



# Neural evidence that inhibition is linked to the affective devaluation of distractors that match the contents of working memory



David De Vito, Naseem Al-Aidroos, Mark J. Fenske\*

Department of Psychology, University of Guelph, Guelph, Ontario, Canada N1G 2W1

## ARTICLE INFO

### Keywords:

Distractor devaluation  
Evaluative codes  
Devaluation-by-inhibition  
Attentional inhibition  
Affective evaluation  
Working memory  
EEG/ERP  
Pd component

## ABSTRACT

Stimuli appearing as visual distractors subsequently receive more negative affective evaluations than novel items or prior targets of attention. Leading accounts question whether this distractor devaluation effect occurs through evaluative codes that become associated with distractors as a mere artefact of attention-task instructions, or through affective consequences of attentional inhibition when applied to prevent distractor interference. Here we test opposing predictions arising from the evaluative-coding and devaluation-by-inhibition hypotheses using an electrophysiological marker of attentional inhibition in a task that requires participants to avoid interference from abstract-shape distractors presented while maintaining a uniquely-colored stimulus in memory. Consistent with prior research, distractors that matched the colour of the stimulus being held in memory elicited a Pd component of the event-related potential waveform, indicating that their processing was being actively suppressed. Subsequent affective evaluations revealed that memory-matching distractors also received more negative ratings than non-matching distractors or previously-unseen shapes. Moreover, Pd magnitude was greater on trials in which the memory-matching distractors were later rated negatively than on trials preceding positive ratings. These results support the devaluation-by-inhibition hypothesis and strongly suggest that fluctuations in stimulus inhibition are closely associated with subsequent affective evaluations. In contrast, none of the evaluative-coding based predictions were confirmed.

## 1. Introduction

Stimuli appearing as visual distractors subsequently receive more negative affective evaluations than novel items or prior targets of attention (Raymond et al., 2003). Such distractor devaluation effects have been found across a wide range of attention tasks requiring stimulus discrimination based on features, categories, spatial location, or temporal position (e.g., Fenske et al., 2004; Goolsby et al., 2009b; Kihara et al., 2011; Martiny-Huenger et al., 2014; Raymond et al., 2005), and have been obtained with a variety of stimuli including simple abstract patterns and shapes (e.g., Raymond et al., 2003), alphabetic and logographic characters (e.g., Martiny-Huenger et al., 2014; Veling et al., 2007), common objects (e.g., Griffiths and Mitchell, 2008), corporate brands (e.g., Duff and Faber, 2011), and human faces (e.g., Raymond et al., 2005). The distractor devaluation effects in these studies have been observed as changes in a number of different subjective emotional judgments, including stimulus valence, likeability, favorability, beauty, cheerfulness, dreariness, pleasantness, and trustworthiness.

### 1.1. The devaluation-by-inhibition hypothesis

The distractor devaluation effect has been taken as evidence that attentional inhibition, presumably applied to prevent distractor interference, has negative affective consequences for associated stimuli (see Fenske and Raymond (2006), Gollwitzer et al. (2014), Raymond (2009) for reviews). Support for this devaluation-by-inhibition hypothesis has come from both cognitive-behavioural and neuroimaging studies. The cognitive-behavioural studies (e.g., Fenske et al., 2005, 2004; Raymond et al., 2005) utilized experimental conditions that are considered to vary in the level of inhibition required for successful task completion. In subsequent evaluations, the stimuli in conditions thought to involve greater inhibition consistently received more negative ratings than those in conditions thought to involve less inhibition. The results of Raymond et al. (2005) and Martiny-Huenger et al. (2014), for example, both supported a key prediction arising from the devaluation-by-inhibition hypothesis; namely, because distractors located close to a target of selective attention have greater interference potential and are therefore subjected to greater levels of inhibition than items further away (Cutzu and Tsotsos, 2003; Hopf et al., 2006; Mounts, 2000), they

\* Corresponding author.

E-mail address: [mfenske@uoguelph.ca](mailto:mfenske@uoguelph.ca) (M.J. Fenske).

should also subsequently receive more negative stimulus ratings than distant distractors.

Other findings from a line of studies involving Go/No-go and Stop-Signal tasks (Fenske et al., 2005; see Gollwitzer et al. (2014) for review; Kiss et al., 2008; Wessel and O'Doherty, 2014) suggest that motor-response inhibition also has deleterious consequences for associated stimuli. This raises the possibility that stimulus devaluation results from neurocognitive inhibition regardless of the specific processing level (i.e., sensory/perceptual, response/action, memory/cognition, etc.) at which it is applied. Converging support for this more general version of the devaluation-by-inhibition hypothesis comes from experimental paradigms designed to explore the inhibition of stimulus memories (e.g., Think/No-think, Directed Forgetting), which have likewise provided results suggesting that stimuli whose memories have been inhibited become affectively devalued (De Vito and Fenske, 2017; Vivas et al., 2016, cf. Janczyk and Wühr, 2012). Taken together, the cognitive-behavioural results obtained following a variety of manipulations of attention-, response-, and memory-related inhibition broadly support the hypothesis that cognitive processes of inhibition have negative affective consequences for associated stimuli.

The devaluation-by-inhibition hypothesis has also received support from the results of neuroimaging investigations into the affective consequences of selective attention (Kiss et al., 2007) and response inhibition (Doallo et al., 2012; Kiss et al., 2008) in tasks using visual stimuli. Kiss et al.'s (2007) electroencephalography (EEG)-based approach, for example, involved an assessment of the N2pc component of the event-related potential (ERP) waveform during periods in which participants ignored a distractor while searching for a stimulus with a pre-specified target feature. They found that differences in the emergence of the N2pc component—widely considered to provide an index of the efficiency with which attention was selectively allocated to a target stimulus in the presence of visual distractors (e.g., Eimer, 1996; Luck and Hillyard, 1994)—were linked to the magnitude of distractor devaluation. The N2pc component on trials containing distractors that subsequently received the most negative ratings emerged significantly earlier than the N2pc on trials containing distractors that subsequently received more positive ratings. To the extent that trial-by-trial variations in N2pc onset reflect the efficiency of attentional selection aided by distractor inhibition (Luck and Hillyard, 1994), these results are consistent with the notion that the level of attentional inhibition applied to a distracting stimulus is linked to the magnitude of its subsequent affective devaluation. Nevertheless, inhibition is not the only factor that has been proposed as an explanation for the distractor devaluation effect.

### 1.2. The evaluative coding hypothesis

A leading alternative to the devaluation-by-inhibition account of distractor devaluation is Dittrich and Klauer's (2012) proposal, based on Eder and Rothermund's (2008) evaluative coding principle, that prior distractors are only rated more negatively than prior targets because of evaluative codes that become associated with targets and distractors solely through attention-task instructions. According to this perspective, merely instructing participants to seek, attend to, select, respond to or otherwise accept some items (targets), and to ignore, avoid, or otherwise reject other items (distractors) is sufficient to impact subsequent ratings because of prior associations linking stimulus approach to positive items and stimulus avoidance to negative items (Chen and Bargh, 1999). Consistent with this view, Dittrich and Klauer observed that stimulus ratings in two experiments were critically affected by manipulations of the attention-task instructions that altered the evaluative meaning of targets and distractors. Indeed, differences in stimulus ratings in their studies were primarily linked to instruction-driven connotations about which items were irrelevant and to-be-rejected or relevant and to-be-accepted, rather than the actual attentional status of a given item as a prior target or prior distractor as would

have been predicted by the devaluation-by-inhibition hypothesis.

In reconciling Kiss et al.'s (2007) electrophysiological findings regarding a link between the N2pc component and distractor devaluation with their evaluative-coding hypothesis, Dittrich and Klauer (2012) argue that fluctuations in the efficiency of selective attention, as reflected by N2pc latency, may impact how well targets and distractors can be perceptually discriminated and therefore how likely each could be clearly coded as something to-be-accepted or to-be-rejected. Thus, the evaluative-coding account posits that the effects of selective attention on subsequent stimulus ratings has nothing to do with distractor inhibition and everything to do with the strength of the association between each item and the evaluative codes that become attached during the selection process. The strength of conclusions about the specific link between distractor inhibition and affective devaluation that can be drawn from Kiss et al.'s (2007) results has also been affected by the emergence of findings that question the specific link between the N2pc and distractor suppression. Indeed, there is growing consensus that the N2pc component, while often emerging under conditions involving distractor inhibition, is not itself an index of inhibition but more directly related to enhanced processing of target information (e.g., Mazza et al., 2009). Fortunately, other advances in cognitive-electrophysiology include the identification of a marker that more directly reflects distractor inhibition during selective attention (e.g., Hickey et al., 2009). This is known as the Pd component—a positive-voltage deflection in the ERP waveform appearing 150–300 ms post-stimulus over regions of visual cortex contralateral to the location of a visual distractor—and is now widely thought to indicate the termination of attention to distracting stimuli and the active suppression of their perceptual processing (see Sawaki and Luck (2014) for review). The benefit of having this more direct neural index of distractor inhibition is that it enables a more direct test of the devaluation-by-inhibition hypothesis. We report the results of this test here, which we conducted using a paradigm that was selected for its ability to simultaneously provide an assessment of competing predictions arising from the evaluative-coding and devaluation-by-inhibition hypotheses.

