Is the Contralateral Delay Activity (CDA) a robust neural correlate for Visual Working Memory (VWM) tasks? A reproducibility study

Yannick Roy | Jocelyn Faubert

Abstract

Visual working memory (VWM) allows us to actively store, update, and manipulate visual information surrounding us. While the underlying neural mechanisms of VWM remain unclear, contralateral delay activity (CDA), a sustained negativity over the hemisphere contralateral to the positions of visual items to be remembered, is often used to study VWM. To investigate if the CDA is a robust neural correlate for VWM tasks, we reproduced eight CDA-related studies with a publicly accessible EEG data set. We used the raw EEG data from these eight studies and analyzed all of them with the same basic pipeline to extract CDA. We were able to reproduce the results from all the studies and show that with a basic automated EEG pipeline we can extract a clear CDA signal. We share insights from the trends observed across the studies and raise some questions about the CDA decay and the CDA during the recall phase, which surprisingly, none of the eight studies did address. Finally, we also provide reproducibility recommendations based on our experience and challenges in reproducing these studies.

Keywords

CDA, contralateral delay activity, EEG, review, VWM

1 | INTRODUCTION

Visual working memory (VWM) allows us to actively store, update, and manipulate visual information surrounding us. Acting as a mental buffer for visual information, VWM has been an active area of research for several decades. While behavioral studies have shown that working memory (WM), of which VWM is a subset, has a limited capacity of only a few items, usually ranging between 3 to 5 (Bays & Husain, 2008; Cowan, 2001), the underlying neural mechanisms remain vague. One ongoing challenge in the field is dissociating WM from attention (Awh et al., 2006; Cowan, 1998; Oberauer, 2002; Olivers et al., 2011). There is no clear consensus yet as how separated or intertwined these two mechanisms really are.

Given the crucial role of VWM in our everyday life, much prior work has tried to understand the neural correlates underlying its functioning via electroencephalogram (EEG). One such VWM neural measurement being studied is the contralateral delay activity (CDA) (McCollough et al., 2007; Vogel & Machizawa, 2004). CDA is a sustained negativity over the hemisphere contralateral...
to the positions of the items to be remembered. A prevailing view of CDA is that it is modulated by the number of items held in WM reaching a plateau at around three or four items. The CDA has been shown to be linked with the number of items held in WM (Luria et al., 2016; Unsworth et al., 2015; Vogel & Machizawa, 2004).

In order to study CDA, the most common task is a change detection task. The sequence of such task typically looks something like the following: the participant is cued with an array of items, varying in numbers but usually between one and eight, balanced on both sides of the visual field. The items then disappear, forcing the participant to hold relevant information in mind and after a short period of time, usually one or two seconds, the participant is quizzed on the visual representation held in WM. For example, a new array of items can be presented and the participant is asked if there has been a change versus the initial items. The participant will answer with a keyboard or a mouse and the results will be logged with different levels of granularity depending on what the study is interested in. There are many variants of such tasks.

Many well-known event-related potentials (ERPs) such as the P300 and N270, benefit from a large volume of studies and replications and are well understood (Fazel-Rezai et al., 2012; Luck, 2014; Polich & Kok, 1995; Rossion et al., 2000). CDA, however, does not benefit from such volume of evidence yet. Our primary goal with this reproducibility study was to answer the following question: Is the Contralateral Delay Activity (CDA) a robust and consistent neural correlate for Visual Working Memory (VWM) tasks? We wanted to know if CDA is a consistent measure across different tasks, subjects, and EEG recording devices, or if it requires a lot of manual cleaning and handcrafting of the data to obtain it. To investigate the robustness of the CDA EEG signal, we looked for CDA-related EEG data sets available online and tried to reproduce their results using a basic independent automated pipeline with no human intervention on the data to extract the CDA. Given that robustness can be interpreted in different ways, we should mention that we use the same definition as in the Framework for Open and Reproducible Research Teaching (FORRT): The persistence of support for a hypothesis under perturbations of the methodological/analytical pipeline. In other words, applying different methods/analysis pipelines to examine if the same conclusion is supported under different analytical conditions (Parsons et al., 2022).

Lastly, before we dive in, we should also define the terms reproducibility and replicability given that researchers from different fields, and even from within the same field, often use them interchangeably. Here we will use the same definitions as in the Recommendations to Funding Agencies for Supporting Reproducible Research by the American Statistical Association (Broman et al., 2017):

- Reproducibility: A study is reproducible if you can take the original data and the computer code used to analyze the data and reproduce all of the numerical findings from the study. This may initially sound like a trivial task but experience has shown that it is not always easy to achieve this seemingly minimal standard.
- Replicability: This is the act of repeating an entire study, independently of the original investigator without the use of original data (but generally using the same methods).

Simply put, one can replicate a study or an effect (outcome of a study) but reproduce results (data analyses). A third term, repeatability, is also used although less often, referring to the same group repeating the same experiment with the same analysis and obtaining the same results.

This current work focuses on reproducibility and not on replicability nor repeatability as we did not collect new data but only used publicly available data sets. Moreover, given the nature of the studies and the data we are handling, we could be even more specific using the terminology computational reproducibility, as defined in the FORRT (Parsons et al., 2022). This computational reproducibility study also intersects with a review given that we extract common trends among studies and explore them from a computational point of view. Note that we initially reproduced the result either with the original code or with a re-written version of it before using the same pipeline for all studies to assess the robustness.

Other fields such as artificial intelligence (AI) have benefited tremendously from good reproducibility frameworks, standards and overall culture. The accelerated pace of innovation and breakthroughs in AI were mainly enabled by accessibility of both data and code. These best practices of sharing both data and code in a reproducible manner are not, unfortunately, the default behavior in psycho- and neuro-related fields. Here, we share how a few groups of researchers have made their data and code available and we hope to inspire others to follow that path.

2 | METHOD

After looking through CDA-related literature, searching for available data sets and asking (i.e., emailing) a few researchers in the field if they were aware of open access CDA-related EEG data sets, we ended up with eight recent CDA-related studies published between 2018 and 2021 with different task paradigms (see Table 1). We do not claim that the list of studies included in this review represents an exhaustive list of CDA papers with available EEG data sets, however, we believe that these eight data sets represent a good sample of the CDA literature as they
used different VWM tasks exploring things such as different set sizes (from 1 to 6 targets), adding new targets after the initial array, retro-cueing, adding targets bilaterally, adding interruptions, halving the targets to create more items to track, and all that using different shapes and colors as stimuli across studies. The search of CDA-related literature was done on Google Scholar using the following search terms in different orders and combinations: CDA, Contralateral Delay Activity, and EEG. From the bibliography section of these articles, a few additional relevant studies were added to the list. Open Science Framework (OSF) with the keywords previously listed was also used to search for additional studies and data sets. It is important to note that six out of the eight studies included here have Edward Awh and Edward Vogel as co-authors, highlighting a lack of independent studies on CDA with openly available EEG data.

