Manipulating Goal States and Brain States: Using EEG and HD-tDCS to Investigate Mechanisms Underlying the Influence of Achievement Goals on Declarative Memory

Yuliya Ochakovskaya

The Graduate Center, City University of New York

How does access to this work benefit you? Let us know!
Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

Part of the Cognitive Psychology Commons

Recommended Citation
https://academicworks.cuny.edu/gc_etds/2830

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
MANIPULATING GOAL STATES AND BRAIN STATES: USING EEG AND HD-TDCS TO
INVESTIGATE MECHANISMS UNDERLYING THE INFLUENCE OF ACHIEVEMENT
GOALS ON DECLARATIVE MEMORY

by

YULIYA OCHAKOVSKAYA

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the
requirements for the degree of Doctor of Philosophy, The City University of New York

2018
Manipulating goal states and brain states: using EEG and HD-tDCS to investigate mechanisms underlying the influence of achievement goals on declarative memory

by

Yuliya Ochakovskaya

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date

Jennifer Mangels
Chair of Examining Committee

Date

Richard Bodnar
Executive Officer

Supervisory Committee:

Elizabeth Chua
Matthew Crump
Jon Horvitz
Elizabeth Tricomi

THE CITY UNIVERSITY OF NEW YORK
ABSTRACT

Manipulating goal states and brain states: using EEG and HD-tDCS to investigate mechanisms underlying the influence of achievement goals on declarative memory

by

Yuliya Ochakovskaya

Advisor: Jennifer Mangels

When it comes to learning factual information, students may benefit from having opportunities where they can learn from their mistakes as opposed to only being asked to study that information. However, the achievement goals that instructors set for their students may influence how students engage with these learning opportunities. Although some instructors may focus students on learning (i.e. mastery goals), others may seek to motivate students by focusing them on doing better than others (i.e. performance goals), which is thought promote greater sensitivity to errors and impair learning. Across two studies, the present dissertation examined the roles of dorsolateral frontal and lateral temporal regions in learning from errors as a function of goals that were hypothesized to underlie different mechanisms and examined whether goals differentially influence learning from errors. Participants were instructed to adopt a mastery or performance goal and answered challenging general knowledge questions. Following their response, they were then presented with accuracy feedback indicating whether their response was correct or incorrect followed by the subsequent learning feedback (i.e. the correct answer). The first study used Event-Related Potentials (ERPs) to examine processing of the correct answer, and it was expected that ERPs indicative of attention and deeper elaborative processing (i.e. superior
frontal positivity) would be more enhanced for mastery compared to performance goals during the learning feedback. Differences over inferior temporal regions were not expected between goals since this region was only thought to index bottom-up semantic processing of the correct answer. The second study employed High-Definition transcranial Direct Current Stimulation (HD-tDCS), where during the same instructions and task, HD-tDCS was applied to either left dorsolateral prefrontal (DLPFC) or lateral temporal regions in order to examine the causal influence of each region on learning as a function of goals. As in the first study, stimulation was expected to benefit learning across both regions, but DLPFC stimulation was expected to primarily benefit learning under mastery compared to performance goals. Learning was examined using immediate and/or week later delayed surprise retests. Both studies showed that learning from errors engaged both dorsolateral frontal and lateral temporal regions, but only in the ERP study was superior frontal positivity modulated by goals. Here, mastery goals led to early and late enhancements and also benefited learning across both retests compared to performance goals. In the second study, differences in learning between goals were not shown at the delayed retest. When it comes to learning from errors, mastery compared to performance goals may promote enhanced attention and elaborative processing of the learning information and benefit learning outcomes. However, behavioral benefits may not be always evident. Factors that may have contributed to these behavioral inconsistencies between the two studies are discussed along with educational implications.
ACKNOWLEDGEMENTS

I am grateful for the support and guidance of my advisor Dr. Jennifer Mangels. The time and effort you devote to mentoring is commendable and I hope all mentors follow by your lead. I am also thankful for the commitment and guidance of Dr. Elizabeth Chua. It was a pleasure to learn from you. I want to thank my committee members for the insightful discussions and the care to help me better understand my work.