### 1.3. Tests of competing predictions

The experimental procedure used in our study is a modified version of that developed by Sawaki and Luck (2011) to investigate the potential involvement of attentional inhibition in protecting the contents of visual working memory from the interfering effects of salient distractors. The main task requires participants to first select one of two visual stimuli and then maintain it in working memory throughout a retention interval. This would allow them to accurately choose this memorized item from a subsequent test display that includes the original plus a rotated version of it. The ability to correctly remember the item on these memory trials is complicated by the appearance of a pair of task-irrelevant distractors during the retention interval—one on each side of fixation. Moreover, the potential for interference is increased by always having one of the distractors be the same colour as the memorized item. While such memory-matching stimuli have been shown to initially capture attention (e.g., Soto et al., 2005), Sawaki and Luck found that their perceptual processing is ultimately actively suppressed, as evidenced by a Pd contralateral to the memory-matching distractor. This finding suggests that the level of suppression required to prevent the memory-matching distractor from interfering with the active contents of working memory was greater than that required to prevent interference from the non memory-matching distractor.

We adapted Sawaki and Luck's (2011) main experimental task to incorporate affective evaluations of stimuli that had previously appeared as either a memory-matching distractor or non-matching distractor during a memory-retention interval. Beyond providing an experimental context for assessing possible links between fluctuations in an electro-cortical index of distractor inhibition and subsequent

distractor ratings, the potential usefulness of this specific task concerns the possible evaluative codes associated with the different types of stimuli (Dittrich and Klauer, 2012; Eder and Rothermund, 2008). For example, the memory-matching distractors in Sawaki & Luck's task share a defining feature and are therefore quite similar to the to-be-accepted and protected memory target. According to the account proposed by Dittrich and Klauer (2012), anything indicated by the task instructions as a to-be-accepted target should be coded as affectively positive based on long-standing associations between behavioural approach and positively-valenced stimuli (Chen and Bargh, 1999). The corresponding impact of an item's to-be-accepted vs. to-be-rejected status on subsequent affective ratings extends to other stimuli that share a defining feature (Goolsby et al., 2009a; Raymond et al., 2003). Therefore, a memory-matching distractor in Sawaki & Luck's task should later receive a relatively positive evaluation, according to the evaluative-coding perspective, based on its similarity to the positively-labeled attentional target being held in memory.

The initial commit-to-memory display in Sawaki and Luck's (2011) task contains both the to-be-memorized item along with another do-not-memorize item. According to Dittrich and Klauer's (2012) evaluative-coding based account, anything the task instructions indicates is to-be-ignored and rejected should be coded as affectively negative based on long-standing associations between behavioural avoidance and negatively-valenced stimuli (Chen and Bargh, 1999). And just as the colour of the attended and memorized item was the same as the subsequent memory-matching distractor, the colour of the ignored do-not-memorize item was always the same as the other (memory non-matching) distractor. Thus, in addition to the potential impact on participants' stimulus ratings of the similarity between memory-matching distractors and the accepted and positively-coded memorized items in our version of the task, the ratings should also—according to the evaluative coding hypothesis—be affected by the similarity between the memory non-matching distractors and the rejected and negatively-coded non-memorized memory-display items. In sum, the evaluative coding hypothesis predicts that memory-matching distractors would subsequently receive more positive affective ratings, and that memory non-matching distractors would subsequently receive more negative affective ratings.

We included previously-unseen novel stimuli in our affective rating trials to provide a baseline against which the evaluative-coding based predictions—more positive ratings for memory-matching distractors, more negative ratings for non-matching distractors—could be properly assessed. These predictions stand in stark contrast to the inhibitory-devaluation based predictions of more *negative* ratings for memory-matching distractors than for novel items. Moreover, the devaluation-by-inhibition hypothesis predicts that trial-by-trial fluctuations in the magnitude of the Pd component may be linked to the level of subsequent devaluation of memory-matching distractors. The possible impact of such Pd fluctuations is less clear from an evaluative-coding perspective. Dittrich and Klauer's (2012) interpretation of Kiss et al.'s (2007) finding of a link between the N2pc and devaluation is that the more effective attentional selection is, as indexed by an earlier N2pc, the more clearly a stimulus can be coded as something to-be-accepted or to-be-rejected. To the extent that distractor suppression also aids stimulus discrimination, a larger Pd component—according to an evaluative-coding perspective—might indicate better differentiation of the memory-matching and non-memory matching distractors and their respective similarity to the positively-coded memorized item vs. the negatively-coded do-not-memorize item. From this perspective, trial-by-trial fluctuations in the magnitude of the Pd component, should be linked to the extent to which the memory-matching distractors receive more positive ratings than novel stimuli and the extent to which the non-memory matching distractors receive more negative ratings than novel stimuli.

## 2. Experiment 1 – cognitive-behavioural

We begin our assessment of opposing predictions arising from the devaluation-by-inhibition and evaluative-coding accounts of the distractor devaluation effect using a purely cognitive-behavioural adaptation of Sawaki and Luck's (2011) experimental procedure. We specifically adapted the procedure to obtain affective ratings of stimuli previously appearing as retention-interval distractors and of previously-unseen novel stimuli. Finding that prior memory-matching distractor ratings are more positive than novel-stimulus ratings would be consistent with an evaluative-coding account of distractor devaluation, as would a finding of more negative ratings for non-matching distractors than for novel-stimuli. In contrast, finding that ratings of prior memory-matching distractors are more *negative* than those of novel stimuli would instead be consistent with a devaluation-by-inhibition account. As for ratings of prior non-matching distractors, distractors with little interference potential are often not significantly devalued (e.g., Martiny-Huenger et al., 2014; Raymond et al., 2005 who argue this is due to the fact that such items evoke such low levels of inhibition), so there is less reason from a devaluation-by-inhibition account to expect that non-matching distractor ratings will reliably differ from those of novel stimuli.

### 2.1. Method

#### 2.1.1. Participants

A total of 52 undergraduate students from the University of Guelph participant pool (Age:  $M=18.4$  years,  $SD=1.44$ ; 35 females, 47 right-handed) volunteered to complete the experiment in exchange for course credit. The size of this sample was chosen to be consistent with the samples of 40–50 participants used in previous cognitive-behavioural investigations of the devaluation-by-inhibition hypothesis (e.g., Fenske et al., 2004; Ferrey et al., 2012; Raymond et al., 2005). The appropriateness of this sample size was confirmed using G\*Power (Faul et al., 2007), which indicated that 42 participants would be required to detect a similar-sized effect on stimulus ratings as in these studies ( $d_z=.5$ ) with a power of .95 at an alpha of .05. All of the participants provided informed consent and all had normal or corrected-to-normal vision.

#### 2.1.2. Apparatus and stimuli

Stimulus presentation and behavioural response collection were controlled by PsychoPy software (Peirce, 2007) running on an Intel Core2Duo computer with a 50.8 cm LCD monitor. All stimuli were presented on a black background in a room with standard fluorescent illumination at a viewing distance of approximately 60 cm.

The stimuli for this experiment were also generated using PsychoPy software (Peirce, 2007). The cue presented at the beginning of each memory trial was a  $1.49^\circ \times 1.13^\circ$  rectangle divided along the vertical midline into a dark grey half (RGB: 117, 117, 117) and a white half (RGB: 255, 255, 255) (see Fig. 1). Arial font was used for the central fixation cross (a '+' symbol), and the rating scale numbers presented during evaluation judgments, each subtending approximately  $.90^\circ \times .90^\circ$  visual angle.

The main stimuli used in the memory-trials and stimulus-rating trials were irregular abstract shapes that each subtended approximately  $4.65^\circ \times 4.65^\circ$  and were designed to each be distinct in appearance. These unique shapes were created by randomly selecting the size (range:  $1.66\text{--}3.32^\circ$  horizontal and vertical visual angle) and number of sides (range: 5–8) of four overlapping polygons. The outline created by these overlapping polygons formed the outer contour of the stimulus, which was then filled with a single uniform colour. Five-hundred and twenty-three different shapes were created using this method. Three different versions of this set of shapes were produced by changing the uniform colour of each shape to green (RGB: 39, 103, 23), orange (RGB: 240, 111, 35), and purple (RGB: 130, 1, 200). Having each of the 523 shapes in each of the three different colours resulted in

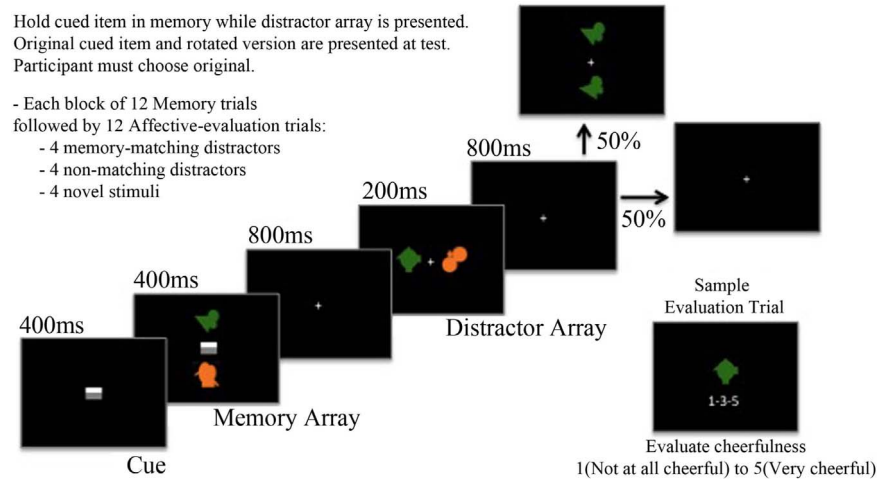


Fig. 1. Experiment 1 trial sequence. Participants memorized a cued item and later, on tested trials, indicated which presented item exactly matched the memorized item.

a final stimulus set of 1569 images.

