate the distribution of alpha for optimising processing in favour of task-relevant relative to irrelevant spatial locations [42, 54], we found no evidence for lateralisation of preparatory alpha when distractors are expected and repeated [54]. By contrast, theta oscillations (4-8Hz) are related to memory and cognitive control and may support maintenance of feature or task-sets and therefore might be a good candidate for mediating expectation suppression. Theta increases in the presence of a distractor [138], and when the salient distractor, is behaviourally suppressed relative to low value and control conditions [82]. Further, prestimulus theta is modulated according to expectations for task-relevant features [139] and stimulus-induced frontocentral theta is associated with expectation suppression and is sensitive to frequent rather than rare repetitions of attended
target stimuli [96]. Finally, even lower frequency oscillations in the delta band (1-4Hz) over visual cortex are associated with temporal expectations of relevant information, with the phase of the oscillation predicative of the quality of subsequent target processing [140]. It is important to further explore the roles of these oscillations in distractor inhibition and relate them more directly to functional mechanisms.

Different Mechanisms: A mutual effect

Although distractor suppression can be considered as the flip side to facilitation, it is more likely that the route to distractor suppression can follow a number of paths. While some circumstances might allow direct inhibition an alterative route would invoke the rapid, flexible but potentially non-specific action of secondary inhibition that may result from a complementary spread of inhibition to unattended features via local inhibitory circuits. Further, we argue that more specific inhibition could be established via less flexible mechanism of expectation suppression with its principles governed by predictive coding is also acting to suppress distraction.

It is worth considering how these potential mechanisms account for reported findings. The crucial distinction we propose is that the models differ in terms of flexibility, specificity and selectivity of suppression. Direct inhibition would be expected to be flexible, and specific for the exact distracting input that is cued by preparatory control mechanism. By contrast, secondary inhibition might be as flexible as direct inhibition as it depends on the initial facilitation of task-relevant items but relatively non-specific as it would inhibit all uncued locations. Finally, expectation suppression may be relatively specific to any repeating stimulus, but requires more explicit task experience to manifest.

Although fundamentally different from one another, the mechanisms might, in some cases, make subtly similar predictions. All would predict facilitation in RT at the behavioural level when prior information on distractors is given. All mechanisms may involve top-down control, although this may
be more long-ranging in the case of top-down secondary inhibition (signals from FPC to lower sensory regions) and more local in the case of expectation suppression (predictions coming from the next ‘level’ in the processing hierarchy). For secondary and expectation mechanisms, behavioural effects should increase with time, with greater predictability and less volatility in the environment (or task), although they may manifest over different time scales (trial-wise explicit cueing may be faster vs accumulating lower-level predictions over trials/experience). This may reflect anatomically similar routes of action, activating frontoparietal networks during distractor cueing and would therefore be equally sensitive to interference of prefrontal cortex through TMS.

Importantly, the mechanisms may differ in the time scale between the balance of excitation and inhibition: Facilitation may precede inhibition in secondary inhibition models [141] but this pattern may be expected to reverse in expectation suppression accounts. On-going tasks demands may differentially impact the mechanisms. For example working memory load may be expected to affect expectation suppression if templates are compared to bottom-up input [77] but leave secondary suppression mechanisms unaffected. Furthermore, the three mechanisms may not be mutually exclusive, with the balance between them varying as a function of on-going demands. While it may be challenging to tease these explanations it will be important to elucidate the different mechanisms, both when they act independently and when they potentially interact to support behaviour.

Acknowledgements

This work was supported by an early career fellowship from St John’s College, Oxford (M.P.N.) and a Medical Research Council Career Development Award (M.G.S). This research was funded by an early career fellowship from St John’s College, Oxford (M.P.N.) and Medical Research Council Career Development Award (MR/J009024/1) to M.G.S. and supported by the NIHR Oxford Health Biomedical Research Centre. The Wellcome Centre for Integrative Neuroimaging is supported by core funding from the Wellcome Trust (203139/Z/16/Z).
References


32


Figure legends

Figure 1. Three general mechanisms for preparatory inhibition, illustrated by spatially defined target/distractor input. [A] Simple task where the participant is required to report the orientation of the grating on the left of the screen. They know in advance that the right side will be a distractor stimulus. [B] Direct Inhibition: neurons within the left visual cortex representing the distractor are specifically inhibited via top-down connections from fronto-parietal cortex. [C] Secondary Inhibition: There is no direct top-down inhibition of left visual cortex, but instead only top-down excitation of right visual cortex. Inhibition arises through biased competition process at lower levels. [D] Expectation suppression: Predictable inputs to the visual system are suppressed through inhibitory connections from within the hierarchy of the visual system in a manner described by predictive coding. In the absence of additional top-down facilitation, representations of the predictable distractor remain suppressed. Green arrows indicate excitatory connections, red circular arrows indicate inhibitory connections.