The completion of this dissertation could not have been possible without the time and effort of the many research assistants. I want to thank Danielle Altman, Maya Baldwin, Stephen Bryson, Katherine Carol, Vien Cheung, Yongwon Cho, Kyle Henson, Kellyann Inniss, Anna Kataeva, Angela Kennedy, Whitney Mhoon-Mock, and Halina Shtravka. I also want to thank Ron Whiteman who during the completion of his PhD always made time to help me with my work.

Finally, I want to thank my parents, Zory and Nailya Ochakovsky, and my dear brother Steven Ochakovsky for their love and continued support.
Table of Contents

Chapter 1: General Introduction

Achievement Goals: Overview.................................................................................................................. 2
Neural substrates of Learning from Errors in Semantic Knowledge: Overview................................. 6
General Aims............................................................................................................................................... 8
The General Knowledge Paradigm........................................................................................................... 8
Defining PGs and MGs............................................................................................................................. 9
Converging Methods.................................................................................................................................. 11

Chapter 2: Mastery goals facilitate learning following errors across test-delay: An ERP Study

Introduction.................................................................................................................................................. 13
Method......................................................................................................................................................... 26
Results......................................................................................................................................................... 39
Discussion................................................................................................................................................... 66
Conclusion................................................................................................................................................ 78

Chapter 3: HD-tDCS over DLPFC and LTC Benefits Learning Following Errors Under Both Mastery and Performance Goals

Introduction................................................................................................................................................ 80
Method......................................................................................................................................................... 88
Results......................................................................................................................................................... 106
Discussion.................................................................................................................................................. 116
Conclusion................................................................................................................................................ 124
List of Tables

1. Pre-task Measures for the Main ERP Sample for Each Goal Condition........................................39
2. Pre-task Measures for the Final Study Sample.........................................................................................89
3. Participant Responses on the Manipulation Check Items for Each Condition..............................107
4. Number of Participants who Reported Receiving Active or Sham Stimulation in Each Condition..........................................................................................................................................................110
5. Median Goal Adoption Ratings..................................................................................................................131

List of Figures

1. Trial Sequence of an Incorrect Response in the General Knowledge Test.................................32
2. Median Ratings of Mastery and Performance Goal Adoption as a Function of Goal Induction Condition.............................................................................................................................................................................41
3. Proportion of Errors Corrected at Retest as a Function of Induced Goal and Retest Delay........43
4. Grand mean ERP Waveforms Time-Locked to Accuracy Feedback and Averaged as a Function of Induced Goal and Feedback Type............................................................................................................................46
5. Selected Grand Mean Waveforms over Fronto-Temporal and Parieto-Occipital Regions for the First-Test Learning Feedback Following Errors, as a Function of Induced Goal, and Sorted as a Function of Whether the Error was Later Corrected or Not on a Later Retest..........................................................................................................................................................................................48
6. Subsequent Memory Effects at Fronto-Temporal and Parieto-Occipital Sites from 500-1000 ms..........................................................................................................................................................................................50
7. Selected Grand Mean Waveforms over a Superior Frontal, Central, and Parietal Electrode for the First-Test Learning Feedback Following Errors, as a Function of Induced Goal, and Sorted as a Function of Whether the Error was Later Corrected or Not on a Later Retest.................................................................53

8. Average Amplitude of Superior Positive-Going ERP Waveforms Time-Locked to Learning Feedback at Frontal, Central and Parietal Sites for the Full Sample..................................................55

9. Selected Grand Mean Waveforms over Inferior Temporo-Parietal Regions for the First-Test Learning Feedback Following Errors, as a Function of Induced Goal, and Sorted as a Function of Whether the Error was Later Corrected or Not on a Later Retest........61

10. Average Amplitude of Inferior Temporo-Parietal ERP Waveforms Time-Locked to Learning Feedback.........................................................................................................................63

11. Montage and Current Mapping of DLPFC and LTC Stimulation.........................................................91

12. Trial Sequence of an Incorrect Response During the First-Test.............................................................96

13. First-Test Flow.........................................................................................................................................97

14. First-Test Accuracy Across the First 15 Items per Block.................................................................112