### 2.1.3. Experimental procedure

All of the materials and procedures used in this experiment were approved by the Research Ethics Board at the University of Guelph. Beyond the use of distinct shapes as stimuli on each trial, the memory-maintenance procedure used by Sawaki and Luck (2011) was also modified to incorporate affective stimulus evaluations. The resulting procedure consisted of eight blocks of trials, each consisting of 12 Memory trials followed by 12 Affective-evaluation trials. Fig. 1 depicts the sequence of stimulus events used in Experiment 1.

Each memory trial began with the presentation of the half-grey/half-white rectangular cue in the center of the screen that indicated the location of the to-be-memorized item (i.e., top or bottom half of the screen). Half of the participants were instructed to use the dark grey half of the rectangle as a cue to determine whether to encode the item on the top or bottom of the screen into memory, while the other half used the white half of the rectangle as their cue. The cue was presented alone for 400 ms, and remained visible throughout the subsequent 400 ms presentation of the memory array. The memory array comprised two abstract shapes, each centered  $4.67^\circ$  degrees above or below display center. The colour of the to-be-memorized item (i.e., the shape appearing in the cued location) was different from the colour of the to-be-ignored item (i.e., the shape appearing in the uncued location). A fixation cross replaced the memory array and remained visible throughout an 1800 ms retention interval. The first 800 ms of the retention interval consisted only of the fixation cross, followed by a 200 ms presentation of two distractors—each centered  $4.67^\circ$  to the left or right of fixation. One of the distractors (memory-matching) was the same colour as the to-be-memorized item from the memory array. The other distractor (non memory-matching) was the same colour as the to-be-ignored item from the memory array. The two memory and two distractor stimuli all had different shapes. The relative locations of the memory-matching and non-matching distractors was counterbalanced to ensure that each appeared to the left of fixation on half the trials and to the right of fixation on half the trials. The specific stimulus colour used for the memory-matching distractors (e.g., green) was the same on every trial. Another stimulus colour (e.g., orange) was used for the non memory-matching distractors, which was also the same on every trial. Stimuli in the remaining colour (e.g., purple) were reserved for use as novel items in the Affective-evaluation trials. The appearance of the distractor array was followed by the final 800 ms of the retention interval, during which only the fixation cross was visible.

The specific colour-to-condition assignments were fully counterbalanced across participants such that each stimulus colour was used equally often for each experimental condition. Having the stimuli

within each experimental condition all be the same unique colour was critical for a fair test of the evaluative-coding hypothesis by ensuring that any affective connotations of being associated with the to-be-accepted and protected class of stimuli (to-be-memorized items and memory-matching distractors) vs. the to-be-ignored and rejected class of stimuli (do-not-memorize items and non-memory-matching distractors) could be clearly linked to a salient defining stimulus feature (Dittrich and Klauer, 2012).

The retention interval was followed by one of two occurrences. On half of the trials, memory maintenance was followed by a memory test. During this memory test, a memory-test array was presented, which contained the to-be-memorized stimulus along with an altered version of that stimulus (randomly rotated  $20^\circ$  clockwise or counterclockwise), each presented  $4.67^\circ$  above or below central fixation. The participant was asked to indicate the location of the to-be-memorized stimulus in the current display by pressing the 'up' or 'down' key on the computer keyboard. The presentation of the memory-test array terminated once the participant made their response or after 3 s elapsed. The memory test was followed by a 1 s blank inter-trial interval.

On the other half of the trials, the experiment immediately proceeded to the 1 s inter-trial interval without providing a memory test. The reason for omitting the memory test is that it exposes participants to memory-matching shapes but not non-matching shapes, confounding our main manipulation with exposure effects in which greater levels of exposure to stimuli enhance ratings of those stimuli (Monahan et al., 2000). By including No-test trials, if such effects were detected, we could then rely solely on this subset of stimuli in which exposure in controlled across matching and non-matching distractors. We randomly intermixed an equal number of Memory-test trials with the No-test trials to ensure that participants were motivated to encode and maintain the to-be-memorized item throughout the retention interval and appearance of the distractor array on each trial.

Each block of 12 Memory trials was followed by a block of 12 Affective-evaluation trials. Each Affective-evaluation trial began with a 500 ms central fixation cross followed by the central presentation of a single to-be-rated stimulus, which remained visible until a rating response was made. Each to-be-rated item was either a previously-unseen Novel shape presented in the colour assigned to the novel condition, or a Memory-matching distractor or Non-matching distractor from the immediately-preceding block of memory trials. Each block of 12 Affective-evaluation trials contained an equal number of shapes (4) from each of these conditions. For the to-be-rated Memory-matching distractors and Non-matching distractors, half were from Memory-test trials and half were from No-test trials. Participants were asked to rate the perceived cheerfulness of each item by pressing one of the numbered keys at the top of a standard computer keyboard using a 5-



point scale (1-Not at all Cheerful, 5-Very Cheerful). Each rating response triggered a 1 s blank inter-trial interval.

Each participant completed 6 practice Memory trials and 6 practice Affective-evaluation trials to familiarize themselves with the task prior to the beginning of the experiment. The entire experimental session lasted approximately 30 min.

## 2.2. Results and discussion

Data from four participants were excluded because of low (< 50%) accuracy on Memory-task responses. The remaining 48 participants (Age:  $M=18.4$  years,  $SD=1.49$ ; 34 females, 47 right-handed) were correct on an average of 68.9% ( $SD=10.5$ ) of the Memory-task trials.

### 2.2.1. Stimulus ratings

Our primary contrasts of interest involved potential differences in affective stimulus ratings due to differences in level of prior inhibition or to differences in instruction-related evaluative associations. We were particularly interested in differences between ratings of novel items and prior memory-matching distractors (Evaluative coding: Memory-matching distractors should be significantly more positive; Inhibitory-devaluation: Memory-matching distractors should be significantly more negative) and between novel items and prior non-matching distractors (Evaluative coding: Non-matching distractors should be significantly more negative; Inhibitory-devaluation: Non-matching distractors may be more negative). However, because errors are also known to have negative affective consequences for associated items (e.g., Chetverikov, 2014; Chetverikov et al., 2015), we only considered ratings of prior distractors from memory-task trials in which a correct response was made. This resulted in the exclusion of 15.3% of the ratings of Memory-matching distractors and 15.8% of the ratings of Non-matching distractors from further analyses. The remaining affective ratings were averaged (means) and submitted to separate 2 (Memory-test status: Not tested or Tested)  $\times$  2 (Distractor status: Novel or Distractor) repeated-measures ANOVAs to test the predicted Novel-Distractor difference for Memory-matching distractors and that for Non-matching distractors.

As shown in Fig. 2, prior Memory-matching distractors received more negative ratings ( $M=2.81$ ,  $SD=.55$ ) than Novel items ( $M=3.11$ ,  $SD=.72$ ). This effect of Distractor status was significant  $F(1,47)=4.61$ ,  $p=.037$ ,  $\eta^2_{\text{partial}}=.089$ , and the lack of a significant Distractor-status  $\times$  Memory-test status interaction [ $F(1,47)=.51$ ,  $p=.513$ ,  $\eta^2_{\text{partial}}=.011$ ] confirmed that the magnitude of this effect was similar following No-test trials (Memory-match – Novel difference:  $-.34$ ) and Memory-test trials (Memory-match – Novel difference:  $-.28$ ). Importantly, the fact that the extra Memory-match stimulus exposure during the memory test did not subsequently lead to enhanced Memory-match distractor ratings when compared to that following No-test trials (i.e., if anything they

were nominally lower,  $M=2.77$  vs. 2.84), suggests that our concern about possible confounding effects of mere exposure may have been unfounded. Taken together, the results of this analysis are inconsistent with the evaluative-coding based prediction of more positive ratings for prior Memory-matching distractors than for Novel items, and are instead consistent with the devaluation-by-inhibition based prediction of more negative ratings for prior Memory-matching distractors than for Novel items.

Prior Non-matching distractors received nominally more negative ratings ( $M=3.03$ ,  $SD=.78$ ) than Novel items ( $M=3.11$ ,  $SD=.72$ ). However, this effect of Distractor status was not statistically significant  $F(1,47)=.21$ ,  $p=.651$ ,  $\eta^2_{\text{partial}}=.004$ , nor did it significantly interact with Memory-test status [ $F(1,47)=1.56$ ,  $p=.219$ ,  $\eta^2_{\text{partial}}=.032$ ], confirming that the magnitude of this effect was similarly small following No-test trials (Non-match – Novel difference:  $-.02$ ) and Memory-test trials (Non-match – Novel difference:  $-.14$ ). Taken together, the results of this analysis are inconsistent with the evaluative-coding based prediction of significantly more negative ratings for prior Non-matching distractors than for Novel items, and are instead consistent with the devaluation-by-inhibition based prediction of either no difference or slightly more negative ratings for prior Memory-matching distractors than for Novel items.