In this section, we detail the studies we have reproduced and provide the resulting CDA signal figure, which matches the CDA figure from the original study. It is important to note that some of the reproduced studies were also looking at different neural correlates, however, for simplicity and readability we focus only on the CDA relevant part of the study. In our analysis, we used all the available data, but did not include any data files that were clearly marked as not being used (for various reasons). All the studies were reproduced using MNE-Python (Gramfort et al., 2013) despite the original code of most studies being in MATLAB. All figures in this section were generated using the following pipeline: (1) load raw data, (2) re-reference and downsample based on the original study's preprocessing steps, (3) filter the data between 1 and 30 Hz, (4) epoch the data, (5) automated removal of bad trials (i.e., epochs) and interpolation of bad channels via autoreject (Jas et al., 2017), (6) obtain each subject’s CDA for each condition by subtracting the averaged ipsilateral electrodes to the averaged contralateral electrodes (i.e., contra minus ipsi), (7) average the CDA for each condition across subjects.

Table 2 shows the high-level information of the data sets to provide the reader an idea about the number of subjects and trials for each study.

### 2.1 Feldmann-Wüstefeld (2021)

In *Neural measures of working memory in a bilateral change detection task* (Feldmann-Wüstefeld, 2021), Feldmann-Wüstefeld and colleagues used a novel change detection task in which both the CDA and the negative slow wave (NSW) can be measured at the same time. They presented memory items bilaterally with different set sizes in both hemifields inducing an imbalance or “net load” as they called it. Their results showed that the NSW increased with set size, whereas the CDA increased with net load. There were three different set sizes participants had to remember: two, four, or six targets. In Figure 1, we can see the five combinations of targets they used (2:0, 3:1, 4:2, 4:0, 5:1). With their nomenclature 2:0 means 2 targets in one hemifield and 0 target in the other hemifield for a net load of 2. 4:2 represents 6 targets total, 4 in one hemifield and 2 in the other for a net load of 2. As we can see, the highest CDA is obtained with 4:0 for a net load of 4. Interestingly, the 5:1 condition shows a higher CDA than 3:1 and 4:2 however lower than 2:0 indicating that having targets in both hemifields reduces the overall CDA amplitude. On the graph, t = 0 s is when the memory display appeared on the screen with the targets, then they stayed visible for 500 ms after which the participant had to remember the targets for 1 second before the probe display appeared for the participant to provide their answer as to confirm if the item in the probe display was indeed part of the targets or not.

| Year | First Author | Title |
|------|--------------|-------|
| 2021 | Tobias Feldmann-Wüstefeld | Neural measures of working memory in a bilateral change detection task |
| 2020 | Nicole Hakim | Perturbing neural representations of working memory with task-irrelevant interruption |
| 2020 | Mario Villena-Gonzalez | Data from brain activity during visual working memory replicates the correlation between contralateral delay activity and memory capacity |
| 2019 | Haley Balaban | Neural evidence for an object-based pointer system underlying working memory |
| 2019 | Eren Gunseli | EEG dynamics reveal a dissociation between storage and selective attention within working memory |
| 2019 | Nicole Hakim | Dissecting the Neural Focus of Attention Reveals Distinct Processes for Spatial Attention and Object-Based Storage in Visual Working Memory |
| 2018 | Kristen Adam | Contralateral delay activity tracks fluctuations in working memory performance |
| 2018 | Tobias Feldmann-Wüstefeld | Contralateral Delay Activity Indexes Working Memory Storage, not the Current Focus of Spatial Attention |
2.2 Hakim et al. (2020)

In Perturbing neural representations of working memory with task-irrelevant interruption (Hakim et al., 2020), Hakim and colleagues investigated the impact of task-irrelevant interruptions on neural representations of working memory across two experiments looking at both the CDA and lateralized alpha power. What they found is that after interruption, the CDA amplitude momentarily sustained but was gone by the end of the trial. On the other hand, lateralized alpha power, which has been used as an effective tool for discerning which visual hemifield is being attended, was immediately influenced by the interrupters but recovered by the end of the trial. Hakim and colleagues suggested that dissociable neural processes contribute to the maintenance of working memory information and that brief irrelevant onsets disrupt two distinct online aspects of working memory and also that task-irrelevant interruption could motivate the transfer of information from active to passive storage, explaining the reason why the CDA drops significantly after interruptions, even on trials with good performance (i.e., the participant remembered the targets correctly). In their 2nd experiment, they go further and test if the expectation of interruption changes the CDA. The full 1.65s epoch is displayed but not the response period which took place after 1.65s (see Figure 2). The CDA was obtained by using only PO7 and PO8 electrodes.

2.3 Balaban et al. (2019)

In Neural evidence for an object-based pointer system underlying working memory (Balaban et al., 2019), Balaban and colleagues argued that to update our representation of the environment, our VWM depends on a pointer system such that each representation is stably and uniquely mapped to a specific stimulus. Therefore, without these pointers, our VWM representations are inaccessible. Via three experiments, they examined whether the pointers are allocated in an object-based, featural, or spatial manner. Their results showed that the separation of an object in two halves invalidated the pointers. It happened in a shape task, where the separation changed both the objects and the task-relevant features, but also in a color task, where the separation destroyed the objects while leaving the task-relevant features intact. They suggested that objects, and not task-relevant features, underlie the pointer system. Two of their three experiments are displayed in Figures 3 and 4, while the third experiment is available in the supplementary material to reduce the length of this manuscript and enhance readability.
For Experiment 1, $t = 0$ s is when the memory display appeared on the screen with the moving objects which either separated in halves after 400 ms or either continued moving as a whole for another 600 ms after which they stopped moving for 300 ms and disappeared. After 900 ms of retention with an empty screen displaying only a fixation cross in the middle, the participant had to give an answer as if the objects being displayed are the same or different than the initial ones. In one condition the shapes were the relevant features (i.e., have some of the shapes changed?) and in a second condition the colors were the relevant features (i.e., have some of the colors changed?). On the graph, the cognitive impact of the separation that happened for half the trials (at 400 ms) is clearly visible in the CDA signal around 200 ms as highlighted by the grayed region. The full 2.2 s epoch is displayed on Figure 3 but not the response period which took place after 2.2 s. The CDA was obtained by using only PO7-PO8, P7-P8 and PO3-PO4 electrode pairs.

For Experiment 2 (Figure 4), the epochs were shorter in time and only the colors were the relevant features. The targets were all moving squares that could either separate after 400 ms or continue moving as a whole. Some trials had two targets and some trials had four targets. In both experiments we see that after the separation, the CDA amplitude increases, as the number of targets to track has now increased.

2.4 | Gunseli et al. (2019)

In EEG dynamics reveal a dissociation between storage and selective attention within working memory (Balaban et al., 2019) Gunseli and colleagues tested the hypothesis that within WM, selectively attending to an item and stopping storing other items are independent mechanisms. In order to make participant drop items from WM, they used a retro-cue to indicate which of the items were the target(s). As opposed to identifying the targets at the beginning of the trial like in most WM studies, here they showed 3 items (bars with different orientations) and then, 1 s later they showed a retro-cue indicating which item is most likely to be tested, or probed, after the retention phase. Their hypothesis was that if the retro-cue is reliable, the participant would drop the other item(s) creating an imbalance between hemispheres and therefore increasing the CDA signal. Whereas if the retro-cue is not reliable the participants would not drop the item(s) and
therefore resulting in a smaller CDA signal. On Figure 5, we see that indeed the CDA is of higher amplitude for trials where the retro-cue is valid 80% of the time in comparison to the other condition where the retro-cue is valid only 50% of the time. The right graph was generated using their MATLAB preprocessed data files (.mat) and the left graph shows our reproduced version from their raw EEG data, with a lowpass filter at 6 Hz, as they mention in their paper to leave the alpha band out of the CDA signal. Unfortunately, the preprocessing code used to generate the MATLAB preprocessed files was not available and it seems like these files benefited from additional manual cleaning because when we use a higher filter (e.g., 20 Hz or 30 Hz) with the same pipeline as other studies, both signals end up with a similar amplitude and the effect is not visible anymore because of a high variability on both signals (50% vs 80%). In their paper they mentioned that using a higher filter (e.g., 40 Hz) did not change the results of the statistical analysis (relative to the 6 Hz filter). We did not perform any statistical analysis, however, we were only able to obtain a visible CDA difference when plotting the grand average with heavy filtering (6 Hz), which helps reduce the variability. Only the PO7-PO8, P7-P8, and O1-O2 electrode pairs were used to generate the CDA.