15. Retest Accuracy Across the First 15 Items per Block........................................................................114
Chapter 1: General Introduction

Instructors may shy away from using challenging assignments or pose challenging questions in their classes that are likely to result in students making errors. Likewise, students may hesitate to volunteer answers that are likely to be wrong and may prefer the instructor simply informs them of the correct information as opposed to making them guess. Instead, students may only be provided with opportunities to make errors when it comes to exams or assignments whose sole purpose is to assess learning of course material. However, when students make errors they are more likely to learn the correct information as opposed to when they are only presented with or asked to study that information (for a review see Metcalfe, 2017). Consequently, it is beneficial to learning when students are actively engaged in coming up with answers, even if they are wrong, so long as they provided with the correct information (Pashler, Cepeda, Wixted, & Rohrer, 2005).

The effort of making educated guesses seems to lead to the strengthening of the relationship between two pieces of information (e.g., a question and a correct answer, a term and its definition, etc.) and various explanations have been proposed to underlie successful learning of those associations (see Metcalfe, 2017). Yet, less is known about the neural mechanisms that support successful learning following errors and whether differences occur as a function of the achievement goals that instructors set for students. Achievement goals are thought to influence how students interpret the value of mistakes and consequently whether they learn from them (Ames & Archer, 1988). The present dissertation examined the neural mechanisms that underlie learning of semantic information following errors in the context of a mastery achievement goal that focused students on developing knowledge versus a performance achievement goal that
focused students on answering more questions correctly in comparison to others. In particular, this dissertation is focused on the complementary roles of lateral prefrontal and lateral temporal cortex in the error correction process and how achievement goals may differentially modulate neural activity in these regions during learning. Investigating differences in neural mechanisms between environmental goals can help researchers make inferences regarding the involvement of putative cognitive processes that underlie these neural regions and explain why differences in learning as a function of goals may appear. Beyond informing instructors and students about the possible advantages and disadvantages of goal environments, differences in cognitive processes can also lead to the development of interventions to promote learning that are tailored to each type of goal environment.

Achievement Goals: Overview

Majority of the work on achievement goals has focused on the goals that students personally endorse that were categorized either as mastery or performance goals (for a review see Elliot, 2005; Anderman & Patrick, 2012). Generally, mastery goals (MGs) are defined as goals with a focus on learning, improving, and/or doing well on a task, whereas performance goals (PGs) are defined as goals with a focus on demonstrating ability and/or doing better than others (for more on differences in operational definitions see Hulleman, Schrager, Bodmann, & Harackiewicz, 2010). Although researchers were initially interested in how personal goals relate to various academically-relevant outcomes, Ames and Archer (1988) later suggested that goals may also be promoted by instructors in developing a MG or PG classroom environment that influences these academically-relevant outcomes as well. They developed classroom MG and PG
criteria based on multiple factors, such as how instructors defined academic success, how they viewed mistakes, and how they evaluated student performance.

In an MG classroom, instructors focus on learning and developing skills through effort, whereas in a PG classroom, instructors focus on student ability in comparison to others through little effort. In addition, mistakes are perceived as part of learning under MG, but as a threat to ability under PG. Both personal and classroom achievement goals are typically measured using surveys where students rate the extent to which they endorse characteristics of MG and PG, or the extent to which they perceive their classroom to be focused on aspects of MG and PG, respectively. In support of their characterizations of MG and PG classroom, Ames and Archer (1988) showed that when students were more likely to perceive their classroom to reflect a PG environment they were more likely to attribute failure to a lack of ability, whereas perceptions of an MG environment were not related to failure attributions. Instead perceptions of an MG classroom were related to reports of effort as being a cause for success. Researchers typically find greater perceptions of classroom MG relate to positive outcomes, such as greater reports of intrinsic motivation (Murayama & Elliot, 2009), interest in course material (Barron & Harackiewicz, 2003), effort, persistence, and use of rehearsal and elaborative learning strategies (Wolters, 2004). However, greater perceptions of a PG classroom are typically linked to negative outcomes, such as reports of less effort (Wolters, 2004), lower levels of intrinsic motivation (Murayama & Elliot, 2009), and less interest in course material (Barron & Harackiewicz, 2003).