Finding significant devaluation of Memory-matching distractors, but not of Non-matching distractors, converges with the results of prior studies in which the stimuli with the greatest potential to interfere with task performance—those closer to the target (Martiny-Huenger et al., 2014; Raymond et al., 2005) or that have become increasingly salient through prior exposure (Frischen et al., 2012)—elicit the strongest negative reactions in subsequent evaluations. These prior results have been taken as evidence confirming the reactive nature of inhibition (e.g., Houghton and Tipper, 1994); the higher the potential for interference, the greater the amount of inhibition that needs to be applied in order to suppress further processing or overt responses. Evidence that memory-matching stimuli initially capture attention (e.g., Soto et al., 2005) could explain why their interference potential may become so great when they are irrelevant to the task at hand. And this, in turn, could explain why they subsequently elicit such strong suppression of any further processing as reflected by Sawaki and Luck (2011) observation of a significant Pd contralateral to memory-matching distractors (and not to non-matching distractors).

To more directly explore this potential link between the Pd-component index of stimulus inhibition and subsequent stimulus evaluation, we conducted Experiment 2 by adapting our approach in Experiment 1 to incorporate concurrent EEG and an assessment of the event-related potential waveform during Memory-trials and the correspondence to stimulus ratings during subsequent Affective-evaluation trials. The value of Experiment 2 is also underscored by the importance of replicating the pattern of stimulus rating results from Experiment 1 in providing sufficient confidence for any conclusions drawn from our findings concerning the competing predictions arising from the evaluative-coding and devaluation-by-inhibition accounts of distractor devaluation.

## 3. Experiment 2 – EEG/ERP

The use of EEG/ERP in Experiment 2 was chosen for its ability to provide the most stringent test to date of the devaluation-by-inhibition hypothesis. There is now strong consensus that the Pd component reflects the termination of attention to distracting stimuli and the active suppression of their perceptual processing (see Sawaki and Luck (2014) for review). Having a direct electro-cortical marker of distractor inhibition allows a direct assessment of the link between inhibition and stimulus devaluation. Here, in addition to a repeated assessment of the stimulus-rating predictions addressed in Experiment 1, we specifically examine whether trial-by-trial fluctuations in the magnitude of the Pd component for Memory-matching distractors are linked to the level

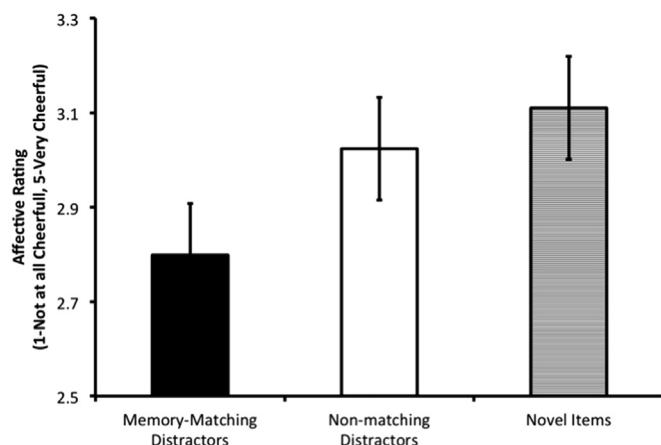


Fig. 2. Mean affective ratings of Memory-matching probes, Non-matching probes and Novel stimuli in Experiment 1. (Error Bars = SE) ( $n=48$ ).

of devaluation as reflected in their subsequent affective ratings. To do this we adopted Kiss et al.'s (2007, 2008) approach of sorting trials of the task in which inhibition is applied (here, the Memory-task) based on whether the putatively-inhibited stimuli (here, the Memory-matching distractors) subsequently received negative ratings or positive ratings. If the level of distractor inhibition is directly linked to level of subsequent distractor devaluation, then a larger Pd component should be found in response to Memory-matching distractors that were later rated as less cheerful as compared to trials leading to more positive ratings of these items. In contrast, the evaluative coding hypothesis might predict, if anything, that trial-by-trial fluctuations in the magnitude of the Pd component would be linked to *enhanced* ratings for Memory-matching distractors because of their similarity to the to-be-accepted/protected memorized stimuli.

### 3.1. Method

#### 3.1.1. Participants

A total of 24 undergraduate students from the University of Guelph (mean age 21.8 years,  $SD=3.23$ , 19 females, 22 right-handed) were recruited through online advertisements and received \$20 as compensation for participating. The size of this sample was chosen to be consistent with the samples of 12–16 participants used in previous demonstrations linking the Pd component to distractor inhibition (e.g., Hickey et al., 2009; Sawaki and Luck, 2011) and in prior neuroimaging (EEG/ERP and fMRI) investigations of the devaluation-by-inhibition hypothesis (Doallo et al., 2012; Kiss et al., 2007, 2008). The appropriateness of this sample size was confirmed using G\*Power (Faul et al., 2007), which indicated that 16 participants would be required to detect similar-sized ERP effects as in these studies ( $d=1.0$ ) with a power of .95 at an alpha of .05. All of the participants provided informed consent before participating and all had normal or corrected-to-normal vision.

#### 3.1.2. Apparatus and stimuli

The apparatus and stimuli in Experiment 2 were the same as in Experiment 1, with the following exceptions. The first of these concerns doubling the number of main experimental stimuli. This was required to accommodate the increased numbers of trials per condition required to ensure stable EEG signal averaging and, given the smaller sample of participants, to provide sufficient statistical power to test for the behavioural differences in stimulus ratings observed in Experiment 1. Therefore, 523 new abstract-shape stimuli were created using the same PsychoPy (Peirce, 2007) algorithm as used for the Experiment 1 stimuli. Combining these new items with the existing stimuli resulted in a stimulus set that consisted of 1046 unique abstract shapes for Experiment 2. Three different versions of this stimulus set were produced in the same way as for those from Experiment 1 by changing the uniform colour of each shape to green (RGB: 39, 103, 23), orange (RGB: 240, 111, 35), and purple (RGB: 130, 1, 200). Having each of the 1046 shapes in each of the three different colours resulted in a final stimulus set of 3138 images.

EEG was recorded using a BioSemi ActiveTwo system (Ag/AgCl electrodes) with 64 scalp electrodes in standard 10–20 placement, and 6 external electrodes: 2 placed on the mastoids to serve as a common reference for all other electrodes, 2 placed at the outer canthi of each eye for calculating the horizontal electrooculogram (HEOG), and 2 placed above/below the right eye for calculating the vertical electrooculogram (VEOG).

#### 3.1.3. Experimental procedure

All of the materials and procedures used in this experiment were approved by the Research Ethics Board at the University of Guelph. The procedure used in this experiment was the same as that used in Experiment 1 with the following exceptions.

The most significant modification was a change from the use of blocks of trials that each consisted of 12 Memory trials followed by 12

Affective-evaluation trials to the use blocks of trials in which each of the 12 Memory trials was immediately followed by the corresponding Affective-evaluation trial. Each to-be-rated item was either a previously-unseen Novel shape, a Memory-matching distractor or a Non-matching distractor from the immediately-preceding memory trial. Equal numbers of abstract shapes from each of these conditions were rated immediately after Memory-test and No-test trials. The number of blocks was also doubled from the eight used in Experiment 1 to a total of sixteen in Experiment 2. Changes were also made to address the data loss in Experiment 1 from having to exclude participants because of poor Memory-task performance. Specifically, we introduced a staircase procedure that was designed to maintain memory-task performance around 75% correct throughout the entire experimental session for each participant. This procedure utilized participants' accuracy across the four most recent Memory-test trials to determine whether, on the next Memory-test trial, to make it easier or harder to recognize the originally-memorized item within the memory-test array. Adjusting the difficulty of the memory test was accomplished by altering the difference between the orientation of the originally-memorized item and that of its rotated-version foil within a range of 20–60°. When recent accuracy dipped below 75%, then the difference in rotation was increased by 5° to make the memorized item easier to recognize. When recent accuracy increased above 75%, then the difference in rotation was decreased by 5° to make the memorized item harder to recognize. When accuracy was exactly at 75%, then the existing rotation difference was maintained.

Less significant changes from Experiment 1 to Experiment 2 included a 1 s presentation of the word “Memory” at display center at the beginning of each Memory trial, and a 1 s presentation of the word “Evaluation” at display center at the beginning of each Affective-evaluation trial. Participants were encouraged to blink when they saw each of these words so they would not need to during stimulus presentation, which was intended to help minimize the exclusion of EEG data from individual trials due to the presence of eye blinks. Other changes included adding jitter to the duration of the retention interval on each Memory trial by randomly varying the duration of the retention periods (700–900 ms) both preceding and following the 200 ms distractor-array. The timing of the sequence of events following the retention interval were also altered to minimize interruption by sensory and cognitive factors during measurement of the Pd component. Specifically, on the half of the memory trials in which memory maintenance was followed by a memory test, the memory-test array remained visible after a memory-test response was provided for a total of 3 s. On the No-test trials the retention interval was followed by a 3 s central fixation display. Participants were instructed to maintain fixation in such cases until the trial timed out.

Minor alterations to the Affective-evaluation trials also included the use of a 4-point rating scale (1-Not at all Cheerful, 4-Very Cheerful) and a change to make the presentation duration of the to-be-rated stimulus be the same throughout each rating-response period. Each to-be-rated stimulus was presented for 1 s followed by a 1 s blank display. Participants could respond at any time during this 2 s response period.