2.5 | Villena-Gonzalez et al. (2020)

In *Data from brain activity during visual working memory replicates the correlation between contralateral delay activity and memory capacity* (Villena-González et al., 2020), Villena-Gonzalez and colleagues replicated the results from Vogel and Machizawa (2004) showing that the amplitude of the CDA correlates with the number of items held in WM using a change detection task with set sizes of one, two, and four. Moreover, they also looked at the individual performances and showed that participants with higher WM capacity (i.e., better performance on the task) also had a higher CDA amplitude. Figure 6 shows a clear difference between one target and two or four targets, however, we see similar amplitudes for two and four targets. Unfortunately, we were not able to reproduce their results showing a clear difference of amplitude between two and four targets. In their paper, they had a significant higher amplitude for four targets, which we failed to reproduce after a few attempts at with different automated preprocessing pipelines. Our version does not invalidate their conclusion as we also see a higher CDA amplitude the larger the set size, however the effect between two and four is not as strong as in their findings. This might indicate that the results could have benefited from extra manual cleaning. The following electrode pairs were used to obtain the CDA: TP7-TP8, CP5-CP6, CP3-CP4, CP1-CP2, P1-P2, P3-P4, P5-P6, P7-P8', P9-P10, PO7-PO8, PO3-PO4, and O1-O2.

2.6 | Hakim et al. (2019)

In *Dissecting the Neural Focus of Attention Reveals Distinct Processes for Spatial Attention and Object-Based Storage in Visual Working Memory* (Hakim et al., 2019), Hakim and colleagues showed that the focus of attention in WM is not a monolithic construct but rather involves at least two neurally separable processes: (a) attention to regions...
in space and (b) representations of objects that occupy the attended regions. On Figure 7, the CDA is clearly visible, showing a slightly higher amplitude for a set size of 4 targets versus 2. The full 1.45 s epoch of the change detection task is displayed but excludes the response period which took place right after. It is important to note that the graph represents the aggregation of 4 sub-experiments with slight variations on the task. The CDA was obtained by using only O1-O2, PO3-PO4, PO7-PO8, P3-P4, and P7-P8 electrode pairs.

2.7 | Feldmann-Wüstefeld et al. (2018)

In *Contralateral Delay Activity Indexes Working Memory Storage, not the Current Focus of Spatial Attention* (Feldmann-Wüstefeld et al., 2018), Feldmann-Wüstefeld and colleagues seek to explore the recent hypothesis from Berggren and Eimer suggesting that the CDA tracks the current focus of spatial attention as opposed to working memory storage (Berggren & Eimer, 2016). Figure 8 shows the CDA results of both their experiments in which they displayed four targets among distractors via two sequential memory displays in a change detection task. The first set of targets is shown at t = 0 s for 200 ms then disappears for 500 ms after which a second set of targets and distractors appear for 200 ms and then disappear for another 500 ms after which the probe display is shown for the participant to confirm if the items currently displayed are the same as the targets. The whole 1.4 s epoch is displayed on Figure 8, leaving the response period out of that figure. A total of four targets were always shown to the participant. The first memory display (i.e., the first set of targets and distractors to be shown) could either have 1, 2, or 3 targets and the second memory display, 700 ms later, could add 3, 2, or 1 targets for a combined total of 4. The targets could be either added in the same hemifield or the different (i.e., opposite) hemifield. The expected result is a higher CDA when the targets are added in the same hemifield as this will increase the “net load” (term they will later use in their 2020 paper) and a lower CDA amplitude when added in the opposite hemifield because it would then decrease the net load. Experiment 1, on the upper row, shows such expected results somewhat perfectly as on the top left graph we see all three CDA signals reaching the same amplitude after both sets of targets are added and equals to four items in the same hemifield. As

---

**FIGURE 7** Reproduced results from Hakim et al. (2019), using our simple comparative pipeline.

**FIGURE 8** Reproduced results from Feldmann-Wüstefeld et al. (2018), using our simple comparative pipeline.
expected, the top right graph shows a slight decrease on the CDA amplitude for $3 + 1$ diff but more interestingly, a CDA of pretty much zero for the $2 + 2$ diff condition where two targets were shown in both hemifields canceling out the CDA signal. For the $1 + 3$ diff condition, we see the CDA flipping side after the new set of three targets is shown on the opposite side. Experiment 2 was similar to Experiment 1 except for the probe display when the participant provides the answer. In Experiment 1 the probe display showed the items at the same spatial location that they were displayed in either the memory display 1 (at $t = 0$ ms) or 2 (at $t = 700$ ms). However, in Experiment 2 the probe display was modeled after Berggren and Eimer (2016) experiment such that the items were spatially translated and interleaved. The authors suggested that given that the retention period of Experiment 1 and 2 were identical (i.e., the 1.4 s displayed on the graph) and that only the probe display was different, the differences observed between both experiment can only be explained by different memory strategies. They therefore suggested that because the mental representation in Experiment 2 is more difficult as the items are not displayed “as is” but translated on the probe display, the participants most likely transferred items of display 1 (i.e., first set of targets) from working memory (WM) into long-term memory (LTM) therefore explaining why in Experiment 2 the CDA seems more affected by the second set of items rather than equally affected from the first and second set of targets as in Experiment 1. A clear example of that difference between experiments is the $2 + 2$ diff condition where the CDA is pretty much zero in Experiment 1 but biased toward the second set of targets in Experiment 2. Only the PO7-PO8 electrode pair was used for the CDA signal.

### 2.8 Adam et al. (2018)

In *Contralateral delay activity tracks fluctuations in working memory performance* (Adam et al., 2018), Adam and colleagues looked at the relationship between the CDA amplitude and working memory performance. Their hypothesis was that if working memory failures are due to decision-based errors and retrieval failures, CDA amplitude would not differentiate good and poor performance trials when load is held constant. If failures arise during storage (i.e., retention phase), then CDA amplitude should track both working memory load and trial-by-trial performance. Their Experiment 1, shown on Figure 9, showed that the CDA amplitude increased with set size but plateaued at three targets showing a similar amplitude for three or six targets. In their Experiment 2, they kept the set size at 6 and compared the good trials (accuracy of 3 or more targets identified correctly out of 6) versus bad trials (accuracy of 2 targets or less identified correctly). Figure 9 (on the right) shows that indeed the amplitude of the CDA correlated with the performance. The O1-O2, OL-OR, P3-P4, PO3-PO4, T5-T6 electrode pairs were used for the CDA.

### 3 RESULTS

Before analyzing the CDA and drawing conclusions on the underlying cognitive functions, we should understand how to best obtain a clean CDA signal in the first place. In this section, we will first discuss what channels and reference(s) various groups have used to obtained their CDA. Second we will discuss some trends we have observed across the reviewed studies.

