A Neural Mechanism for Surprise-related Interruptions of Visuospatial Working Memory

Jan R. Wessel

Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA 52242, USA and Department of Neurology, University of Iowa Hospitals and Clinics, Iowa City, IA 52245, USA

Address correspondence to Jan R. Wessel, University of Iowa Hospitals and Clinics, Department of Neurology, 200 Hawkins Drive, 1042 ML, Iowa City, IA 52242, USA. Email: jan-wessel@uiowa.edu

Abstract

Surprising perceptual events recruit a fronto-basal ganglia mechanism for inhibition, which suppresses motor activity following surprise. A recent study found that this inhibitory mechanism also disrupts the maintenance of verbal working memory (WM) after surprising tones. However, it is unclear whether this same mechanism also relates to surprise-related interruptions of non-verbal WM. We tested this hypothesis using a change-detection task, in which surprising tones impaired visuospatial WM. Participants also performed a stop-signal task (SST). We used independent component analysis and single-trial scalp-electroencephalogram to test whether the same inhibitory mechanism that reflects motor inhibition in the SST relates to surprise-related visuospatial WM decrements, as was the case for verbal WM. As expected, surprising tones elicited activity of the inhibitory mechanism, and this activity correlated strongly with the trial-by-trial level of surprise. However, unlike for verbal WM, the activity of this mechanism was unrelated to visuospatial WM accuracy. Instead, inhibition-independent activity that immediately succeeded the inhibitory mechanism was increased when visuospatial WM was disrupted. This shows that surprise-related interruptions of visuospatial WM are not effected by the same inhibitory mechanism that interrupts verbal WM, and instead provides evidence for a 2-stage model of distraction.

Key words: change detection, motor inhibition, stop-signal task, surprise, working memory

Introduction

Surprising events are distracting—that is, they can disrupt ongoing cognitive processes such as working memory (WM). For example, when holding a phone number in mind, an unexpected noise can lead to forgetting the number. Recent research has shown that surprise-related interruptions of such verbal types of WM are mediated by a fronto-basal ganglia inhibitory mechanism, which is otherwise recruited to stop action (Wessel et al. 2016).

This fronto-basal ganglia inhibitory mechanism is best known from studies of motor inhibition (see Aron et al. 2014; Kenemans 2015 for reviews), which use the stop-signal task (SST, Logan and Cowan 1984). In the SST, the inhibitory mechanism is recruited to rapidly cancel an impending action when an external signal explicitly instructs the participant to stop.

However, recent studies have shown that this mechanism is also recruited following surprising perceptual events, such as unexpected tones. Specifically, one study showed that the activity of this mechanism is related to motor inhibition after surprising events (Wessel and Aron 2013), and another study showed that it also mediates interruptions of verbal WM (Wessel et al. 2016). In the latter study, strings of letters were maintained in WM and sine-wave tones were played immediately before the contents of WM were probed. On some trials, the tones were replaced by surprising birdsong segments. Independent component analysis (ICA, Jutten and Herault 1991) of scalp electroencephalogram (EEG) data showed that surprising tones elicited activity within the same independent component that indexes motor inhibition in the SST (from here onwards referred to as the “motor inhibition independent...
component”, MI-IC). Critically, the degree of this inhibitory activity mediated a surprise-related decrement of WM accuracy: tones that were more surprising (as quantified trial-by-trial using a Bayesian model) were more likely to be followed by mistakes on the WM probe; and greater surprise-related activity within MI-IC explained this adverse influence of surprise on WM. This was tested using a single-trial general linear model, which showed a significant correlation between MI-IC EEG activity on the one hand, and an interaction term that quantified the relationship between the degree of surprise on each trial and that trial’s WM accuracy on the other hand. Furthermore, the activity of the MI-IC statistically mediated the influence of the degree of surprise on each trial on WM accuracy. These findings showed that the same inhibitory mechanism that is explicitly recruited to interrupt motor activity in the SST is also automatically and rapidly recruited by surprising events to interrupt some types of cognitive activity (in this case, verbal WM). This is in line with the original conceptualization of the SST, which proposed that it probes the ability to “inhibit thought and action” (Logan and Cowan 1984).

In the current study, we followed this line of research by testing whether the inhibitory mechanism is also recruited to interrupt other types of cognitive representations—specifically, non-verbal types of WM such as visuospatial WM (Vogel and Machizawa 2004). Doing so is important for 2 reasons. First, it would shed a light on the generality of this mechanism for the interruption of cognitive processing. Studies have shown that the influence of the inhibitory mechanism can be broad and non-selective. In addition to its ability to inhibit both motor activity and verbal types of WM, studies have shown that it can even inhibit motor activity that is unrelated to the specific task: measurements of corticospinal excitability using transcranial magnetic stimulation show that when actions are successfully stopped, motor excitability is suppressed even in task-unrelated motor effectors, including those that are at rest (Badry et al. 2009; Cai et al. 2012; Wessel et al. 2013). Crucially, the same is true for unexpected events: Just like action stopping, surprising tones lead to a “global” suppression of motor excitability in task-unrelated effectors (Wessel and Aron 2013). We hypothesize that the global effects of the mechanism are explained by the properties of the neural network underlying the inhibitory mechanism. This network includes the presupplementary motor area (pre-SMA, Li et al. 2006; Aron et al. 2007; Boehler et al. 2014) and the subthalamic nucleus of the basal ganglia (Kuhn et al. 2004; Ray et al. 2009; Benis et al. 2014; Obeso et al. 2014). We believe that the subthalamic nucleus can "globally" inhibit thalamocortical loops that underlie ongoing motor—and potentially cognitive—representations (Wessel and Aron). Hence, it is of great importance to investigate the boundary conditions of this mechanism: Which types of representations does it affect, and which types remain unaffected?