Each participant completed 18 practice Memory-task plus Affective-evaluation trials to familiarize themselves with the task prior to the beginning of the experiment. To familiarize participants with the procedure for Memory-task trials in which their memory would not be tested, the final 6 practice trials included 3 No-test memory trials. The entire experimental session lasted approximately 60 min.

#### 3.1.4. Neuroimaging analysis

All data analyses were conducted using EEGLab Toolbox (Delorme and Makeig, 2004) as well as ERPLab Toolbox (Lopez-Calderon and Luck, 2014). The data sets were re-referenced according to the average of the mastoids and resampled to 250 Hz. The EEG data was segmented into 600 ms epochs time-locked to the onset of the distractor array with a 100 ms baseline, and the mean across the epochs was used to

calculate averaged ERP waveforms.

The horizontal electrooculogram (HEOG) and the vertical electrooculogram (VEOG) were both low-pass filtered (30 Hz) and both used to remove eye-movement or blink artifacts from the data sets. Artifact rejection was completed using the moving window peak-to-peak threshold and step-like function rejection algorithms provided by ERPLAB. Epochs were removed from the data analysis if the HEOG exceeded  $-80$  or  $80 \mu\text{V}$ , if the voltage of the VEOG exceeded  $80 \mu\text{V}$ , and through visual inspection. An average of 3.5% (range: 0–12.5%) of trials were removed from each data set because of artifacts.

The average of two electrode pairs PO7/PO8 and P7/P8 were used to calculate the Pd component. First, an averaged waveform was calculated by subtracting the average voltage at electrode sites over the hemisphere ipsilateral to the Memory-matching distractor (e.g. PO7 and P7 for left-visual field Memory-matching distractors) from those over the contralateral hemisphere (e.g. PO8 and P8). The mean amplitude of this averaged difference waveform from 250 to 300 ms following the presentation of the distractor array was calculated and used as the amplitude of the Pd component.

### 3.2. Results and discussion

Every participant was correct on at least 57% of the Memory-test trials ( $M=71.0\%$ ,  $SD=9.4$ ). The staircase procedure utilized in Experiment 2 was therefore effective to the extent that no data was lost from having to exclude participants due to poor Memory-task accuracy.

#### 3.2.1. Stimulus rating results

As in Experiment 1, our primary behavioural contrasts of interest involved potential differences in affective stimulus ratings due to differences in level of prior inhibition or to differences in instruction-related evaluative associations. To avoid the potential confounding effects of negative affect for error-associated items (e.g., Chetverikov et al., 2015), we excluded ratings of stimuli that immediately followed Memory-test trials in which there was a failure to provide a correct response. This resulted in the exclusion of 14.5% of the ratings of Memory-matching distractors, 14.8% of the ratings of Non-matching distractors, and 15.5% of the ratings of Novel items from further analyses. The remaining affective ratings were averaged (means) and submitted to separate 2 (Memory-test status: Not tested or Tested)  $\times$  2 (Distractor status: Novel or Distractor) repeated-measures ANOVAs to assess the predicted Novel-Distractor difference for Memory-matching distractors (Evaluative coding: Memory-matching distractors should be significantly more positive vs. Inhibitory-devaluation: Memory-matching distractors should be significantly more negative) and for Non-matching distractors (Evaluative coding: Non-matching distractors should be significantly more negative vs. Inhibitory-devaluation: Non-matching distractors may be slightly more negative).

As shown in Fig. 3, prior Memory-matching distractors received more negative ratings ( $M=2.50$ ,  $SD=.29$ ) than Novel items ( $M=2.81$ ,  $SD=.50$ ). This effect of Distractor status was significant  $F(1,23)=6.35$ ,  $p=.019$ ,  $\eta^2_{\text{partial}}=.216$ , and the lack of a significant Distractor-status  $\times$  Memory-test status interaction [ $F(1,23)=1.61$ ,  $p=.218$ ,  $\eta^2_{\text{partial}}=.065$ ] confirmed that the magnitude of this effect was similar following No-test trials (Memory-match – Novel difference:  $-.28$ ) and Memory-test trials (Memory-match – Novel difference:  $-.35$ ).

Prior Non-matching distractors received nominally more negative ratings ( $M=2.68$ ,  $SD=.47$ ) than Novel items ( $M=2.81$ ,  $SD=.50$ ). However, this effect of Distractor status was not statistically significant  $F(1,23)=.68$ ,  $p=.419$ ,  $\eta^2_{\text{partial}}=.029$ , nor did it significantly interact with Memory-test status [ $F(1,23)=1.90$ ,  $p=.182$ ,  $\eta^2_{\text{partial}}=.076$ ], confirming that the magnitude of this non-significant effect did not reliably differ following No-test trials (Non-match – Novel difference:  $-.08$ ) and Memory-test trials (Non-match – Novel difference:  $-.19$ ).

The results of these analyses closely replicate the results of

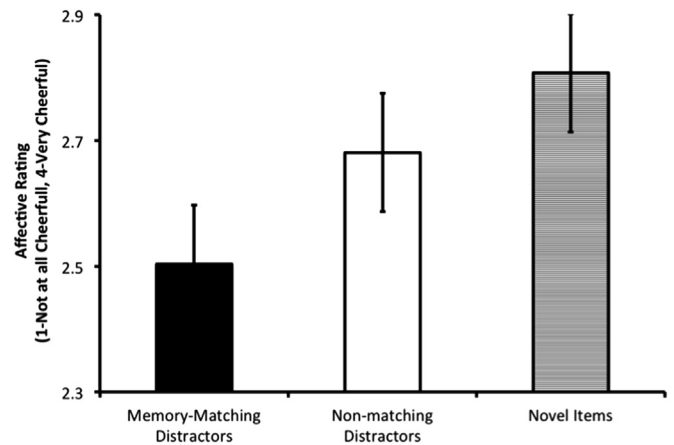


Fig. 3. Mean affective ratings of Memory-matching distractors, Non-matching distractors and Novel stimuli in Experiment 2. (Error Bars=SE) ( $n=24$ ).

Experiment 1. Taken together, these findings are one again inconsistent with the evaluative-coding based predictions of relatively enhanced ratings for prior Memory-matching distractors and relatively diminished ratings for prior Non-match distractors, but are consistent with the devaluation-by-inhibition based predictions of significant devaluation of prior Memory-matching distractors and less or absent devaluation of Non-matching distractors.

It is noteworthy that none of the stimulus-rating effects due to potential differences in level of prior inhibition or instruction-related evaluative associations significantly interacted with the effects of including or omitting the memory test in either experiment. The consistent similarity in the pattern of stimulus ratings following No-test and Memory-test trials across both experiments and the absence of any other evidence of potential confounding effects of mere exposure due to the memory tests give us confidence in the decision to collapse across Test-status for the EEG/ERP analyses and the assessment of links between the Pd component and magnitude of distractor devaluation. The primary advantage of doing this, of course, is the extra statistical power that comes with increasing trials per condition.

#### 3.2.2. EEG/ERP results

Our assessment of the Pd component was based on the analysis used by Sawaki and Luck (2011). The mean difference in amplitude between electrode sites contralateral to the Memory-matching distractor and those ipsilateral to the Memory-matching distractors from 250 to 300 ms post distractor-array was  $1.04 \mu\text{V}$  (Fig. 4). The results of a

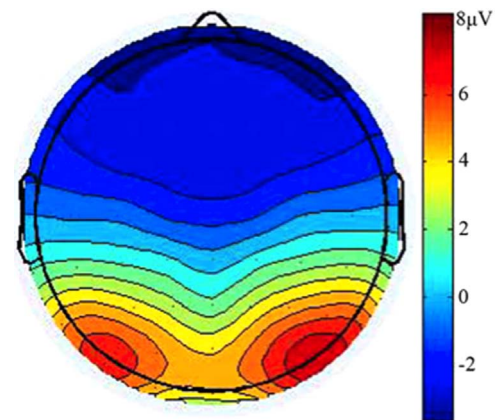


Fig. 4. Topographic scalp map of mean activation during the 250–300 ms post-distractor-array time interval used to calculate the Pd component. Data were rearranged so that the left side of the map represents the electrodes ipsilateral to the Memory-matching distractor while the right side represents the electrodes contralateral to the Memory-matching distractor. ( $n=24$ ).