It is important to note that definitions of MG and PG, especially when it comes to classroom goals, typically incorporate multiple constructs in relation to a goal. For example, Barron and Harackiewicz (2003) defined classroom MG with survey items that asked about being challenged to learn new things, having interest, and instructors being willing to help
among other items. They defined classroom PG with survey items that asked about participating when having the correct answer and impressing the instructors, among other items. The challenge with definitions that are composed of multiple constructs is that it’s not clear which factor relates to an outcome. Interestingly, later examination of personal achievement goals showed that when PGs are defined in terms of demonstrating ability (e.g. appearing smart) they are linked to worse outcomes, such as poorer grades, but when PGs are defined in terms of normative comparison (e.g. doing better than others) they are linked to better outcomes, such as better grades (Hulleman et al., 2010). Less is clear about which definition of an MG is most beneficial as there has been greater variability in defining MGs unlike PGs and it appears that definitions that are actually void of goal language (e.g. I feel successful when I learn something interesting) may be best (Hulleman et al., 2010). When it comes to inducing goals in laboratory tasks, differences in outcomes between MG definitions that are typically defined as self-referenced (e.g., improve on your previous score) or task-referenced (e.g. master the task) have yet to be examined due to a low number of studies being available under each definition (Van Yperen, Blaga, & Postmes, 2015).

Findings from survey research conducted in the classroom and from laboratory memory tasks during which goals were induced suggest that an MG environment may lead to a greater use of learning strategies (Wolters, 2004) and deeper elaborative learning (Ikeda, Castel, & Murayama, 2015; Murayama & Elliot, 2011b). In line with this, Avery, Smillie, and de Fockert (2013) showed that MGs led to greater reports of using explicit step-by-step strategies that were more dependent on working-memory during a math problem solving task, whereas PGs led to greater reports of implicit short-cut strategies that were less dependent on working-memory. Similarly, other laboratory studies showed that PGs compared to MGs impair working memory
performance (Avery & Smillie, 2013) possibly as a result of concerns regarding task performance in relation to others (Crouzevialle, Smeding, & Butera, 2015). However, Avery and colleagues (2013) also showed that math problem solving was poorer under MG compared to PG if participants had to concurrently maintain a sequence of letters in working-memory so that they could later recall a letter from that sequence after completion of the math problem. The authors suggested that this disruption occurred because MGs relied more on working memory processes compared to PGs. Interestingly, Avery and colleagues (2013) provided participants with positive feedback throughout the task informing participants that they were accomplishing the goals sets for them in the MG or PG induction. Thus, positive feedback may have alleviated concerns regarding one’s own performance under PG and this may have neutralized negative effects of PG, or even led to benefits in task performance compared to MGs. Indeed, other studies have found that participants under PG showed improvements on a reading comprehension and analogy task when they were told they had performed better than others compared to when they were told they performed below average (Cianci, Klein, & Seijts, 2010; see also Crouzevialle et al., 2015). Taken together, these studies suggest that MGs may bias working-memory processing more toward the task, whereas PGs may bias working-memory processing toward concern about one’s own performance in relation to others.

However, the mechanisms that directly underlie learning as a function of goals have been largely unexplored. Recently, Mangels and colleagues (2017) used event-related potentials (ERPs) to examine the mechanisms that underlie learning of semantic information following errors as a function of goals. Under an MG or PG induction, participants were asked to answer challenging general knowledge questions that were followed by the correct answer (i.e. learning information) and later retested on that same information to examine error correction. The authors
showed that MGs led to greater fronto-temporal negativity thought to index deeper semantic processing during learning of semantic information, whereas PGs led to greater parieto-occipital negativity thought to index shallower visual processing during learning of the semantic information. However, they did not find behavioral differences in learning of semantic information between their MG and PG instruction. Using functional magnetic resonance imaging (fMRI), Lee and Kim (2014) also did not show behavioral differences as a function of personal MG or PG endorsement on a rule-finding task, but did show that personal endorsement of MGs compared to PGs was associated with greater frontal activity during presentation of accuracy feedback thought to reflect the use of cognitive control. Thus, despite differences in neural mechanisms that are associated with greater benefits under MG than PG (Lee & Kim, 2014; Mangels, Rodriguez, Ochakovskaya, & Guerra-Carrillo, 2017), when it comes to laboratory tasks benefits of MG compared to PG on behavioral outcomes are not always as evident (see also Crouzevialle, Smeding, & Butera, 2015; Murayama & Elliot, 2011; Van Yperen et al., 2015).