The second reason to study whether visuospatial WM is subject to the same inhibitory mechanism that can disrupt verbal WM is that there is ongoing debate about whether different types of WM are maintained in similar ways, or whether their neural representations differ. While some studies suggest that WM systems might be shared across sensory domains (Saults and Cowan 2007; Morey et al. 2011), several studies of neural activity suggest that verbal and visuospatial WM might be represented in different ways (see D’Esposito and Postle 2015 for a recent review). Of specific importance in the context of the current study is the fact that verbal WM relies on subvocal rehearsal (Morra 2015), that is, the maintenance of information via “inner-speech” (Pihan et al. 2000; Narayanan et al. 2005; Baddeley 2012). Hence, the maintenance of verbal WM could have a motoric component, and studies have shown that subvocal rehearsal indeed relates to activity in motor areas for language production (e.g., Paulesu et al. 1993). Studies of visuospatial change detection, on the other hand, have shown that visuospatial WM is maintained in close proximity to the sensory areas that encode them in the first place (e.g., Goldman-Rakic 1987; Vogel and Machizawa 2004; Ester et al. 2009; Lara and Wallis 2014)—that is, visuospatial WM may not involve motor areas. Hence, the question of whether the surprise-related interruption of verbal and non-verbal (visuospatial) WM rely on the same neural system can not only speak to exactly how general the influence of the inhibitory mechanism is, but it can also potentially provide insights into potential differences and/or commonalities between different types of WM. Specifically, because the inhibitory mechanism that mediates surprise-related interruptions of verbal WM is primarily known as a motor inhibition mechanism, it might only be able to affect WM information whose maintenance involves the motor system, whereas it might be unable to affect “non-motor” types of WM. Some evidence of lateralization of the respective mechanisms is also interesting in this regard: Neuroimaging studies have suggested a potential lateralization of different types of WM to separate hemispheres, with verbal WM predominantly activating left prefrontal regions, and spatial WM predominantly activating the right hemisphere (Smith et al. 1996; D’Esposito et al. 1998). Since the motor inhibition mechanism recruited in the SST appears to be right-lateralized (Aron et al. 2014), this could also suggest a differential influence of this mechanism on different types of WM (cf. also Clark et al. 2007).

To test whether the inhibitory mechanism affects visuospatial WM, we developed a new experimental paradigm that is highly similar to the task used in the prior study that tested the influence of surprise on verbal WM (Wessel et al. 2016). Instead of verbal information (letter strings), however, participants encoded and maintained visuospatial information (the colors of an array of randomly arranged squares) across a delay interval. The WM aspect of this paradigm sought to most accurately replicate the original change-detection tasks used in Vogel and Machizawa (2004), which is one of the most common tests of visuospatial WM. Just as in the previous verbal WM study, the end of the delay interval was demarcated by a tone that preceded the WM probe. During training (and for most of the experiment), this tone was a stereotypical sine-wave tone. However, on a subset of trials in the experiment, this tone was surprisingly replaced by a brief birdsong segment.

Just like in the study of verbal WM (Wessel et al. 2016), we found that behaviorally, greater degrees of surprise associated with the tones lead to greater decrements in visuospatial WM accuracy. We then used the exact same procedure as in the verbal WM study to test the neural activity of the inhibitory mechanism (i.e., the MI-IC) in relation to the surprise-related WM decrement. After the WM task, participants performed a visual SST, in which they sometimes had to stop an impending action. We computed an ICA on the combined dataset of both tasks, and extracted the MI-IC for each participant—that is, the IC that reflected the motor inhibition process in the SST. We then investigated the MI-IC’s activity in relation to the surprising tones on single trials in the WM task: both event-related potentials (ERPs) and event-related spectral perturbations (ERSPs) were tested in relation to the degree of surprise carried by each tone, and its relation to the decrement in WM. As for verbal WM (Wessel et al. 2016), this was done using a single-subject,
single-trial GLM, which related the trial-level EEG response of the MI-IC to a trial-by-trial measurement of surprise, each trial’s WM accuracy, and the interaction between those 2 factors (i.e., a term that hence expresses the influence of surprise on WM accuracy).

We predicted that surprising tones in the WM task would elicit activity of the inhibitory mechanism, indicated by increased activity of the MI-IC. As outlined above, we had no directed prediction regarding the association between the activity of the inhibitory process and the surprise-related decrement in visuospatial WM, since a case can be made for either outcome: If the effect of the inhibitory mechanism is very general, it could conceivably affect visuospatial WM, leading to the same outcome as in the study of verbal WM (Wessel et al. 2016). In this case, the MI-IC response would show a significant positive correlation with the interaction term that expresses the relationship between the degree of surprise and the WM accuracy (tones with higher degrees of surprise that are followed by greater decrements of WM should be accompanied by greater MI-IC activity). This was the case for verbal WM in the previous study. However, since visuospatial WM is potentially maintained in a different neural system than verbal WM, it might not be susceptible to interruptions by the same neural mechanism. In this case, there should be no relationship between the MI-IC’s activity and the interaction term. However, we would then expect a positive correlation of the MI-IC with the regressor that expresses the degree of surprise independently of WM accuracy (i.e., the trial-to-trial surprise regressor), as surprising events should still elicit activity of the inhibitory mechanism (Wessel and Aron 2013). In case of the latter outcome, we aimed to investigate other candidate mechanisms that have been shown to be invoked by surprising events—specifically, the early auditory N1 (Naatanen and Picton 1987) and the later centro-parietal P3b (Polich 2007).

Materials and Methods

Participants

Participants were recruited from the University of Iowa community using a dedicated campus-wide email list for research recruitment. To ensure sufficient statistical power to detect an effect, we aimed to collect the same amount of participants as in our prior study of surprise-related interruptions of verbal WM (N = 20). We inadvertently collected one additional participant, resulting in an N of 21 (mean age: 27.67, SEM: 2.46, 12 female, 2 left-handed). Participants participated in exchange for a prorated hourly payment of $15. The study was approved by the local Institutional Review Board at the University of Iowa (#201511709).

Materials

Stimuli were displayed on an IBM-compatible desktop computer running Fedora Linux and Psychtoolbox 3 (Brainard 1997) under MATLAB 2015a (TheMathWorks), connected to a 17-inch Dell flat-screen monitor. Responses were made via a connected USB keyboard.

WM Task

The WM task was designed as a hybrid of the verbal WM task used to test the influence of surprise on verbal WM (Wessel et al. 2016) and the classic change detection paradigm used to probe visuospatial WM (Vogel and Machizawa 2004). From the verbal WM task in Wessel et al. (2016), it retained the exact timing and layout of the surprising events and the general task structure, and from the visuospatial change detection task of Vogel and Machizawa (2004), it retained the exact layout and timing of the WM stimuli, delay interval, and probe.

On each trial, participants saw a fixation cross (400 ms), followed by a directional cue (black arrow displayed above the fixation cross for 200 ms), followed by an array of colored squares displayed for 100 ms (Fig. 1). Participants were instructed to retain the squares on the cued side in WM across a subsequent delay interval. In the classic change-detection paradigm, that delay interval is typically 900 ms long. This was the critical interval in the current study as well (i.e., all comparisons were made based on this interval). However, a key aspect of the WM task used to test the effect of surprising events on verbal WM was that the tones must convey some information about the task (otherwise, they can easily be ignored). Hence, just as in the verbal WM task, our task featured 4 different delay intervals (the classic 900 ms, as well as additional intervals of 700, 1100, and 1300 ms duration). Participants were instructed that on each trial, a tone would inform them of the impending end of the delay interval, and that it would be helpful for them to attend to the tone, as the intervals were of different duration. The tone was played 300 ms before the end of each interval, for a duration of 200 ms. At the end of the interval, a second array of squares was displayed in the same position of the first array, and the participants were supposed to make a judgment of whether one of the squares on the cued side of the array had changed in color. The next trial began after the response was made.