Note that we did not look at any eye-tracking data and assumed that the subjects respected the instructions of fixating the middle of the screen. Many studies had the eye-tracking data available, however we did not use it nor have we excluded any trials based on eye-movement. This does certainly impact the results we obtained when compared to the original curves of the authors in their paper. Given that such data were not available for all data sets, we decided to not consider them at all.
3.1 EEG channels

The contralateral delay activity (CDA), as its name implies, is a difference in the activity between the left and right hemisphere. The signal is obtained by subtracting one or more contralateral channels to equivalent ipsilateral channel(s) to the attended side. Unsworth et al., mentioned in their 2015 paper that it is now standard procedure for measuring the CDA to use posterior parietal, lateral occipital, and posterior temporal electrodes (PO3, PO4, T5, T6, OL, and OR), citing the work of Vogel and Machizawa (2004) and Vogel et al. (2005). Looking at the electrodes used in the reproduced studies from 2018 to 2020, there seems to be a slight change toward favoring PO7/PO8 as the best electrode pair. Table 3 shows the selected channels used by the reviewed studies to obtain the CDA. Adam et al. (2018), used the recommendations from Unsworth, 2015, without the PO7/8 pair, while all the others included at least the PO7/8 pair. Feldmann-Wüstefeld et al. (2018), Feldmann-Wüstefeld (2021), and Hakim et al. (2020), used only that pair for their CDA results and did not average with any other pairs. Interestingly, in her 2019 study, Hakim had use PO7/8 but also P7/8, P3/4, PO3/4, and O1/2, which were not included in the 2020 analysis. Villena-Gonzalez et al. (2020) used the most electrode pairs of the reviewed studies, including parietal, occipital, temporal and central sites. It is worth noting that the PO7/8 is not an electrode pair in the standard 32-electrode 10–20 placement. This most likely explains why Adam et al. (2018) did not include this pair given that they use a system with 20 electrodes with 10–20 locations.

In order to better understand the signal shape and amplitude coming from the various electrode pairs, we looked at each pair for each condition of each study. Regardless of what the study actually used for their analysis, we used all channel pairs available in the raw data. We included here only the breakdown for Balaban (2019) and Villena-Gonzalez et al. (2020) on Figure 10a,b, respectively. All the other studies followed the trend of Balaban (2019) showing a stronger CDA from the parietal sites with PO7/8 being the best candidate (i.e., strongest CDA amplitude) across studies. Villena-Gonzalez et al. (2020), is the only one showing different results with a very strong frontal lateralized activity (see Figure 10b). We initially thought these results were odd, until we analyzed our own data recorded for another project while writing this review, for which we also observed frontal activity being way stronger than parietal activity. Our task is a 3D-MOT task (see Faubert, 2013).

To confirm our empirical estimation of PO7/8 being the best electrode pair for CDA, we calculated the effect size for each available pair of electrodes by comparing contralateral versus ipsilateral activity for each study. All the data sets had the PO7/8 pair available but Adam et al. (2018). Aside from Villena-Gonzalez et al. (2020), the largest effect size was always observed from a parietal electrode pair. PO7/8, PO3/4, and P7/8 dominated the top three electrode pairs in most studies. PO7/8 had the largest or second largest effect size for 5 of the 8 studies. The breakdown is available on the repository online.

Finally, the approach for averaging the channel pairs also varied. Some groups subtracted the pairs first and then averaged across pairs, while others averaged one side first and then subtracted to the average on other side. The order in which the averaging and subtracting happens does not really matter if no other operation is done on the signal and if they are all referenced to the same reference. For example, if we are trying to use the pairs PO7-PO8, P7-P8, and PO3-PO4, we could either start by subtracting each pair individually (e.g., signal from PO8 minus signal from PO7) and then averaging the 3 signals obtained, giving us the resulting CDA or we could average one side together (PO7, P7, and PO3) and subtract it from the average signal from the other side (PO8, P8, and PO4). Both approaches would give the same CDA for a given trial.

| Study                        | Channels                                      |
|------------------------------|-----------------------------------------------|
| Feldmann-Wüstefeld et al. (2018) | PO7/8                                         |
| Feldmann-Wüstefeld (2021)     | PO7/8                                         |
| Hakim et al. (2020)           | PO7/8                                         |
| Gunseli et al. (2019)         | PO7/8, P7/8, O1/2                             |
| Balaban (2019)                | O1/2, OL/R, P3/4, PO3/4, T5/6                 |
| Adam et al. (2018)            | O1/2, PO3/4, PO7/8, P7/8, P3/4                |
| Hakim et al. (2019)           | O1/2, PO3/4, PO7/8, P3/4, P7/8                |
| Villena-Gonzalez et al. (2020)| TP7/8, CP5/6, CP3/4, CP1/2, P1/2, P3/4, P5/6, P7/8, P9/10, PO7/8, PO3/4, O1/2 |
3.2 | EEG reference(s)

The EEG signal being an electric potential difference between two electrodes, namely the electrode of interest and the reference, it is no surprise that the reference plays an important role in EEG studies. Changing the reference can drastically change the signal obtained and therefore influence the conclusion of a study. The ideal reference would be a neutral point with no electrical activity to which we could measure a difference of potential being only the activity of interest. Unfortunately, no such point exists on the body, leaving the problem of finding a good reference unsolved or open to interpretation. Best practices in EEG studies include (1) using the average of left and right mastoids or (2) to re-reference the signal offline to the average of all channels or (3) using one or multiple electrode(s) on the midline such as Cz or (4) using a mathematical reference such as the reference electrode standardization technique (REST) or a Laplacian approach (Yao et al., 2019). Here in the reviewed studies, all of them used to the average of left and right mastoids except for Feldmann-Wusterfeld 2018 and 2021 which used the average of all electrodes.

Given that CDA is a difference between left and right hemispheres, the reference is not as important as in most evoked-related potential (ERP) studies. The choice of reference will impact the visual inspection and the cleaning of the data, however, when it comes to obtaining the CDA itself, because we are subtracting one electrode to the other, the reference gets somewhat canceled and therefore does not impact the resulting CDA signal as much. For example, if the contralateral electrode is PO8(-Ref) and the ipsilateral electrode is PO7(-Ref), then CDA (contralateral - ipsilateral) is (PO8 - Ref) - (PO7 - Ref) which is equivalent to PO8 - PO7 as the Ref cancels out. It is therefore still important to consider the reference for cleaning the signal but to keep in mind that the CDA itself is not as affected as much by the reference as typical, non-latelarized, ERPs. This actually makes the CDA signal even more robust to noise, assuming that most EEG noise would be visible on both channels, and be canceled out during the subtraction. A noise that would survive that subtraction would be a noise seen only by one electrode or one ‘side’ of the head.

3.3 | CDA decay

If the amplitude of the CDA correlates with the number of items being tracked, one could expect the amplitude to remain somewhat stable for the whole retention or tracking duration. However, in all the replicated studies we can observe the amplitude of the CDA declining way before the end of the retention phase even when the participants did not lose track of the item(s).