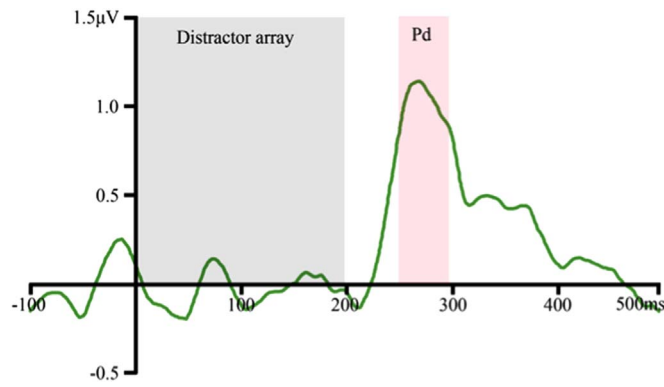


Fig. 5. Averaged ERP waveform of the difference in voltage between electrode sites contralateral to the Memory-matching distractor vs. those ipsilateral to the Memory-matching distractor (Electrodes: PO7/PO8 and P7/P8). For all statistical analyses, recorded voltages were not filtered; however, for illustrative purposes, voltage data in Figures were high-pass (0.1 Hz) and low-pass (30 Hz) filtered. ( $n = 24$ ).

one-sample  $t$ -test showed that this mean amplitude of the difference between the voltage contralateral and ipsilateral to the Memory-matching distractors was significantly greater than zero,  $t(23) = 4.83$ ,  $p < .001$  (Fig. 5). This pattern closely resembles the results of Sawaki and Luck (2011). Moreover, the observation of a Pd contralateral to memory-matching distractors suggests that the level of suppression required to prevent the memory-matching distractor from interfering with the contents of working memory was greater than that required to prevent interference from the non-matching distractor (Hickey et al., 2009). The demonstration that the condition in Experiment 2 that elicited a direct electro-cortical marker of distractor inhibition was the only condition that also elicited significant distractor devaluation provides strong support for the devaluation-by-inhibition hypothesis. It is less clear how such a finding can be explained from an evaluative-coding perspective.

If active suppression represented by the Pd component is indeed linked to the affective devaluation of Memory-matching distractors, then fluctuations in the magnitude of the Pd component should also be linked to the magnitude of distractor devaluation. To assess this potential link, we followed the approach developed by Kiss et al. (2007, 2008) for investigating how ERPs obtained during a task involving stimulus/response inhibition can be predictive of subsequent affective ratings of the stimuli in that task. First, we limited our analysis to trials in which the Memory-matching distractor encountered in the Memory task was the to-be-rated item in the subsequent Affective-evaluation task. Then we sorted these trials into two categories based on whether the Memory-matching distractor received a relatively negative rating (i.e., Low-rating: a 1 or 2) or a relatively positive rating (i.e., High-rating: a 3 or 4). Separate averaged waveforms were calculated using these different groups of trials for each participant, from which we were able to calculate the mean Pd amplitude. The amplitudes of these High-rating and Low-rating Pd components were then compared using a paired-samples  $t$ -test. As shown in Fig. 6, the magnitude of the Pd component on Low-rating trials was significantly greater than that on High-rating trials  $t(23) = 2.32$ ,  $p < .05$ . In other words, the level of inhibition applied to Memory-matching distractors that subsequently received the most negative ratings was significantly greater than the level of inhibition applied to Memory-matching distractors that subsequently received more positive ratings. This strongly suggests that trial-by-trial fluctuations in the level of inhibition applied to Memory-matching distractors is linked to the level of their subsequent affective devaluation.

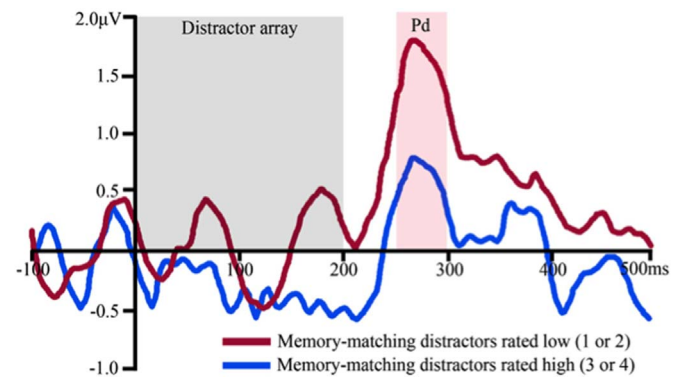


Fig. 6. Averaged ERP waveforms of the differences in voltage between electrode sites contralateral to the Memory-matching distractor vs. those ipsilateral to the Memory-matching distractor, separated by subsequent affective rating (Low, High) of the Memory-matching distractor (Electrodes: PO7/PO8 and P7/P8). ( $n = 24$ ).

#### 4. General discussion

Previously ignored stimuli typically receive more negative affective evaluations than novel stimuli or the targets of attention/response (e.g., Raymond et al., 2003), but it remains unclear exactly how this occurs. Does inhibition applied to prevent distractor interference have negative affective consequences for associated stimuli? Or are prior distractors only rated more negatively because of evaluative codes— affective connotations conveyed solely by experimental instructions—linking negative affect to categories of stimuli that are to be ignored, avoided, or otherwise rejected? We directly compared and contrasted competing predictions arising from the devaluation-by-inhibition and evaluative-coding accounts of the distractor devaluation effect in two experiments. The results of both experiments provide strong support for the hypothesis that distractor devaluation is directly linked to stimulus inhibition and is not solely due to the influence of instruction-based evaluative codes.

Participants in both experiments rated the cheerfulness of abstract shapes that were either previously-unseen novel items or had previously appeared as distractors during the retention interval of a working memory task. Those distractors that matched the colour of a stimulus being held in memory received significantly more negative ratings than the novel stimuli in both experiments, whereas ratings of non-matching distractors did not reliably differ from those of novels. Because all of the to-be-memorized stimuli shared the same defining feature as all of the memory-matching distractors, the devaluation of memory-matching distractors stands in stark contrast to the evaluative-coding based prediction that items from that stimulus category should typically receive enhanced ratings because of the instruction-based positive affective connotations that such items are the sort of things to be accepted and protected (Dittrich and Klauer, 2012). Finding no difference between ratings of novel items and non-matching distractors in both experiments is also inconsistent with the other evaluative-coding based prediction; namely, that non-matching distractors should become significantly devalued from sharing a defining feature and any accompanying negative affective associations with visual stimuli in the initial memory arrays that participants were cued to reject as do-not-memorize items.

In contrast, the stimulus rating results of Experiments 1 and 2 confirmed the predictions arising from the devaluation-by-inhibition hypothesis. Finding significant devaluation of memory-matching distractors, for example, resonates well with prior evidence that memory-matching distractors are subjected to significant levels of attentional inhibition (Sawaki and Luck, 2011) because of their relatively greater task-interference potential due to attention being initially drawn to sensory information that matches the contents of working memory (Soto et al., 2005). Our replication in Experiment 2 of Sawaki and



Luck's (2011) demonstration that the Pd component—an established electro-cortical marker of the termination of attention to a stimulus and its subsequent inhibition (Sawaki and Luck, 2014)—occurs in response to the memory-matching distractors (and not to non-matching distractors) confirms that inhibition is indeed applied to memory-matching distractors. Finding significant devaluation only of stimuli from the condition in which there was direct evidence of stimulus inhibition confirms that prior inhibition is directly associated with changes in stimulus affect. So closely replicating this pattern of stimulus rating results across two independent samples of participants, despite slight changes in the experimental procedure and context (e.g., wearing an EEG cap), speaks to the reliability of our findings.

We also demonstrated in Experiment 2 that the level of distractor inhibition indicated by Pd magnitude was significantly greater during memory maintenance for memory-matching distractors that subsequently received the most negative ratings than for those receiving more positive ratings. This provides the clearest neural evidence to date that distractor inhibition is associated with distractor devaluation. There are many reasons why participants vary in their ability to exert attentional control from one point to another within a given task—fatigue, mind-wandering, etc.—and these momentary changes can impact participants' ability to resist distraction (e.g., Leber, 2010). The possible impact of such trial-by-trial fluctuations in Pd magnitude on distractor evaluations, while straightforward to understand from a devaluation-by-inhibition account, is less clear from an evaluative-coding perspective. But, to the extent that attentional mechanisms aid discrimination of items from different stimulus categories, as suggested by Dittrich and Klauer (2012), then the evaluative coding hypothesis might predict, if anything, that trial-by-trial fluctuations in the magnitude of the Pd component should have been linked to enhanced ratings for distractors matching the to-be-accepted/protected memorized stimuli. There was no evidence of this.

Our failure to confirm any of the evaluative-coding based predictions converges with the other results to show that the effects of selective-attention tasks in determining affective responses to the stimuli encountered in such tasks can not all be explained solely as an artefact of the attention-task instructions (see also Gollwitzer et al. (2014)). Martiny-Huenger et al. (2014), for example, found devaluation of prior distractors from a focused-attention task that used a single stimulus set comprising unique but otherwise homogenous Chinese characters. Participants in their studies were required to make judgments about whether a central target was symmetrical while being flanked by task-irrelevant distractors that were themselves either symmetrical or asymmetrical. Thus individual items could not be simply labeled as good/to-be-accepted and bad/to-be-rejected based on task instructions because there were no distinct categories or defining stimulus features that delineated targets from distractors. Moreover, in line with our finding of significant devaluation only for memory-matching distractors and Raymond et al.'s (2005) finding of greater devaluation for distractors in close proximity to a target of attention, they too found significant devaluation only for distractors with high potential for task interference, including those associated with an incompatible response and those in close proximity to the target. This means that—even if specific evaluative codes could have been differentially linked to targets and distractors—the same affective associations would have applied to all distractors, making it difficult to explain why only some were devalued while others were not. In contrast, such differences in level of stimulus devaluation are easily explained by understanding that some stimulus/response representations require greater levels of inhibition to avoid task interference, and that the distractors that receive the most inhibition are also those that later show the greatest levels of distractor inhibition.