Participants performed 432 trials of this experiment for the main task. Ninety trials each were trials with standard tones at each of the 4 delay intervals. The remaining 72 trials were trials with a 900 ms delay interval (the standard delay interval in the classic change-detection task), but instead of the standard tone, a random, unique birdsong segment was played (the segments were normalized in volume to the 200 ms, 600 Hz sine-wave standard tone, and cropped to a duration of 200 ms). To be consistent, these were the same tones used in prior studies of the effect of surprising events on verbal WM, Wessel et al. (2016). Thus, the overall probability of a surprising tone was 1/6 (72 out of 432 trials). Within each of the 5 different types of trials (standard tones at the 4 delay intervals, surprising tones at the 900 ms interval), the number of trials with and without a change of color within the cued array was identical. The side of the cued array (left, right) was pseudo-randomly determined on each trial with equal probability. The design of the WM arrays was designed to be identical to Vogel and Machizawa (2004). In detail, each individual square was displayed at a visual angle of 0.65°, with a minimum space of 2° between each square. Each array of squares was presented within a possible area of 4 × 7.3° to the left and right of the fixation, with the center of the area 3° to the left and right of the cross. Possible colors were green, blue, magenta, yellow, white, black and red, with no color repeating more than twice within each array. The colors of the non-cued side of the array never changed between the first array and the probe array. Participants responded to the probe array by pressing upward arrow key for perceived matches, and the downward arrow key for perceived non-matches.

Finally, participants practiced the task before the actual experiment. In addition to familiarizing the participants with the mechanics of the task, the practice served 2 additional purposes. First, since the practice only used the standard tone, it established an expectation space for the actual experiment,
and allowed the surprising tones to be maximally unexpected. Second, it allowed us to gauge the number of arrays each participant could handle while 1) performing above chance, and 2) performing below ceiling. To this end, the participants started with a load of 3 (i.e., 3 squares were displayed in each hemisphere), which was then adjusted based on the performance in the practice blocks. If a participant’s accuracy in the preceding block was above 80%, the load was increased by 1. If accuracy was below 60%, the load was reduced by 1. Participants performed 6 blocks of practice, with 12 trials each. This procedure has proven to be effective in the study of the effects of surprising events on verbal WM (Wessel et al. 2016), and again proved to be successful here (see Results).

**Stop-signal Task**

We used a standard version of the visual SST. Trials began with a fixation cross (500 ms duration), followed by a white leftward or rightward arrow (go-stimulus). Participants had to respond as fast and accurately as possible to the arrow using their left and right index finger (the respective response-buttons were q and p on the QWERTY keyboard). On 33% of trials, a stop-signal occurred (the arrow turned from white to red) at a delay after the go-stimulus (stop-signal delay, SSD). The SSD, which was initially set to 200 ms, was dynamically adjusted in 50 ms increments to achieve a p(stop) of 0.5: after successful stops, the SSD was prolonged; after failed stops, it was shortened. This was done independently for leftward and rightward go-stimuli. Trial duration was fixed at 3000 ms. Six blocks of 50 trials were performed (200 go, 100 stop).

**Procedure**

After arrival at the laboratory, participants signed written informed consent and provided demographic information. After the EEG cap was affixed to their head, the experimenter handed the participants written instructions for the WM task, and ensured that the task was fully understood. After performing the WM task training and the main WM task, participants were then handed written instructions for the SST. They practiced the SST for one block of 24 trials (16 go, 8 stop), and then performed the main stop-signal experiment. After the experimentation concluded, the experimenter removed the EEG cap and debriefed the participant.

**Behavioral Analysis**

For the SST, we calculated go-trial reaction time (GoRT), failed stop-trial reaction time (FsRT) and mean rate of stopping (p(stop)). These values were used to test the predictions of the race model, which predicts that GoRT will be longer than FsRT, and to test whether the SSD staircase method was effective in each participant. To this end, participants were excluded if their GoRT was faster than their FsRT, or if their p(stop) was outside the range of 0.4–0.6. We calculated stop-signal reaction time (SSRT) based on the block-wise integration method (based on the arguments put forward in Boehler et al. 2012; Verbruggen et al. 2013). To be consistent with our previous papers, which used the mean method (Verbruggen and Logan 2008), we also report the SSRT results based on that method.

For the WM task, we calculated the accuracy on the probe as a percentage of correct responses in each condition. Specifically, we compared WM accuracy after surprising tones in the 900 ms delay condition to the WM accuracy after standard tones in the same condition using a 2-sided t-test at P < 0.05 to test any effects of the surprising trials across the entire experiment (i.e., regardless of the degree of surprise). More importantly, we tested the role of surprise in WM accuracy by splitting the surprising trials by whether participants responded correctly or incorrectly to the WM probe, and compared the surprise values of trials in either condition. Our...
hypothesis was that WM trials with incorrect responses after surprising tones would have been more surprising than WM trials with correct responses after surprising tones. The surprise value of each individual surprising tone $i$ was calculated using a Bayesian method that compares the expected probability of occurrence of an unexpected tone prior to the current trial to the probability of occurrence an unexpected tone after the current event (using the Kullback–Leibler divergence):

$$\text{Surprise}_i = \log\left( \frac{p_{\text{unexpected}}(1\ldots i)}{p_{\text{unexpected}}(1\ldots i-1)} \right)$$

These values were tested against each other using a 2-sided $t$-test at $P < 0.05$. This measure interprets surprise as an objective quantity that reflects a discrepancy between the exact observed prior and posterior likelihood of a given event. While it is unclear whether humans’ predictions strictly adhere to these exact, objective probabilities, it is almost impossible to gauge the subjectively perceived likelihood of an event on each trial without influencing the nature of the unexpected event (strictly speaking, prompting the participants to actively indicate whether they “expect” an unexpected event on a given trial would likely change the processing of any “unexpected” event that were to occur). An upside of this measure is that it can be clearly estimated independently of the data, purely based on environmental properties.