Figure 2. [A] Explicit cueing of distractor likelihood. Participants made target discrimination judgements after one of three central arrows cued target hemifield, and provided either advanced knowledge of the presence (Ci, central) or absence (Ci, left) of a distractor subsequently presented in the opposite hemifield, or no information (Ci, right). Following the cue, a single target checkerboard, or a target with a distractor on the other side. Subjects indicated the presence of a deviant square in the target checkerboard. Results indicated that the latency cost of adding a distractor was significantly reduced on trials with advance knowledge of distractor location (grey bar in right pair) [43]. [B] The benefit of predicting distractors. Subjects were asked to attend two target locations for the presentation of a number stimulus. In some blocks, there was a high probability (80%) of accompanying distractor letters (shown here, in the upper panel), whereas in other blocks the probability of distractors was low (20%). Although distractors generally impaired behaviour, this cost was reduced on expected distractor trials (lower panel) [21]. Behavioural evidence for distractor inhibition. [C] Previewing distractors reduces the search costs to the single feature search level [50]. Subjects were presented with a search array in which the target was either defined by a feature singleton (pop-out) or a conjunction of features (conjunction; see upper panel). Prior presentation of
a subset of distractors dramatically reduced the usual cost of the additional distractors in the conjunction task (lower panel).

Figure 3 Suppressing expected distractors [A]. Evidence that distractor suppression is not under direct top-down control. In this study, subjects were informed about the cue context (target position, distractor position, or uninformative) before each block (left panel). Each trial consisted of a sequential pair of displays: the cue (its location around the fixation cross indicative of the quadrant location of the forthcoming target or distractor) and the target stimulus (oriented grating). On 50% of the trials a distractor (two overlapped gratings) accompanied the target. In the Flexible Cueing task cues varied predicted location on a trial-wise basis. In the Blocked Cueing task cues to a predicted location were valid for the duration of the block of trials. RT data in the right panel shows subjects cannot use the knowledge of forthcoming distractor location to improve task performance in Flexible Cueing task but can when that information is stable across the course of many trials in the Blocked task [54]. [B] Distractor suppression is flexible under some circumstances. In this study distractor locations (indicated by position of red D) were implicitly associated with cued target placeholders (indicated by arrow; four examples of implicitly associated target-distractor locations are illustrated in the left panel). If the location of a distractor is validly linked to the location of a cued target (Tvalid-Dvalid), distractor locations can be flexibly inhibited compared to if the distractor appears at an invalidly implicitly cued location (Tvalid-Dinvalid) right panel, adapted from [55]).

Figure 4. Distractor suppression reduces neural activity. [A] In this study subjects discriminated between actors and political figures based on either the face stimulus (face-target) or the written name (face-distractor) (left panel). Stimuli could be category-congruent (face and name matched) or category-incongruent (face and name mismatched). Results are shown in the two right panels illustrating the mean FFA activation for current congruent and incongruent trials as plotted as a
function of previous-trial congruency (x-axis) for face-target and face-distractor condition [90]. EEG results from [54]. Repeated distractors result in a diminution of the P1, contralateral to distractor locations (left panels), and a reduction of the N2pc (right panel).

Figure 5. Modulating alpha power for distractor inhibition [A] As alpha power increases, the firing rate decreases. These data were obtained from intracranial recording in sensorimotor regions of the monkey [reproduced from 103]. [B] A schematic diagram explaining why neuronal firing might be negatively related to alpha power. To the left, the group of neurons is firing continuously. The alpha oscillations emerges due to a rhythmic silencing of the neuronal firing (adapted from [115]. [C] In a Sternberg working memory task, distractors were introduced in the retention interval. The occurrence of either strong or weak distractors could be anticipated. The time-frequency of power demonstrated that alpha power was higher when strong distractors could be anticipated [adapted from 117]. [D] In an attention task, oscillatory activity was entrained by TMS contra- or ipsilateral to the target occurring in the left or right hemifield. The detection ability of the target increased when the ipsilateral hemisphere was entrained at 10 Hz, but not 5 and 20 Hz [reproduced from 126].