In later section 3.6 we look at the CDA when subjects lost track of one of more item(s) (i.e., bad trials) but all figures previously showed with the reproduced results and discussed in the methods section were generated from good performances only, meaning that the subjects did remember all the items for the whole duration of the trial and provided good answers at the end. As we can see on Figures 1, 3, 4, 5, and 7 the CDA is starting to decrease way before the end of the trials. In Hakim (2020), they added interruptions trying to interfere with WM processes and CDA indeed dropped significantly shortly after such interruptions. And yet, despite the important drop in amplitude the participants still provided correct answers. While none of the eight studies addressed directly what seems to be
a natural CDA decay, some suggested (e.g., Hakim., 2020; Feldmann-Wüstefeld et al., 2018) that the items could be transferred from short term memory to long term memory therefore affecting the CDA amplitude observed here.

### 3.4 Recall

One interesting phenomenon that most studies did not mention in their paper is the recall, or response, period (i.e., when the participant is providing the answer). Interestingly, in most studies, we are observing a re-increase in the amplitude of the EEG signal, sharing similarities with the CDA observed during the identification and retention (or tracking) periods. One potential reason most studies do not look at that period is because eye movements are generally not controlled in this phase, introducing artifacts in the EEG signal. Moreover, the induced artifacts are most likely lateralized and not normally distributed as the participants might be fixating and moving their eyes to the side of interest, complicating the dissociation between eye movements and cognitive processes in the EEG signal.

Figures 11–14 show the CDA graphs but with a longer epoch this time, including two seconds after the end of the retention phase. The right-most dashed line represents when the participant was asked to provide an answer. The figures all show a re-increase in the signal amplitude also followed by a decay of the signal as discussed in the last section. Only two studies did not show such re-increase of a CDA-like signal during recall: Villena-Gonzalez et al. (2020) and Gunseli et al. (2019). The graphs are available online in our repository with the analysis.

In order to investigate if the re-increase of the signal is mainly driven by eye-movements or by cognitive processes, we looked at different channel pairs with the assumption that the frontal pairs, especially the ones further from the midline, would be the most affected by eye-movements and such artifacts would then leak to central and parietal channels. We show the breakdown of channel pairs for Balaban et al. (2019) Experiment 1 on Figure 15, Feldmann-Wüstefeld et al. (2018) Experiment 1 on Figure 16, and Hakim et al. (2019) on Figure 17. The different channel pairs available in the data set, not only the one used for the CDA, were plotted for the whole trial and extending over the response period.

**FIGURE 11** Feldmann-Wüstefeld (2021) - Recall/Probe. Same CDA as on Figure 1 but with a longer epoch, showing a CDA re-increase during recall (t > 1.5 s).

**FIGURE 12** Feldmann-Wüstefeld et al. (2018) - Recall/Probe. Same CDA as on Figure 8 but with a longer epoch, showing a signal amplitude re-increase during recall (t > 1.5 s) for Experiment 1.
period. What we observe is that indeed there seems to be a mixture of cognitive processes and eye-movement artifacts. For Balaban et al. (2019), we see the re-increase in amplitude in the same parietal channels while the frontal channels remain unaffected during the early part of recall, suggesting that the underlying activity is not only from eye-movements but similar to the one during the retention period eliciting the CDA. For Feldmann-Wüstemfeld et al. (2018), on the 2nd and 3rd graph of Figure 16, we see a big change in amplitude for the F7-F8 pair during recall, which is most likely driven by eye-movements rather than cognitive processes. However,
on the 1st graph, for the $1 + 3$ same condition, the F7-F8 pair is not being impacted as much and have a similar amplitude as the parietal channels, suggesting that the activity observed from the parietal channels is mostly cognitive while being slightly affected by eye-movement artifacts. It is not very likely that the eye-movement artifacts would affect more the parietal channels than the frontal ones. In Hakim et al. (2019) on Figure 17, we observe a very strong increase of amplitude on the F7-F8 pair suggesting that the response period is highly contaminated by eye-movement artifacts making it quite difficult to draw any conclusion on the re-increase in amplitude in the parietal channels as it could simply be driven by the eye-movement artifacts.

Here we provide some exploratory graphs, however, it is too early to draw any strong conclusions on the nature of the CDA-like amplitude re-increase during recall, without a deeper analysis of eye-tracking and EOG data which was outside of the scope of this reproducibility study.

3.5 CDA amplitude versus number of items

Looking at the reproduced results, it seems fair to conclude that indeed the amplitude of the CDA correlates with the number of items tracked by the participants up to a plateau of 3 to 4 items. This correlation and plateau has been discussed in previous CDA studies (e.g., Luria et al., 2016; Unsworth et al., 2015; Vogel & Machizawa, 2004). Figure 6 from Villena-Gonzalez et al. (2020), data shows a clear difference in CDA amplitude between one, two, and four items. Figure 9 from Adam et al. (2018), data shows a clear difference between one and three or one and six but a very similar amplitude for three and six items, aligned with some sort of CDA amplitude plateau around three or four items. Figure 7 from Hakim et al. (2019), data shows a small difference between two and four items.

The work from Feldmann-Wüstefeld et al. (2018) (Figure 8) shows that even after the initial identification and tracking phases, the CDA can be increased by adding...
items on the same side or decreased by adding items on the opposite side. Similarly, the work from Balaban et al. (2019) (Figures 3 and 4) shows that when a whole item splits in two separate parts that need to be tracked, the CDA increases. Feldmann-Wüstefeld (2021) (Figure 1) used a bilateral task with different loads on both sides creating a net load of either two or four items and what their work shows is that different combinations of net load of two or four produce different CDA amplitudes. The more objects on the opposite side, the lower the CDA amplitude despite the same net load. However, the results are aligned with the CDA theory saying that the more objects the higher the amplitude. The 4:0 condition shows a higher amplitude than the 2:0 condition. The 5:1 shows a higher amplitude than the 3:1.  

### 3.6 CDA amplitude versus individual performance

If the CDA amplitude correlates with the number of items held in memory, the CDA should be indicative of the performance of a specific trial. This obviously holds true only if the working memory failures occur during the retention phase and not during the recall phase to provide the answer. If such memory fault occurs during the initial identification phase, the CDA would not drop per se but instead reach a lower amplitude than the expected peak amplitude should the participant have tracked the correct amount of items. However, if the participant mistakenly identifies the wrong target (for example in a multiple-object tracking task) and confidently holds and tracks the correct number of items, despite some of them being wrong, this kind of mistake would result in the same CDA amplitude as if the right items were kept in memory. While the current state of CDA literature lacks the trial-by-trial analysis to evaluate the role CDA plays in performance, several studies have looked at the individual differences in CDA amplitude versus working memory performance or capacity. For example, Unsworth and colleagues in 2015 (Unsworth et al., 2015) replicated the work from Vogel and Machizawa (2004) showing a correlation between CDA and performance on a change detection task where high working memory individuals had larger CDA ($r = 0.30$; Unsworth, 2015) and concluded their study
saying that CDA is a reliable and valid individual measure of working memory that predicts behavioral performance on visual working memory tasks. Of the studies reviewed here, only three looked at the direct correlation between individual working memory performance and CDA amplitude. Most studies looked at the relationship between task conditions and performance, as well as the relationship between task conditions and CDA amplitude; however, only three looked at the direct correlation between performance and CDA amplitude (Adam et al., 2018; Feldmann-Wüstefeld, 2021; Villena Gonzalez et al., 2020). Among these three studies, working memory performance, or capacity, was evaluated differently based on the task and the granularity in the answers provided by the participants.