**EEG Recording**

EEG was recorded using a 62-channel electrode cap connected to 2 BrainVision MFlplus amplifiers (BrainProducts). Two additional electrodes were placed on the left canthus (on the lateral part of the orbital bone of the left eye) and under the left eye. The ground was placed at electrode Fz, and the reference was placed at electrode Pz. EEG was digitized at a sampling rate of 500 Hz.

**EEG Preprocessing**

Data were preprocessed using custom routines in MATLAB 2015a (TheMathWorks). ICA were performed using functions from the EEGLAB toolbox (Delorme and Makeig 2004). After import into MATLAB, both datasets were merged (the stop-signal dataset was appended to the WM task data), and the continuous time-series were filtered using symmetric 2-way least-squares finite impulse response filters with a high pass cutoff of 0.5 Hz and a low pass cutoff of 50 Hz. The continuous time-series were then visually inspected for channels with non-stereotypic artifacts, which were excluded from further processing. The remaining data were visually inspected for segments with non-stereotypic artifact activity (e.g., muscle artifacts), which were removed from further analysis of the continuous data. After artifact removal, the data were re-referenced to the common average, and subjected to a temporal infomax ICA decomposition algorithm (Bell and Sejnowski 1995), with extension to subsymmetric sources (Lee et al. 1999). The resulting component matrix was screened for components representing eye-movement artifacts using outlier statistics. The IC selection was visually inspected for accuracy of the automated classification, and artifact components were removed. The remaining components were subjected to further analyses.

**EEG Analysis Logic**

The EEG analysis logic is the same as in Wessel et al. (2016). For a theoretical overview, see Wessel (2016). In short, to test whether the brain network for motor inhibition is related to the effect of surprising events on WM (which was the case for verbal WM in the previous study), the data from both the WM task and the SST were subjected to the same ICA in each subject. ICA then disentangles the independent signal components represented in the combined EEG signal mixture at the scalp, as those ICs correspond to independent neural processes (Onton et al. 2006). We then use clear-cut statistical criteria to identify a single IC in each participant that indexed motor inhibition in the SST (we call this the MI-IC). This is done based on a priori knowledge about the properties of the neural process for motor inhibition as measured by EEG (see below for a list of criteria).

Importantly, none of the WM task portion of the data is taken into account for this selection process. After the MI-IC is selected from the SST, its activity in the WM task portion is then investigated. As for verbal WM in Wessel et al. (2016), we aimed to test 1) whether surprising events recruit the MI-IC (i.e., whether activity in the MI-IC is increased on surprising vs. non-surprising tones), and 2) whether the MI-IC’s activity after surprising tones is related to any changes in accuracy on the WM probe.

The core of the inferential logic applied in this procedure is as follows: Because ICA identifies all independent neural source components that contribute to the EEG data and disentangles them from one another, identifying a component that represents one process (e.g., motor inhibition in the SST) and showing that its activity is related to aspects of other tasks (e.g., the occurrence of surprising tones in the WM task) shows that the same neural process, produced by the same neural generator, is active in both cases. In the particular case of the current study, it would mean that surprising events recruit the motor inhibition process, and, depending on the analysis of the MI-ICs relation to WM accuracy after surprise, do so potentially in relation to WM accuracy.

**MI-IC Component Selection**

The MI-IC was selected from each participant’s SST data based on a priori knowledge about the properties of the inhibition process in the SST and its neural concomitants. Prior studies have shown that motor inhibition is indexed by a low-frequency (delta/theta frequency band, 1–8 Hz, Huster et al. 2013) component at fronto-central electrodes (Kok et al. 2004), which likely reflects the activity of the pre-SMA part of the stopping network (Huster et al. 2011; Albert et al. 2013). This prior research has shown that the onset of the fronto-central P3-ERP component in particular indexes successful stopping (Bekker et al. 2005; Schievarens et al. 2015): Its amplitude is increased for successful compared with failed stop-trials (Kok et al. 2004), and its onset correlates highly with the speed of stopping, measured by SSRT (in a sample of $N > 60$, a correlation in excess of $r = 0.6$ was found between the onset of the P3 in individual subjects and their SSRT, Wessel and Aron 2015).

Hence, we utilized our prior knowledge of these properties of the neural marker of successful stopping in the SST to identify components in fully automated fashion (this procedure was identical to Wessel et al. 2016). From each individual participants ICA, we first selected each component whose weight matrix had its maximal rectified weight at one of the fronto-central electrodes (FCz, Cz, C1, C2, FC1, FC2). We then averaged
those components’ back-projected channel-space activity at these fronto-central electrodes within the 500 ms time-period following the stop-signal, and correlated this event-related average activity to the event-related average activity of the overall EEG data (i.e., the EEG data based on the back-projection of “all” non-artifact ICs for that participant) in that time range. The component that showed the highest correlation with the overall ERP was selected as the MI-IC. The results section and Fig. 2 show the properties of the selected IC across the group, and confirm that the same properties (earlier onset for successful vs. failed stopping, high correlation of neural signal onset to SSRT, increased activity in the delta- and theta- frequency bands) that have been found for the stopping process in other studies are also present for the selected MI-ICs for the current study.

Event-related Potentials

To identify the MI-IC (see above), ERPs were calculated at the fronto-central electrode cluster (FCz, Cz, FC1, FC2, C1, C2) by averaging the mean activity at these electrodes and segmenting the data with respect to the onset of the stop-signal on successful and failed stop-trials (a time period of −100 to 700 ms surrounding the stop-signal was chosen, and a baseline correction was applied based on the average activity −100 to 0 ms before the stop-signal). To illustrate the presence of the stop-signal P3 relative to go-trials (Fig. 2A), data were also extracted from matched go-trials for each condition (successful and failed stop-trials). Matching of go-trials was done as follows: For every stop-trial, we selected the next go-trial in the trial sequence, and we extracted the ERP from that trial time-locked to the time point that corresponded to stimulus onset + the current SSD (this is the time point at which the stop-signal would have occurred on that go-trial, based on the state of the current staircase). Since we are comparing activity within individual independent components for each participant, the potential differences in motor activity between the individual trial-types should not matter, as motor activity and inhibition-related activity will be represented in separate ICs.