Our demonstration that the magnitude of the Pd component—the level of distractor suppression—is linked to distractor ratings obtained later on, converges nicely with Kiss et al.'s (2007) demonstration that differences in the emergence of the N2pc component—widely consid-

ered to provide an index of the efficiency with which attention could be selectively allocated to a target stimulus in the presence of visual distractors (e.g., Eimer, 1996; Luck and Hillyard, 1994)—are also linked to the magnitude of distractor devaluation. Although it is becoming increasingly clear that the N2pc is not itself an index of inhibition, per se, but is more directly related to processing of target information (Mazza et al., 2009), to the extent that target selection is aided by effective distractor suppression it makes sense that the N2pc component might typically emerge under conditions involving distractor inhibition. The added benefit of now being able to measure the Pd component as a more direct neural index of distractor inhibition is that it enables a more direct test of the devaluation-by-inhibition hypothesis. And the results of our assessment of the link between the Pd marker of stimulus inhibition and subsequent affective evaluations clearly suggest that inhibition is indeed associated with changes in stimulus affect.

While there is growing evidence that the Pd component provides an index of the direct suppression of cortical representations of distracting stimuli, the exact neural source of the top-down inhibitory signals that target visual cortex has yet to be determined. Middle frontal gyrus may be a candidate, however, because trial-by-trial fluctuations in middle frontal gyrus activity and Pd magnitude both predict the ability to resist the impact of visual distractors (Leber, 2010; Sawaki et al., 2012). Future research might use fMRI with our version of Sawaki and Luck's (2011) experimental procedure to establish whether middle frontal gyrus activation level during the distractor-array presentation is linked to subsequent stimulus ratings, just as we found the Pd magnitude to be. Such a finding would resonate well with the fMRI results of Doallo et al.'s (2012) study that combined a response-inhibition task with an affective-evaluation task to investigate the link between inhibition and stimulus devaluation. They found that activity in middle frontal gyrus was greatest during periods that required inhibition for successful task performance (i.e., during No-go trials in a Go/No-go response-inhibition task), and that the magnitude of this inhibition-related activity was linked to the subsequent level of affective devaluation of the associated stimuli.

Moreover, results such as Doallo et al.'s (2012) provide important clues about how inhibitory signals from top-down control areas might lead to changes in stimulus evaluations. These clues include evidence that inhibition-related activity in lateral prefrontal cortex couples with activity in key emotion-related regions, including orbitofrontal cortex and amygdala (also see Berkman et al. (2009)). The findings of Doallo et al. suggest further that the pattern of activity in emotion-related regions produced through their coupling with inhibition-related regions at the time inhibition is applied can be reinstated at a later time in the context of an affective evaluation task. Importantly, the magnitude of these changes in orbital-frontal cortex and amygdala activity for previously inhibited (No-go) items were associated with the magnitude of stimulus devaluation as shown behaviourally by subjective affective ratings. A formal neural-network model of distractor devaluation (Fragopanagos et al., 2009) suggests that similar connectivity between prefrontal top-down control areas and fronto-limbic emotion areas may be involved in determining the negative affective consequences of attentional inhibition for visual distractors. Thus, while our present results have provided the most direct neural evidence to date linking attentional inhibition to stimulus devaluation, future neuroimaging investigations will be useful for confirming the specific regions and sequence of neurocognitive events through which inhibitory mechanisms of attention alter the coding and representation of stimulus value.

Future research is also needed to verify that inhibition itself is what causes a change in stimulus value (Raymond, 2009), rather than another process that consistently accompanies inhibition. The primary functional value of inhibition, for example, lies in its critical role as a mechanism for resolving conflict and potential interference from competing signals in a variety of cognitive and neural operations (see Munakata et al. (2011) for review). Indeed, our results—a Pd compo-

nent and significant devaluation of the distractors made most salient through their resemblance to memory targets—strongly supports the view that inhibition is a reactive process that is applied to distracting or otherwise inappropriate stimulus/response representations in direct proportion to the level of potential interference that might otherwise occur (Houghton and Tipper, 1994). However, just as there is growing evidence for a link between inhibition and stimulus devaluation, there is also growing evidence of a link between stimulus/response conflict and negative affect (Dreisbach and Fischer, 2012; Fritz and Dreisbach, 2013, 2015). Thus, whereas some theoretical accounts have emphasized the inextricable link between conflict and inhibition in determining stimulus-linked affective response (e.g., Ferrey et al., 2015), others have suggested that it is solely conflict, and not inhibition, that is responsible for stimulus devaluation effects (Chetverikov and Kristjánsson, 2016). Finding a way to disentangle the relative influence of different cognitive mechanisms, such as these, in determining affective responses to stimuli encountered in complex visual tasks presents a specific challenge for ongoing research. The importance of such work, however, is underscored by the growing range of stimuli, tasks, emotional judgments, choices and behaviour that are impacted by these critical links between cognition and emotion.

### Author note

This research was supported by the Natural Science and Engineering Research Council of Canada, the Canada Foundation for Innovation, and the Ontario Ministry of Research and Innovation. Correspondence concerning this article should be addressed to Mark J. Fenske, Ph.D., Department of Psychology, University of Guelph, Guelph, Ontario, Canada N1G 2W1. Email: [mfenske@uoguelph.ca](mailto:mfenske@uoguelph.ca).