In Adam et al. (2018), they used a whole report task with a set size of six items and noted the participants’ answer accuracy between zero and six targets identified correctly. With this granularity, they looked at the correlation between CDA amplitude and accuracy (from 0/6 to 6/6, steps of 1) and showed individual differences in working memory performance with a correlation of $r = -0.26$ and $p = 0.028$. They also noted that the magnitude of the effect is relatively small, but consistent with previously observed effects in the literature, citing Unsworth (2015) work. With their EEG and behavioral data we were able to reproduce a similar result with a slightly higher and more significant correlation ($r = -0.34$, $p = 0.001$). It is important to note that the classes are quite unbalanced as there are very few trials with an accuracy score of 0/6 (all wrong) or 6/6 (all good) and two thirds (66%) of the trials have an accuracy of either 2/6 or 3/6. In Feldmann-Wüstefeld (2021), they used the K score from Cowan (2001) as a performance measure for each subject, where $K = N \times (hit\ rate - false\ alarm\ rate)$ and reported a correlation $r = -0.43$, $p = 0.026$ in their study. With their EEG and behavioral data we were able to reproduce a similar result with a slightly higher correlation ($r = -0.53$, $p = 0.016$).

Villena-Gonzalez et al. (2020) used the formula from Pashler (1988), $K = S \times ((H- F)/(1- F))$, where $H$ is the observed hit rate, $F$ the false alarm rate, and $S$ is the higher set size (maximum number of to-be-remembered items). In their study, instead of looking at the correlation with WMC and CDA amplitude, they looked at the increase in CDA amplitude between set size of two and set size of four items and reported $r = 0.448; p = 0.0159$ showing that participants with high WMC (i.e., better performance) showed larger amplitude increase in CDA between two and four items, compared with participants with low WMC. Unfortunately, their behavior files were not available with their EEG data, preventing us from calculating the WMC with the same formula. In the EEG files, a binary trigger identifying good and bad answers was available so we were able to calculate the accuracy for each participant. Unfortunately, when correlating the difference in amplitude between $ss = 2$ and $ss = 4$ with accuracy for each participant we found no significant correlation and therefore were unable to reproduce the effect of CDA on individual performance. It is worth noting that according to a recent study by Ngiam and colleagues, a substantial number of subjects and trials is required to detect a significant CDA difference between set sizes (approximately 400 trials with 25 subjects for approximately 80% statistical power) (Ngiam et al., 2021). Therefore, according to Ngiam’s study, the Villena-Gonzalez design is quite underpowered.

The figures of the correlations, statistical analyses and distributions are available in the supplementary material.

3.7 | Subjects variability

Given the weak but consistent correlation between CDA amplitude and individual performance, we looked at the
individual level to explore the variability on CDA across participants. On figures 19, 18, 21, and 20 in supplementary material, we did plot the CDA amplitude of the top 5 and bottom 5 participants of four different studies reviewed here (Adam et al., 2018; Villena-Gonzalez et al., 2020; Balaban et al., 2019; Feldmann-Wüstefeld, 2021), including three different conditions for each. These figures are exploratory and no further statistics were performed; however, we believe these give more perspective on the CDA shape and its variability across subjects and across studies. The blue graphs on left represent the CDA of the 5 participants with the best performance on the task (top 5) and the orange graphs represent the CDA for the 5 participants with the worst performance (bottom 5). The top left graphs in blue on the first row represent the best participant, performance wise, and the bottom right graphs in orange the last row represent the participant with the lowest performance score. By visually looking at the graphs, it would be difficult to decipher a clear trend of CDA amplitude on performance, explaining the weak correlation. It is important to note here that the y axis (CDA amplitude in microvolt) was not fixed as the peak-to-peak amplitude varies quite significantly between participants and finding a one-size-fits-all range ends up hiding the shape of the CDA, which is what we seek to showcase here. We invite the reader to pay a close attention to the value on the y axis before drawing any conclusion.

4 | DISCUSSION

As shown on previous figures, all the reproduced studies showed a clear CDA across different VWM paradigms. While it is now fair to say that the CDA amplitude correlates with the number of items in WM, the shape and amplitude of the CDA require more investigation to be better understood. For example, in Balaban et al. (2019), when the items are separated, the CDA increases and reaches a higher peak than for integrated shape as expected if the CDA amplitude correlates with the number of items in working memory. However, by the end of the retention phase, just before being probed, the two conditions have pretty much the same CDA amplitude, which would normally indicate that at that point in time there is the same number of items in working memory, which is not the case. Moreover, in Feldmann-Wüstefeld et al. (2018) Experiment 1, we see a cumulative effect of the CDA amplitude, whether from external stimuli or during recall. As mentioned before, the CDA decay can either come from an amplitude increase in the ipsilateral electrodes, reducing the difference between contra minus ipsi, or from a decrease of the amplitude in the contralateral electrodes. A combination of both is also possible. Fukuda, Woodman, and Vogel suggested that the waning CDA amplitude over time is actually a result of a selective increase in the negativity of the ipsilateral electrodes, while activity in the contralateral electrodes appears mostly unchanged (Fukuda et al., 2015). If the decay is caused by a decrease of amplitude in the contralateral electrodes, it would be coherent with WM models where oscillatory processes are keeping the information ‘alive’ via cell assembly firing together synchronizing lower frequencies in the theta and alpha bands (4 Hz to 12 Hz) with higher frequencies (40 Hz). The cell assembly passively decays, requiring regular updates before it decays too much and the information is lost (Lisman & Idiart, 1995; Luck & Vogel, 2013).

While we did not investigate further the CDA-like signal during recall, we thought it would be relevant to highlight it in this review as this period was not looked at in any of the reproduced studies and could help shed light on the neural correlates involved in VMW and underlying the CDA. For example, in their 2018 study, Feldmann-Wüstefeld and colleagues mentioned that the difference between their Experiment 1 and Experiment 2 is the memory strategy the participants might have used. Interestingly, when we compare the recall/response period of Experiment 1 we see the CDA amplitude re-increasing; however, in Experiment 2 we do not see the CDA re-increasing, as if indeed a different recall strategy was used. Interestingly, the CDA-like signal during recall seems to have the same amplitude for all conditions which is not the case during the retention phase (see Figures 11–14 top-left corner). If the observed EEG signal represents cognitive processes and not simply eye-movement artifacts, a possible explanation could be that during the initial identification of the items, the indexing happens one by one in a serial fashion, increasing the CDA in a cumulative way for each item being indexed, explaining the different amplitudes for different number of items. During recall, the update or refreshing of working memory is internally induced and could bring back all the items at once, hence showing the same CDA amplitude for all conditions. Moreover, the rate of ascend of the signal is significantly higher during recall, as if the items were brought back in a more parallel fashion to memory. Here we shared our early speculative ideas, however, given that we did not look at eye-tracking data nor exhaustively remove trials based on EOG and eye-movements during the response period, a more throughout examination would be required before drawing any strong conclusions on the meaning of that CDA-like signal.

As mentioned earlier, this sample of studies does not represent an exhaustive list of CDA studies with publicly available EEG data sets and this review could suffer from a bias given that many of the authors of selected studies are present and/or previous colleagues and collaborators.
There are several barriers for researchers to make their research reproducible; however, one that can be alleviated is the lack of knowledge and awareness about the available tools and best practices for reproducibility. In this study, we thought we would share what others have done and how they have done it, with the hope of helping other fellow researchers to do the same. Hopefully this study will reduce the fear of the extra steps required to make your study reproducible as well as highlight the value of taking these extra steps enabling one’s work to generate a greater impact on the field.