Furthermore, ERPs were extracted around tones in the WM task. Epochs were generated beginning from −300 to 700 ms around tone onset, and averaged separately for trials in the

<table>
<thead>
<tr>
<th>Component</th>
<th>Weight (Average)</th>
<th>Weight (S.E.M.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>STOP SIGNAL</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>µV</td>
<td>STOP SIGNAL</td>
<td>SUCCESSFUL STOP</td>
</tr>
<tr>
<td>Frequency</td>
<td>Successful stop</td>
<td>600</td>
</tr>
<tr>
<td>Hz</td>
<td>P3 onset (ms)</td>
<td>P3</td>
</tr>
<tr>
<td>Successful stop trials</td>
<td>13</td>
<td>30</td>
</tr>
<tr>
<td>P3 onset - SSRT correlation</td>
<td>r = 0.47; p &lt; .05</td>
<td></td>
</tr>
<tr>
<td>TIME-FREQUENCY PROPERTIES</td>
<td>Successful stop trials</td>
<td></td>
</tr>
<tr>
<td>% change</td>
<td>STOP</td>
<td>STOP</td>
</tr>
</tbody>
</table>

Figure 2. Summary of the MI-IC identified based on the SST. The components show the typical properties of the MI-IC known from previous studies. (A) Grand-average ERP of each component, backprojected into channel-space, measured at fronto-central electrodes (FCz, FC1, FC2, Cz, C1, C2; see electrode location image in top right corner), separately for successful stop-trials (green), go-trials matched to successful stop-trials (go-trials for which the SSD staircase was at the same position, black), failed stop-trials (red) and go-trials matched to failed stop-trials (yellow). Transparent line denotes 1 SEM. Bottom left inlay shows the topographical distribution of the component weight maps (rectified to show positivity at Cz), alongside its SEM. (B) Participants’ P3 onset was earlier for successful compared with failed stop-trials, again replicating previous studies, P < 0.0001. (C) Correlation between the statistical onset of the P3 ERP for each participant and their SSRT (according to the block-wise integration method). Each participant is represented by a black dot. Replicating previous studies, the 2 measures correlated highly at r = 0.5, P < 0.05. (D) Time-frequency plot of the event-related spectral properties of the MI-IC, on successful stop-trials, oscillatory power was increased in the delta (1–4 Hz) and theta (5–8 Hz) frequency bands (again replicating previous studies). Power is displayed in units of percent change from baseline (i.e., the average activity in the 300 ms prior to the stop-signal).
900 ms delay condition with standard tones to trials in the 900 ms delay condition with surprising tones. The following ERPs were generated: To test the relationship between the MI-IC and the surprising tones, ERPs were generated based on a backprojection of the MI-ICs to the fronto-central electrode cluster (see above). Furthermore, to test the relationship between other classic surprise-related ERPs not accounted for by the MI-IC, we also generated the early fronto-central N1 potential by calculating the same ERP at the same electrodes, yet based on the backprojection of all other non-artifact ICs that were not the MI-IC (i.e., the non-artifact EEG data with the contribution of the motor inhibition process removed). Lastly, we also generated the centro-parietal P3b by averaging the same non-MI-IC backprojection at electrode CPz.

**MI-IC P3 Onset in the SST**

To test whether the properties of the neural process that indexes motor inhibition matched those of the MI-ICs selected from the SST in the current experiment, we detected the onset of each participant’s fronto-central P3 back-projection from the MI-IC. This was done as follows: The difference wave between stop- and matched go-trials were computed (independently for successful and failed stop trials). The entire difference wave was tested for differences from zero using sample-by-sample t-tests between single-trial stop- and go-trial amplitudes at a 2-sided \( p < 0.05 \) (false-discovery rate (FDR) corrected using the procedure proposed by Benjamini et al. 2006). Then, the peak of the P3 difference wave in the time period 200–400 ms following the stop-signal was detected. This peak sample was significantly different from 0 in all cases (\( p < 0.05 \), FDR-corrected). Working “backwards” from that sample, we detected the sample closest to the onset of the stop-signal at which the positive difference wave was still significantly different from 0 (at \( p < 0.05 \), FDR-corrected). Put differently, we defined the P3 onset in each participant as the point after the stop-signal at which the stretch of significant samples that included the peak of the P3 began. We then compared this onset between successful and failed stop-trials across participants using a paired-samples t-test, and correlated this onset with SSRT across participants.

**Event-Related Spectral Perturbation**

The ERSP was calculated using the absolute of the Hilbert transform as an analytic signal for each center frequency ranging from 1 to 30 Hz, with a frequency window of \( +/- 0.5 \) Hz. Values for the analytic signal were converted into \( % \)-change compared with a 300 ms pre-trial baseline. To illustrate the MI-ICs’ time-frequency response on successful stop-trials, ERSPs were extracted from 0 to 700 ms after the stop-signal. For the WM task analysis, ERSPs were extracted from 0 to 500 ms after tone-onset.

**Single-trial GLM**

To investigate the relation between the MI-ICs’ activity and the 3 explanatory factors (the degree of surprise of each tone, WM accuracy after surprising tones, and the interaction between the 2), we constructed individual GLMs for each participant. WM accuracy on each surprising trial was coded with 1 for misses or false alarms, and 0 for hits or correct rejections. Surprise was computed as above. Both variables were demeaned, and then their interaction was calculated by multiplication. De-meaning the variables controls for the fact that surprise and WM failure are correlated, and results in a valid regression model (the variance inflation factor for each condition was below 10 in all participants). These 3 variables were used as regressors to model each individual time “frequency sample point of the tone-related MI-IC ERSP and ERP on each surprising trial, resulting in a vector/matrix of beta weights for each participant, with the size being the same as the standard ERP (a single vector with 500 elements corresponding to 1000 ms of data at 500 Hz) and ERSP response (a 30 \( \times \) 250 matrix corresponding to 30 frequencies and 500 ms of data at 500 Hz). This matrix was then tested for significance as described in the next section. The final GLM has the following format for the ERSP:

\[
\text{ERSP}_{\text{f,s}} = \beta_{\text{intercept}} + \beta_{\text{surprise}} \times X_1 + \beta_{\text{WMaccuracy}} \times X_2 + \beta_{\text{surprise, WMaccuracy}} \times X_3 + \epsilon \quad \text{with } f \in \{1 : 30\},
\]

frequencies in Hz, and \( s \in \{1 : 250\} \), samples at 500 Hz.

and the following format for the ERP:

\[
\text{ERP}_{\text{f,s}} = \beta_{\text{intercept}} + \beta_{\text{surprise}} \times X_1 + \beta_{\text{WMaccuracy}} \times X_2 + \beta_{\text{surprise, WMaccuracy}} \times X_3 + \epsilon \quad \text{with } s \in \{1 : 500\},
\]

samples at 500 Hz.