### References

- Berkman, E.T., Burklund, L., Lieberman, M.D., 2009. Inhibitory spillover: intentional motor inhibition produces incidental limbic inhibition via right inferior frontal cortex. *NeuroImage* 47 (2), 705–712. <http://dx.doi.org/10.1016/j.neuroimage.2009.04.084>.
- Chen, M., Bargh, J.A., 1999. Consequences of automatic evaluation: immediate behavioral predisposition to approach or avoid the stimulus. *Personal. Soc. Psychol. Bull.* 25, 215–224.
- Chetverikov, A., 2014. Warmth of familiarity and chill of error. *Cogn. Emot.* 28 (3), 385–415.
- Chetverikov, A., Jóhannesson, O.I., Kristjánsson, A., 2015. Blaming the victims of your own mistakes: how visual search accuracy influences evaluation of stimuli. *Cogn. Emot.* 29 (6), 1091–1106.
- Chetverikov, A., Kristjánsson, Á., 2016. On the joys of perceiving: affect as feedback for perceptual predictions. *Acta Psychol.* 169, 1–10. <http://dx.doi.org/10.1016/j.actpsy.2016.05.005>.
- Cutzu, F., Tsotsos, J.K., 2003. The selective tuning model of attention: psychophysical evidence for a suppressive annulus around an attended item. *Vis. Res.* 43 (2), 205–219. [http://dx.doi.org/10.1016/S0042-6989\(02\)00491-1](http://dx.doi.org/10.1016/S0042-6989(02)00491-1).
- De Vito, D., Fenske, M.J., 2017. Suppressing memories of words and familiar objects results in their affective devaluation: evidence from Think/No-Think tasks. *Cognition* 162, 1–11. <http://dx.doi.org/10.1016/j.cognition.2017.01.020>.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134 (1), 9–21. <http://dx.doi.org/10.1016/j.jneumeth.2003.10.009>.
- Dittrich, K., Klauer, K.C., 2012. Does ignoring lead to worse evaluations? A new explanation of the stimulus devaluation effect. *Cogn. Emot.* 26 (2), 193–208. <http://dx.doi.org/10.1080/02699931.2011.570313>.
- Doallo, S., Raymond, J.E., Shapiro, K.L., Kiss, M., Eimer, M., Nobre, A.C., 2012. Response inhibition results in the emotional devaluation of faces: neural correlates as revealed by fMRI. *Soc. Cogn. Affect. Neurosci.* 7 (6), 649–659. <http://dx.doi.org/10.1093/scan/nsr031>.
- Dreisbach, G., Fischer, R., 2012. Conflicts as aversive signals. *Brain Cogn.* 78 (2), 94–98.
- Duff, B.R.L., Faber, R.J., 2011. Missing the mark. *J. Advert.* 40 (2), 51–62.
- Eder, A.B., Rothermund, K., 2008. When do motor behaviors (mis)match affective stimuli? An evaluative coding view of approach and avoidance reactions. *J. Exp. Psychol. Gen.* 137 (2), 262–281. <http://dx.doi.org/10.1037/0096-3445.137.2.262>.
- Eimer, M., 1996. The N2pc component as an indicator of attentional selectivity. *Electroencephalogr. Clin. Neurophysiol.* 99 (3), 225–234. [http://dx.doi.org/10.1016/S0921-884X\(96\)95711-2](http://dx.doi.org/10.1016/S0921-884X(96)95711-2).
- Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A., 2007. GPOWER: a general power analysis program. *Behav. Res. Methods* 39 (2), 175–191. <http://dx.doi.org/10.3758/BF03193146>.
- Fenske, M.J., Raymond, J.E., 2006. Affective influences of selective attention. *Curr. Dir. Psychol. Sci.* 15 (6), 312–316. <http://dx.doi.org/10.1111/j.1467-8721.2006.00459.x>.
- Fenske, M.J., Raymond, J.E., Kessler, K., Westoby, N., Tipper, S.P., 2005. Attentional inhibition has social-emotional consequences for unfamiliar faces. *Psychol. Sci.* 16 (10), 753–758. <http://dx.doi.org/10.1111/j.1467-9280.2005.01609.x>.
- Fenske, M.J., Raymond, J.E., Kunar, M.A., 2004. The affective consequences of visual attention in preview search. *Psychon. Bull. Rev.* 11 (6), 1055–1061. <http://dx.doi.org/10.3758/BF03196736>.
- Ferrey, A.E., Burleigh, T.J., Fenske, M.J., 2015. Stimulus-category competition, inhibition, and affective devaluation: a novel account of the uncanny valley. *Front. Psychol.* 6, 1–15. <http://dx.doi.org/10.3389/fpsyg.2015.00249>.
- Ferrey, A.E., Frischen, A., Fenske, M.J., 2012. Hot or not: response inhibition reduces the hedonic value and motivational incentive of sexual stimuli. *Front. Psychol.: Emot. Sci.* 3, 575. <http://dx.doi.org/10.3389/fpsyg.2012.00575>.
- Fragopanagos, N., Cristescu, T., Goolsby, B.A., Kiss, M., Eimer, M., Nobre, A.C., Taylor, J.G., 2009. Modelling distractor devaluation (DD) and its neurophysiological correlates. *Neuropsychologia* 47 (12), 2354–2366. <http://dx.doi.org/10.1016/j.neuropsychologia.2008.09.018>.
- Frischen, A., Ferrey, A.E., Burt, D.H.R., Pistchik, M., Fenske, M.J., 2012. The affective consequences of cognitive inhibition: devaluation or neutralization? *J. Exp. Psychol. Hum. Percept. Perform.* 38 (1), 169–179. <http://dx.doi.org/10.1037/a0025981>.
- Fritz, J., Dreisbach, G., 2013. Conflicts as aversive signals: conflict priming increases negative judgments for neutral stimuli. *Cogn. Affect. Behav. Neurosci.* 13, 311–317.
- Fritz, J., Dreisbach, G., 2015. The time course of the aversive conflict signal. *Exp. Psychol.* 62 (1), 30–39. <http://dx.doi.org/10.1027/1618-3169/a000271>.
- Gollwitzer, P.M., Martiny-Huenger, T., Oettingen, G., 2014. Affective consequences of intentional action control. In: *Advances in Motivation Science*, Vol. 1, pp. 49–83. <http://doi.org/10.1016/bs.adms.2014.08.002>.
- Goolsby, B.A., Shapiro, K.L., Raymond, J.E., 2009a. Distractor devaluation requires visual working memory. *Psychon. Bull. Rev.* 16 (1), 133–138. <http://dx.doi.org/10.3758/PBR.16.1.133>.
- Goolsby, B.A., Shapiro, K., Silvert, L., Kiss, M., Fragopanagos, N., Taylor, J., Raymond, J., 2009b. Feature-based inhibition underlies the affective consequences of attention. *Vis. Cogn.* 17 (4), 500–530. <http://dx.doi.org/10.1080/13506280801904095>.
- Griffiths, O., Mitchell, C.J., 2008. Negative priming reduces affective ratings. *Cogn. Emot.* 22 (6), 1119–1129. <http://dx.doi.org/10.1080/02699930701664930>.
- Hickey, C., Di Lollo, V., McDonald, J.J., 2009. Electrophysiological indices of target and distractor processing in visual search. *J. Cogn. Neurosci.* 21 (4), 760–775. <http://dx.doi.org/10.1162/jocn.2009.21039>.
- 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. *Proc. Natl. Acad. Sci.* 103 (4), 1053–1058. <http://dx.doi.org/10.1073/pnas.0507746103>.
- Houghton, G., Tipper, S.P., 1994. A model of inhibitory mechanisms in selective attention. In: Dagenbach, D., Carr, T.H. (Eds.), *Inhibitory Processes in Attention, Memory, and Language*. Academic Press, Inc, San Diego, CA.
- Janczyk, M., Wühr, P., 2012. Inhibition does not always cause emotional devaluation. *Exp. Psychol.* 59 (6), 372–378.
- Kihara, K., Yagi, Y., Takeda, Y., Kawahara, J.I., 2011. Distractor devaluation effect in the attentional blink: direct evidence for distractor inhibition. *J. Exp. Psychol. Hum. Percept. Perform.* 37 (1), 168–179. <http://dx.doi.org/10.1037/a0019948>.
- Kiss, M., Goolsby, B.A., Raymond, J.E., Shapiro, K.L., Silvert, L., Nobre, A.C., Eimer, M., 2007. Efficient attentional selection predicts distractor devaluation: event-related potential evidence for a direct link between attention and emotion. *J. Cogn. Neurosci.* 19 (8), 1316–1322. <http://dx.doi.org/10.1162/jocn.2007.19.8.1316>.
- Kiss, M., Raymond, J.E., Westoby, N., Nobre, A.C., Eimer, M., 2008. Response inhibition is linked to emotional devaluation: behavioural and electrophysiological evidence. *Front. Hum. Neurosci.* 2, 13. <http://dx.doi.org/10.3389/fpsyg.2008.09.013.2008>.
- Leber, A.B., 2010. Neural predictors of within-subject fluctuations in attentional control. *J. Neurosci.* 30 (34), 11458–11465. <http://dx.doi.org/10.1523/JNEUROSCI.0809-10.2010>.
- Lopez-Calderon, J., Luck, S.J., 2014. ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Front. Hum. Neurosci.* 8, 1–14. <http://dx.doi.org/10.3389/fnhum.2014.00213>.
- Luck, S.J., Hillyard, S.A., 1994. Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*. <http://doi.org/10.1111/j.1469-8986.1994.tb02218.x>.
- Martiny-Huenger, T., Gollwitzer, P., Oettingen, G., 2014. Distractor devaluation in a flanker task: object-specific effects without distractor recognition memory. *J. Exp. Psychol.: Hum. Percept. Perform.* 40 (2), 613–625.
- Mazza, V., Turatto, M., Caramazza, A., 2009. Attention selection, distractor suppression and N2pc. *Cortex* 45 (7), 879–890. <http://dx.doi.org/10.1016/j.cortex.2008.10.009>.
- Monahan, J.L., Murphy, S.T., Zajonc, R.B., 2000. Subliminal mere exposure: specific, general, and diffuse effects. *Psychol. Sci.* 11 (6), 462–466. <http://dx.doi.org/10.1111/1467-9280.00289>.
- Mounts, J.R., 2000. Evidence for suppressive mechanisms in attentional selection: feature singletons produce inhibitory surrounds. *Percept. Psychophys.* 62 (5), 969–983. <http://dx.doi.org/10.3758/BF03212082>.
- Munakata, Y., Herd, S.A., Chatham, C.H., Depue, B.E., Banich, M.T., O'Reilly, R.C., 2011. A unified framework for inhibitory control. *Trends Cogn. Sci.* 15, 453–459. <http://dx.doi.org/10.1016/j.tics.2011.07.011>.
- Peirce, J.W., 2007. PsychoPy-Psychophysics software in Python. *J. Neurosci. Methods* 162 (1–2), 8–13. <http://dx.doi.org/10.1016/j.jneumeth.2006.11.017>.
- Raymond, J.E., 2009. Interactions of attention, emotion and motivation. *Prog. Brain Res.* 176, 293–308. [http://dx.doi.org/10.1016/S0079-6123\(09\)17617-3](http://dx.doi.org/10.1016/S0079-6123(09)17617-3).

- Raymond, J.E., Fenske, M.J., Tavassoli, N.T., 2003. Selective attention determines emotional responses to novel visual stimuli. *Psychol. Sci.* 14 (6), 537–542. <http://dx.doi.org/10.1046/j.0956-7976.2003.psci.1462.x>.
- Raymond, J.E., Fenske, M.J., Westoby, N., 2005. Emotional devaluation of distracting patterns and faces: a consequence of attentional inhibition during visual search? *J. Exp. Psychol. Hum. Percept. Perform.* 31 (6), 1404–1415. <http://dx.doi.org/10.1037/0096-1523.31.6.1404>.
- Sawaki, R., Geng, J.J., Luck, S.J., 2012. A common neural mechanism for preventing and terminating the allocation of attention. *J. Neurosci.* 32 (31), 10725–10736. <http://dx.doi.org/10.1523/JNEUROSCI.1864-12.2012>.
- Sawaki, R., Luck, S.J., 2011. Active suppression of distractors that match the contents of visual working memory. *Vis. Cogn.* 19 (7), 956–972. <http://dx.doi.org/10.1080/13506285.2011.603709>.
- Sawaki, R., Luck, S.J., 2014. How the brain prevents and terminates shifts of attention. In: Mangun, G.R. (Ed.), *Cognitive Electrophysiology of Attention: Signals of the Mind*. Elsevier, New York, NY, pp. 16–29. <http://dx.doi.org/10.1016/B978-0-12-398451-7.00002-6>.
- Soto, D., Heinke, D., Humphreys, G.W., Blanco, M.J., 2005. Early, involuntary top-down guidance of attention from working memory. *J. Exp. Psychol.: Hum. Percept. Perform.* 31 (2), 248–261. <http://dx.doi.org/10.1037/0096-1523.31.2.248>.
- Veling, H., Holland, R.W., van Knippenberg, A., 2007. Devaluation of distracting stimuli. *Cogn. Emot.* 21 (2), 442–448. <http://dx.doi.org/10.1080/02699930600843007>.
- Vivas, A.B., Marful, A., Panagiotidou, D., Bajo, T., 2016. Instruction to forget lead to emotional devaluation. *Cognition* 150, 85–91.
- Wessel, J., O'Doherty, J., 2014. Stimulus devaluation induced by stopping action. *J. Exp. Psychol.: Gen.* 143 (6), 2316–2329.