Moreover, the FAIR principles (Wilkinson et al., 2016) and the Brain Imaging Data Structure (BIDS) (Gorgolewski et al., 2016) both provide guidelines and standards on how to acquire, organize, and share brain-related data and code. As the amount of recorded brain data keep increasing around the world and become more openly accessible it is important to have best practices to reduce the friction and wasted time. In addition to reproducibility aiming at validating or invalidating scientific evidences, data mining leveraging new approaches such as artificial intelligence (AI) and deep learning (DL) can benefit from having access well-documented and openly accessible brain data sets (Roy et al., 2019). Trainees are also benefiting tremendously from reproducible experiments. Unintuitively however, most young trainees venturing in a new field think that it will be quick and easy to reproduce someone else’s work and then take it from there to either modify or improve it. Unfortunately, they quickly realize that despite being theoretically easy, it is not. There are always a multitude of complications from different operating system (OS) compatibility issues to versions of software libraries to missing parts of the code or lack of documentation making it hard to understand the code and its logic. A better reproducibility culture and best practices can help reduce significantly such friction and waste of time.

4.1 Reproducibility recommendations

(1) EEG preprocessing: in most published papers, the preprocessing steps are well described; however, many studies have a “visual inspection” part to remove some of the contaminated data. This subjective step can make reproducing the results difficult and we encourage researchers to also share the preprocessed EEG files of each participants in order to be able to compare where the results start diverging when failing to reproduce the results. This obviously increases the size of data being stored and shared but adds a lot of value.

(2) Excluded subjects: in their documentation file (e.g., README.md), researchers should make clear which subjects have been excluded from the analysis. Usually, in the published paper there is a mention like “3 subjects were removed because of ...”, while this informs the reader about the number of subjects that were included, if the raw data of these subjects are included in the data folder, it puts the burden on someone trying to reproduce the study to find the excluded ones. We have encountered four ways to deal with this issue while reproducing the studies for this review. (a) Not sharing the raw data of excluded subjects. In most cases we would not recommend this as there might be useful parts in the data. It really depends on the reason for excluding the participants from the study. (b) Making a specific Excluded folder and inserting the excluded subjects’ data in it. (c) Identifying the excluded participant in the documentation file. Sometimes that information is available somewhere in the code as a if statement like if fname == “participant_X”; however, this information should be brought forward and featured in the main documentation file. (d) Identifying the excluded participants in the published paper. This works well when one or two participants are excluded with a mention like “participant #12 was excluded because of ...” but does not scale well for multiple participants. (3) Document triggers/events; the triggers/events allowing to epoch the data properly should be clearly identified. Such information is usually a mix of hardware triggers from the EEG file and information from a behavioral file (e.g., csv file). For example, in most studies the side and the set size of each trial is available in the EEG file (saved as a signal with the EEG hardware) but the performance of the trial is available in a behavioral file. Such structure should be clearly described the main documentation file. (4) Performance scoring code; there are many ways to assess performance (e.g., K Score) and it would be beneficial to share the corresponding code to avoid confusion. In most studies, the formula is described in the article but the corresponding code is not available. (5) Online repository platform; we recommend using OSF to share the data. OSF is a free, open platform to support your research and enable collaboration, as stated on their website. There is a limit of 50GB for the free tier which might not be enough for most EEG studies; however, the cost of additional storage capacity is fairly low and has a one-time payment which works well with grant funding. All of the reviewed studies, but one, used OSF to share both the data and the code. Villena-Gonzalez used Mendeley Data instead. Another option is OpenNeuro, a free and open platform for validating and sharing BIDS-compliant MRI, PET, MEG, EEG, and iEEG data, as one can read on their website. (6) README file; we recommend having a README.md or .pdf file in the root folder explaining the crucial information about the data with reproducibility in mind. We provide as good
example from Adam et al. (2018) below. We took a print screen of the beginning of the file. Their README file provides a clear explanation of the files and the data contained in these files saving precious time of investigating to figure it out. Beware of the copy/paste across experiments! Their first sentence shows a good example of a copy/paste that was not edited. In this case, the information is not misleading and the mistake is pretty obvious but in other contexts it can be very counterproductive. (7) Raw data; researchers should include the raw data from the EEG-recording device and not an exported version from EEGLab or else. Even if the data have not been modified and have been exported as they are, it creates doubt and questions for no added value. The preprocessed data will be exported from tools like EEGLab or .mat files; however, the raw data files should be the one coming from the recording device directly.

Finally, our objective with this review, aside from better understanding the cognitive mechanisms at play during a VWM task and the key role of CDA, was to evaluate the CDA reliability as a potential candidate for brain-computer interfaces (BCIs). Many other ERPs and neural correlates are being used in brain-computer interface paradigms; however, to our knowledge, the CDA has never been used in a BCI context. We have, however, seen a recent interest in using machine learning for CDA classification with regard to the number of items held in WM (e.g., Adam et al. (2020)). We believe that within certain contexts, CDA could be used to enhance passive BCIs (Zander & Kothe, 2011).

5 | CONCLUSION

With this study we were able to look at CDA across different VWM tasks, from different groups of subjects, recorded by different groups of researchers using different EEG equipment. By using the same simple independent pipeline for all studies, we have shown that CDA is a robust neural correlate of visual working memory. As for its exact role and meaning, more research is required.

Moreover, having access to all these data sets allowed us to look beyond the usual numerical CDA mean amplitude over a window of interest but to also observe two phenomena that are understudied and underdocumented. First, the decay happening shortly after the CDA peaks while the participant still has to maintain the information. Second, the CDA-like signal during the response period. For the latter, it can be difficult to control efficiently in order to dissociate eye-movements and other artifacts from cognitive processes in the EEG signal during the response period; however, we believe it deserves more investigation as it could help shed more light on the CDA and working memory.

Finally, all the code and figures generated for this review are available online on our repository. Many more figures that we did not include in the manuscript to keep it as short and concise as possible for better readability, are available on the repository.

AUTHOR CONTRIBUTIONS
Yannick Roy: Conceptualization; data curation; formal analysis; methodology; software; validation; visualization; writing – original draft. Jocelyn Faubert: Funding acquisition; investigation; project administration; supervision; writing – review and editing.

FUNDING INFORMATION
This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC/CRDC) (reference number: RDPJ 514052–17) and an NSERC Discovery fund.

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest.

DATA AVAILABILITY STATEMENT
The EEG data used for the review is publicly available via the original articles and the code we produced is available on GitHub: https://github.com/royyannick/EEG-CDA-Review.