**EEG Significance Testing**

In all cases (average ERSPs, average ERPs, ERSP regression weights, ERP regression weights), testing was done using sample-by-sample paired t-tests at a 2-sided \( p < 0.05 \) (false-discovery rate (FDR) corrected using the procedure proposed by Benjamini et al. 2006). Then, the peak of the P3 difference wave in the time period 200–400 ms following the stop-signal was detected. This peak sample was significantly different from 0 in all cases (\( p < 0.05 \), FDR-corrected). Working “backwards” from that sample, we detected the sample closest to the onset of the stop-signal at which the positive difference wave was still significantly different from 0 (at \( p < 0.05 \), FDR-corrected). Put differently, we defined the P3 onset in each participant as the point after the stop-signal at which the stretch of significant samples that included the peak of the P3 began. We then compared this onset between successful and failed stop-trials across participants using a paired-samples t-test, and correlated this onset with SSRT across participants.

**Bayes Factor Analysis**

In case that—unlike for verbal WM—there is no relation between the MI-IC activity and surprise-related WM decrements (i.e., the surprise * WM accuracy interaction) or between MI-IC activity in WM accuracy in general, we planned to perform a Bayes Factor analysis to precisely quantify the support for the alternative hypothesis (i.e., the hypothesis that there is a relationship). To this end, we selected the highest t-value out of the entire ERSP spectrum/ERP epoch for the GLM analysis with the interaction regressor (i.e., the right panel in Fig. 4A for ERSP and in Fig. 4B for the ERP) and the WM accuracy regressors (the middle panels in Fig. 4). We then subjected those 4 t-values to a Bayes Factor analysis using the Bayes Factor Package for R (Rouder et al. 2009). We furthermore reduced the scale of the prior from the default \( \pi_0 \approx 0.707 \) to 0.5. This increases the Bayes Factor, as it takes account smaller effect sizes as potential evidence for the alternative hypothesis (i.e., that there is a relationship between the MI-IC ERSP/ERP and the respective regressors). The combination of these 2 factors (lowered scale of the prior and selection of the single-highest t-value from the entire ERSP time-frequency range/ERP time-range without any correction for multiple comparisons) leads to a very strong bias towards high Bayes Factors, that is, towards finding stronger
Results

Behavior: SST

FsRT (mean: 474 ms, SEM: 14 ms) was faster than GoRT (mean: 561 ms, SEM: 15 ms) for each individual participant (t-test of the group averages: t(20) = 14.78, P < 10^{-4}), indicating the validity of the race model. Mean p(stop) was 0.52 (SEM: 0.005), and all participants' p(stop) was between 0.4 and 0.6, indicating the effectiveness of the SSD staircasing procedure. Mean SSRT was 219 ms (according to the block-wise integration method) and 240.5 ms (according to the mean method). Mean SSD was 321 ms (SD: 19 ms).

Behavior: WM Task

Regardless of surprise value, probe accuracy in the 900 ms delay condition was numerically decreased after surprising tones compared with standard tones when averaged across the entire experiment (79% vs. 82%), yet this difference was not significant (t(20) = 1.674, P = 0.109, 2-sided, d = 0.4). More importantly, when comparing the degree of surprise of each surprising tones between trials with correct and trials with incorrect responses on the WM probe, we found that surprising tones followed by incorrect WM responses were significantly more surprising than surprising tones followed by correct WM responses, with large effect size: t(20) = 2.74, P = 0.013, d = 1.02). This is the same result (with comparable effect size) as was found for verbal WM in Wessel et al. (2016). For the sake of completeness, the accuracy on the 700, 1100, and 1300 ms delay intervals (standard tones only) was 82%, 82%, and 80.5%, respectively. As for verbal WM, reaction times to the probe did not differ depending on whether the tones were surprising or not: t(20) = 0.097, P = 0.93, d = 0.008.

MI-IC Indexes Motor Inhibition in the SST

As can be seen from Fig. 2, the MI-IC selection reflected the known properties of the neural signature of motor inhibition in the SST: the onset of the fronto-central P3 after the stop-signal was significantly earlier on successful compared with failed stop trials (t(19) = 8.94, P < 10^{-7}). Moreover, the onset of the P3 on successful stop-trials highly correlated with SSRT across participants; r = 0.47, P = 0.036 for the block-wise integration method, r = 0.48, P = 0.032 for the mean method (one outlier was removed from both analyses based on a criterion of Cook’s d > 1. With this outlier included, the correlation was even larger; r = 0.63 for the mean method, and r = 0.56 for the integration method).

MI-IC Accounts for the P3a After Surprising Tones in the WM Task

We first tested whether the activity of the MI-IC was related to the processing of surprising tones in the WM task, regardless of subsequent accuracy on the WM probe. Our previous study has shown that the MI-IC is active following surprising events and accounts for the fronto-central “novelty-P3a” (Wessel and Aron 2013). Figure 3 shows that the same was true for the current study. Figure 3B shows that frequencies in the delta (1–4 Hz), and theta (5–8 Hz) frequency bands of the MI-IC were significantly increased on surprising versus standard tones.

MI-IC Activity Relates to Surprise, but not WM Accuracy

To test the main hypothesis of the paper, we used the single-trial sample-by-sample GLM on both the fronto-central MI-IC ERP/ERSP. In Wessel et al. (2016), the ERSP GLM showed that the MI-IC did not relate to the degree of surprise or the WM accuracy by themselves, but was only related to the “interaction” of the 2 factors. In other words, activity of the MI-IC was increased specifically when high degrees of surprise lead to verbal WM failure (furthermore, the activity of the MI-IC statistically mediated the effect of surprise on verbal WM). The current analysis showed that the same was not the case for visuospatial WM: Neither WM accuracy nor the interaction term was significantly positively (or negatively) correlated with the MI-IC ERSP response. In turn, the ERSP response of the MI-IC after surprising tones clearly related to solely the degree of surprise on each trial. Specifically, 4 Hz (delta) and 6 Hz (theta) band activity related directly to the surprise term of the GLM (Fig. 4A). This means that while the activity of the MI-IC clearly relates to the degree of surprise, there was no effect of MI-IC activity on surprise-related decrements in WM, which is in contrast to the findings of our prior study with verbal WM (Wessel et al. 2016).

To ensure that no other type of activity of the MI-IC was in any way related to WM accuracy we also analyzed the ERP in the same manner as the ERSP. The ERP analysis confirmed the analysis of the ERSP: The fronto-central MI-IC ERP, moreover in the P3a time-range, strongly positively correlated with the degree of surprise (i.e., greater surprise was followed by more positive MI-IC P3a amplitudes). In contrast, just like for the ERSP, the ERP amplitude was unrelated to both WM accuracy and the interaction between surprise and WM (Fig. 4B).

Bayes Factor Analysis

The highest t-value for the relationship between MI-IC activity and WM accuracy across the entire time-frequency spectrum of the MI-IC ERSP was 2.06, which, at a sample size of 21, and a scaling of the prior of r = 0.5, results in a maximum Bayes Factor of 1.51 for the ERSP. For the surprise * WM accuracy interaction, the maximum t-value was 1.65, and the resulting Bayes Factor was 1.44. According to common convention (Jeffries 1961), any Bayes Factor below 5 is “barely worth mentioning” (which is the lowest category and includes a Bayes Factor of 0, i.e., no measured difference at all). In contrast, the (significant) MI-IC ERSP relationship with the pure surprise term produced a Bayes Factor in excess of 10.18, denoting “strong” evidence for the alternative hypothesis, according to Jeffries.

For the ERP, the WM accuracy term produced a maximum Bayes Factor of 1.62 and the interaction term produced a maximum Bayes Factor of 4.1 (i.e., both again were <5), whereas the pure surprise term produced a maximum Bayes factor of 1607 (indicating “decisive” evidence for the alternative hypothesis according to Jeffries). Taken together, the Bayes Factor analysis shows that, even when taking into the account the highest t-value out of 500 tests (ERP) and 7500 (ERSP), respectively, and with a very low threshold for evidence for the alternative hypothesis, no evidence for a relationship between the surprise * WM accuracy regressor and either the ERSP or the ERP could be found (while the pure surprise term yielded “strong” and “decisive” evidence, respectively).
Non-MI-IC Activity and Its Relation to Surprise-Related WM Failure

Lastly, because we found that unlike for verbal WM, the activity of the MI-IC was unrelated to the surprise-related interruption of visuospatial WM, we tested other candidate processes that could relate to surprise-related decrements in visuospatial WM (as WM was clearly affected by surprise behaviorally). We investigated 2 potential candidates, both of which are well-established signatures of novelty processing: The early fronto-central N1, which is increased for unexpected auditory events (Naatanen and Picton 1987), and the P3b (Polich 2007), which is an independent part of the P3a/P3b complex after novelty (Debener et al. 2005a). To test the involvement of the processes reflected in these ERPs in surprise-related WM decrements, we split the ERP amplitudes after surprising tones by WM accuracy, and tested the 2 resulting trial types (surprising tones followed by correct or incorrect WM responses) against each other. To make sure that any finding would not be conflated with any MI-IC activity, we quantified the activity of these ERPs solely based on the (non-artifact) independent components that were not the MI-IC. This analysis showed that the N1, just like the activity of the MI-IC, was unrelated to WM accuracy (none of the samples in the N1 time-range, and in fact, no samples in that ERP epoch whatsoever, were different between the 2 different trial types, Fig. 5B). In contrast, the centro-parietal P3b showed a straightforward relationship to visuospatial WM accuracy: In the time range from 400 to 430 ms, the P3b to surprising tones was clearly increased on trials in which surprising events lead to visuospatial WM failure. The centro-parietal P3b is more active when surprising events lead to visuospatial WM failure. The centro-parietal P3b is more active when surprising events lead to visuospatial WM failure.

Discussion

Our study shows that while surprising events negatively affect visuospatial WM, this decrement is clearly not mediated by the same mechanism that interrupts verbal WM. We used the same method used in a recent study to show the effect of a neural mechanism for motor inhibition on verbal WM (Wessel et al. 2016) and the same stimulus material used in classic studies of that type of non-verbal WM (Vogel and Machizawa 2004). While—just like for verbal WM—incorrect WM responses
after surprising tones were related to higher degrees of surprise, this decrement in WM accuracy was unrelated to neural activity of the inhibitory mechanism.

Importantly, our study had enough power to show a potential difference, as it had the same (actually, a slightly larger) sample size than our prior study of verbal WM, and none of the effects were close to significance. Additionally, a Bayes Factor analysis showed no evidence for the alternative hypothesis that there is a relationship between surprise-related WM decrements and MI-IC activity, even using very lenient criteria that take into account even small amounts of evidence. Furthermore, our study included multiple "negative" controls: Activity of the motor inhibition mechanism was clearly increased following surprising tones (Fig. 3), and even more so, its activity related to the trial-by-trial level of surprise (Fig. 4). Additionally, we found that the WM decrement after surprising events was directly related to another classic neural index of surprise processing: the centro-parietal P3b. The P3b after surprising tones was significantly increased when WM was interrupted compared with when it was not.

While the motor inhibition mechanism accounts for the fronto-central P3a after surprise (Fig. 3A), our results are clearly in line with research that shows that the P3 potential after surprising events is actually a conglomerate of 2 independent processes, one reflected in the earlier P3a, and one reflected in the subsequent P3b (Debener et al. 2005a; Polich 2007). It is debated which exact contributions to novelty processing these ERPs represent, as their underlying processes are unclear. However, their relative timing supports the idea that inhibition-related interruptions of cognitive processing (represented in the P3a) is followed by additional, independent processing of the surprising event (P3b). Importantly, this is in line with our theoretical conceptualization of how cognitive processes can be interrupted. Based on the timing of the neural processes following surprising events, and their differential relationship to cognitive interruptions, we believe that surprising tones elicit a sequence of processing that starts with the detection of the surprising event (which can happen as early as 50 ms in the case of surprising tones, Gamble and Woldorff 2015), which triggers an immediate inhibition of ongoing motor and cognitive processes (indexed in EEG by activity of the MI-IC). This is then followed by a reorienting of attention to the surprising event, which allows for an “active engagement” with the event. The latter process could be reflected potentially in the P3b (Donchin 1981; Kok 2001). This is in line with the fact that studies of the autonomic nervous system found higher correlations between the P3b and autonomic arousal compared with the P3a, which indicates that the P3a might be unrelated to the autonomic (re-)orienting response to surprising events (Sokolov 1963), whereas the P3b might be one of the central nervous comitants of the orienting response (Rushby et al. 2005). This is also in line with other conceptualizations of the P3a and P3b after novelty, according to which the P3a reflects a “disruption of frontal attention engagement”, whereas the “P3b originates when temporal-parietal mechanisms process the stimulus information for memory storage” (Polich and Criado 2006).
A key finding of our study is that visuospatial WM is not susceptible to interruptions by the same type of neural system that interrupts verbal WM after surprising events. There is currently ongoing debate whether verbal and visuospatial WM engage overlapping systems during maintenance of information or not (Park et al. 2002; Kane et al. 2004; Alloway et al. 2006). Since we found the mechanism that mediates the surprise-related interruptions of “verbal” WM to be unrelated to the effects of surprise on “visuospatial” WM, our results indicate that the 2 types of WM might be maintained in different ways. This is corroborated to some extent by the respective neural indices that are tied to their maintenance. Visuospatial WM elicits a contralateral delay activity (CDA) (Vogel and Machizawa 2004) over parieto-occipital cortices, that is, in close proximity to where these stimuli are initially encoded (visual cortex). In contrast, the maintenance of verbal (or verbalizable) WM might relate to motor structures (Paulesu et al. 1993; Petrides et al. 1993), and potentially involves thalamocortical loops (Chang et al. 2007; Chatham et al. 2014; Chatham and Badre 2015). This offers 2 potential explanations as to why the inhibitory mechanism investigated here can interrupt verbal WM, but seems to leave visuospatial WM unaffected.