ORCID
Yannick Roy © https://orcid.org/0000-0003-4408-5221

REFERENCES
Adam, K. C., Robison, M. K., & Vogel, E. K. (2018). Contralateral delay activity tracks fluctuations in working memory performance. Journal of Cognitive Neuroscience, 30(9), 1229–1240. https://doi.org/10.1162/jocn_a_01233
Adam, K. C., Vogel, E. K., & Awh, E. (2020). Multivariate analysis reveals a generalizable human electrophysiological signature of working memory load. Psychophysiology, 57(12), e13691. https://doi.org/10.1111/psyp.13691
Awh, E., Vogel, E. K., & Oh, S.-H. (2006). Interactions between attention and working memory. Neuroscience, 139(1), 201–208. https://doi.org/10.1016/j.neuroscience.2005.08.023
Balaban, H., Drew, T., & Luria, R. (2019). Neural evidence for an object-based pointer system underlying working memory. Cortex, 119, 362–372. https://doi.org/10.1016/j.cortex.2019.05.008
Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. Science, 321(5890), 851–854. https://doi.org/10.1126/science.1158023
Berggren, N., & Eimer, M. (2016). Does contralateral delay activity reflect working memory storage or the current focus of spatial attention within visual working memory? Journal of Cognitive Neuroscience, 28(5), 810–815. https://doi.org/10.1177/0898929515581242
Hakim, N., Feldmann-Wüstefeld, T., Awh, E., & Vogel, E. K. (2020). Recommendations to funding agencies for supporting reproducible research. *American Statistical Association, 2*. [https://www.amstat.org/asa/files/pdfs/pol-reproducibleresearchrecommendations.pdf](https://www.amstat.org/asa/files/pdfs/pol-reproducibleresearchrecommendations.pdf)

Cowan, N. (1998). *Attention and memory: An integrated framework*. Oxford University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87–114. [https://doi.org/10.1017/S0140525X0003922](https://doi.org/10.1017/S0140525X0003922)

Faubert, J. (2013). Professional athletes have extraordinary skills for rapidly learning complex and neural dynamic visual scenes. *Scientific Reports, 3*(1), 1–3. [https://doi.org/10.1038/srep01154](https://doi.org/10.1038/srep01154)

Fazel-Rezai, R., Allison, B. Z., Guger, C., Sellers, E. W., Kleih, S. C., & Pfurtscheller, G. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *NeuroImage, 12*(3), 558–569. [https://doi.org/10.1006/nimg.2000.0909](https://doi.org/10.1006/nimg.2000.0909)

Cowan, N. (1998). *The magical number 4 in short-term memory: A reconsideration of mental storage capacity*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Perturbing neural representations of working memory with task-irrelevant interruption. *Journal of Cognitive Neuroscience, 32*(3), 558–569. [https://doi.org/10.1162/jocn_a_01481](https://doi.org/10.1162/jocn_a_01481)

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017). Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage, 159*, 417–429. [https://doi.org/10.1016/j.neuroimage.2017.06.030](https://doi.org/10.1016/j.neuroimage.2017.06.030)

Lisman, J. E., & Idiart, M. A. (1995). Storage of 7±2 short-term memories in oscillatory subcycles. *Science, 267*(5203), 1512–1515. [https://doi.org/10.1126.science.7878473](https://doi.org/10.1126/science.7878473)

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT Press.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences, 17*(8), 391–400. [https://doi.org/10.1016/j.tics.2013.06.006](https://doi.org/10.1016/j.tics.2013.06.006)

Luria, R., Balaban, H., Awh, E., & Vogel, E. K. (2016). The contralateral delay activity as a neural measure of visual working memory. *Neuroscience & Biobehavioral Reviews, 62*, 100–108. [https://doi.org/10.1016/j.neubiorev.2016.01.003](https://doi.org/10.1016/j.neubiorev.2016.01.003)

McCullough, A. W., Machizawa, M. G., & Vogel, E. K. (2007). Electrophysiological measures of maintaining representations in visual working memory. *Cortex, 43*(1), 77–94. [https://doi.org/10.1016/S0010-9452(08)70447-7](https://doi.org/10.1016/S0010-9452(08)70447-7)

Ngiam, W. X., Adam, K. C., Quirk, C., Vogel, E. K., & Awh, E. (2021). Estimating the statistical power to detect set-size effects in contralateral delay activity. *Psychophysiology, 58*(5), e13791. [https://doi.org/10.1111/psyp.13791](https://doi.org/10.1111/psyp.13791)

Obrauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 411–421. [https://doi.org/10.1037/0278-7393.28.3.411](https://doi.org/10.1037/0278-7393.28.3.411)

Olivers, C. N., Peters, J., Houtkamp, R., & Roelofsma, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences, 15*(7), 327–334. [https://doi.org/10.1016/j.tics.2011.05.004](https://doi.org/10.1016/j.tics.2011.05.004)

Parsons, S., Azevedo, F., Elsherif, M. M., Guay, S., Shahim, O. N., Govaert, G. H., Norris, E., O’Mahony, A., Parker, A. J., Todorovic, A., Pennington, C. R., Garcia-Pelegrin, E., Lazić, A., Robertson, O., Middleton, S. L., Valentini, B., McCuaig, J., Baker, B. J., Collins, E., ... Aczel, B. (2022). A community-sourced glossary of open scholarship terms. *Nature Human Behaviour, 6*(3), 312–318. [https://doi.org/10.1038/s41562-021-01269-4](https://doi.org/10.1038/s41562-021-01269-4)

Pashler, H. (1988). Familiarity and visual change detection. *Psychology & Psychophysiology, 44*(4), 369–378. [https://doi.org/10.3758/BF03210419](https://doi.org/10.3758/BF03210419)

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: An integrative review. *Biological Psychology, 41*(2), 103–146. [https://doi.org/10.1016/0301-0511(95)05130-9](https://doi.org/10.1016/0301-0511(95)05130-9)

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., ... Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: An electrophysiological account of face-specific processes in the human brain. *Neuroreport, 11*(1), 69–72. [https://doi.org/10.1097/00001756-200001170-00014](https://doi.org/10.1097/00001756-200001170-00014)

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J. (2019). Deep learning-based electroencephalography analysis: A systematic review. *Journal of Neural Engineering, 16*(5), 051001. [https://doi.org/10.1088/1741-2552/ab260c](https://doi.org/10.1088/1741-2552/ab260c)

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2015). Working memory delay activity predicts individual differences in cognitive abilities. *Journal of Cognitive Neuroscience, 27*(5), 853–865. [https://doi.org/10.1162/jocn_a_00765](https://doi.org/10.1162/jocn_a_00765)
Villena-González, M., Rubio-Venegas, I., & López, V. (2020). Data from brain activity during visual working memory replicates the correlation between contralateral delay activity and memory capacity. *Data in Brief*, 28, 105042. https://doi.org/10.1016/j.dib.2019.105042

Vogel, E., & Machizawa, M. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, 428(6984), 748–751. https://doi.org/10.1038/nature02447

Vogel, E., McCollough, A., & Machizawa, M. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, 438, 500–503. https://doi.org/10.1038/nature04171

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1), 1–9. https://doi.org/10.1038/sdata.2016.18

Yao, D., Qin, Y., Hu, S., Dong, L., Vega, M. L. B., & Sosa, P. A. V. (2019). Which reference should we use for EEG and ERP practice? *Brain Topography*, 32(4), 530–549. https://doi.org/10.1007/s10548-019-00707-x

Zander, T. O., & Kothe, C. (2011). Towards passive brain–computer interfaces: Applying brain–computer interface technology to human–machine systems in general. *Journal of Neural Engineering*, 8(2), 025005. https://doi.org/10.1088/1741-2560/8/2/025005

**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**Appendix S1 Supporting Information**

**How to cite this article:** Roy, Y., & Faubert, J. (2023). Is the Contralateral Delay Activity (CDA) a robust neural correlate for Visual Working Memory (VWM) tasks? A reproducibility study. *Psychophysiology*, 60, e14180. https://doi.org/10.1111/psyp.14180