The first explanation holds that the inhibitory mechanism is in fact largely specific to the motor realm, being able to non-selectively affect motor representations. In consequence, verbal WM can be targeted by this mechanism only since it draws upon the motor system, which is susceptible to the global interruption effectuated by the inhibitory system. In contrast, visuospatial WM, which does not rely on motor structures, and may be maintained in more sensory regions, would be not affected by this mechanism.

The second possible explanation would be that while verbal WM could be maintained in reverberant thalamocortical loops, visuospatial WM might not involve the same system. If that is the case, the purported anatomy of the inhibitory mechanism could explain why verbal WM can be affected by it, whereas visuospatial WM cannot. We hypothesize that the global influence of the inhibitory mechanism is a result of the following processing chain: after unexpected events (or stop-signals), the cortical regions of the brain network underlying the inhibitory mechanism (pre-SMA and right inferior frontal cortex) signal the need for inhibition to the subthalamic nucleus of the basal ganglia (which could potentially happen very rapidly through purported frontal-basal ganglia “hyper-direct” pathway, Nambu et al. 2002). The subthalamic nucleus, in turn, can inhibit ongoing motor representations by non-selectively interrupting thalamocortical drive. This is done through a non-selective excitation of the output nuclei of the basal ganglia, particularly the internal globus pallidus (some anatomical evidence indicates that the subthalamic nucleus could have such broad projections to the pallidum, Hazrati and Parent 1992; reviewed by Wessel and Aron). According to our proposed framework, any representation maintained in thalamocortical loops might therefore be susceptible to interruption by the inhibitory mechanism. If visuospatial WM contents, like the stimuli used here, are in fact not maintained in such loops, and are rather maintained through ongoing activity near or in the visual regions that encoded the stimulus material in the first place, this would explain why the mechanism does not affect this type of WM.

Our wider theory of how surprising events disrupt cognition proposes a 2-stage model: After the surprising event is detected, the inhibitory mechanism immediately interrupts ongoing representations (potentially by shutting off thalamocortical drive through non-selective projections in the basal ganglia, see above). This first step, in turn, enables a subsequent shift of attentional focus aimed at engaging and processing the surprising event. Taken together with our recent study (Wessel et al. 2016), the results reported in this study provide some preliminary evidence for this model: The findings show that different types of cognitive representations (visuospatial WM and verbal WM) are interrupted by 2 separate mechanisms, which are indexed by independent neural signatures: The P3a, which is accounted for by the MI-IC, and the MI-IC-independent P3b. These 2 processes occur in immediate succession and have, in fact, often been studied as a common phenomenon, owing to their ubiquitous co-occurrence after surprising events. We propose that the P3a reflects the inhibitory stage of our model, while the later P3b reflects the attentional reorienting (or any other type of “active engagement”) to the surprising event. In real-world situations, both factors could independently contribute to surprise-related distraction, depending on the nature of the current cognitive representations. Future studies could aim to test this theory by differentially relating the different signatures to attentional reorienting, active engagement with the unexpected event (e.g., incidental recall), and automatic surprise-related inhibition, respectively. Furthermore, they could also test whether other types of cognition are differentially affected by separate surprise-related processes.
One caveat to our interpretation could be the modality of the surprising events. In both cases (Wessel et al. 2016 and the current study), the surprising events were auditory in nature. It is conceivable that maintenance of verbal WM involves an auditory component (since the information is verbalizable), which could be absent for visuospatial WM. Since the surprising event happens in the auditory domain, it could therefore interfere with verbal WM, while visuospatial WM might be unaffected. However, we believe that this explanation is both contradicted by our data, and not in line with the neural activity elicited by surprising events. First, our current dataset showed that surprising tones in fact interfered with visuospatial WM: Surprising trials with incorrect responses were associated with stronger degrees of surprise compared with correct responses. Second, studies of the neural concomitants of unexpected events have shown that surprising events engage overlapping brain networks, regardless of their sensory domain (Downar et al. 2002), which include the brain areas underlying the inhibitory mechanism (Wessel and Aron).

For future studies, it would be informative to investigate the effect of surprising events on electrophysiological activity that reflects the active maintenance of the visuospatial WM contents, that is, the CDA (Vogel and Machizawa 2004). The prediction would be that the CDA amplitude on trials with surprising events would be reduced after the surprising tone—specifically, to an extent that should be in direct relation to the amplitude of the P3b. However, to do this, one would have to separate the trials in which a WM error was made because of the surprising event from the trials in which a WM error was made because of other factors (e.g., inaccurate encoding, or any other factors that cause errors to be made even on trials without surprising tones). In our dataset, the individual trial-to-trial signal-to-noise ratio of the CDA was insufficient to warrant such an analysis. However, future experiments could aim at different ways to separate such trials to collect a “pure” sample of trials in which the interruption can be clearly attributed to the occurrence of the surprising tone.

In summary, we have investigated the activity of an inhibitory neural mechanism that is active following surprising events and scales with the degree of surprise. However, unlike prior studies with verbal WM, our study showed that this mechanism does not affect ongoing representations in visuospatial WM. In turn, we have shown that an independent mechanism, indexed by the centro-parietal P3b ERP, is related to the interruption of visuospatial WM: When surprising events disrupted ongoing visuospatial WM, the P3b was increased. Hence, this study provides evidence for a multistage model of distraction, in which different types of cognitive representations can get interrupted by different (though likely cascaded or otherwise related) neural processes.

Funding
J.R.W. is supported by the Aging Mind and Brain Foundation at the University of Iowa.

Notes
The author would like to thank Hailey Billings, Kylie Dolan, and Cailey Parker for help with EEG data collection, Edward Ester for helpful discussions, and Nandakumar Narayanan, Eliot Hazeltine, and 2 anonymous reviewers for feedback on the final manuscript. Conflict of Interest: None declared.

References
Debener S, Makeig S, Delorme A, Engel AK. 2005a. What is novel in the novelty oddball paradigm? Functional significance of...