Neural substrates of Learning from Errors in Semantic Knowledge: Overview

The updating of general knowledge through corrective feedback can be considered a specific case of associative learning in declarative memory, which has been studied extensively with regard to neural substrates (see Kim, 2011). Yet, relatively few studies have examined knowledge updating processes in particular. Past research on correction of errors in factual knowledge (i.e., declarative memory) implicates cortical regions over the lateral temporal cortex (LTC) (e.g. Butterfield & Mangels, 2003) and frontal cortex (Mangels, Butterfield, Lamb, Good, & Dweck, 2006; Pine, Sadeh, Ben-Yakov, Dudai, & Mendelsohn, 2018). However, these cortical
regions likely make separate, but interactive contributions to this knowledge updating process. Previous ERP studies have linked the memory-encoding benefits of LTC to the bottom-up activation of pre-existing semantic representations of the correct answer, given that this activity predicted later memory only for familiar answers and not for unfamiliar answers (Butterfield & Mangels, 2003; Mangels, Hoxha, Lane, Jarvis, & Downey, 2018; Whiteman & Mangels, 2016). Furthermore, LTC has been implicated both in semantic processing of verbal information (Gold et al., 2005; Gold & Buckner, 2002; Kirchhoff, Wagner, Maril, & Stern, 2000) and in the successful encoding of verbal information, when participants were instructed to memorize information (Ezzyat et al., 2018; Wittig, Jang, Cocjin, Inati, & Zaghloul, 2018).

Simply encoding the correct answer is necessary, but not sufficient for updating knowledge in semantic memory; this answer must be properly associated with the question to which it is related. Thus, elaborative processing may also support learning following errors (Cyr & Anderson, 2015; Huels & Metcalfe, 2012; Knight, Hunter Ball, Brewer, DeWitt, & Marsh, 2012), with the dorsolateral prefrontal cortex (DLPFC) serving as a strong candidate for this elaborative processing (for a review see Blumenfeld & Ranganath, 2007). For example, DLPFC has been linked to memory success of paired stimuli when participants had to elaborate on or make judgments regarding the association of the stimuli in each pair (Blumenfeld, Parks, Yonelinas, & Ranganath, 2011; Leach, McCurdy, Trumbo, Matzen, & Leshikar, 2018; Sandrini, Cappa, Rossi, Rossini, & Miniussi, 2003; but see Hawco, Berlim, & Lepage, 2013). Thus, this region is likely to be implicated in the strengthening the relationship between a question and the correct answer following errors, but through top-down elaborative processing of semantic representations activated in LTC (see also Summerfield et al., 2006).
General Aims

The goal of these present experiments is to investigate the neural mechanisms that underlie learning of semantic information following errors as a function of achievement goals. To this aim, this dissertation will focus on 1) the role of regions within the frontal and temporal lobes in learning following errors; 2) how activity in these regions is modulated by Mastery (MG) and Performance Goal (PG) inductions. As previously discussed, the wording of goal-relevant instruction, particularly with regard to what aspects of the task they emphasize can influence outcomes. Yet, the effects of MGs and PGs on behavioral outcomes are not always evident. Thus, the two experiments that comprise this dissertation will employ the same MG and PG instructions and use the same general knowledge test-feedback-retest paradigm, but will use different, yet converging methods. Specifically, study 1 will use ERPs and study 2 will use High-Definition transcranial Direct Current Stimulation (HD-tDCS) that provide complementary approaches. We will provide an overview of the paradigm, instructions and predictions for each study in the sections that follow.

The General Knowledge Paradigm

The general knowledge test-feedback-retest paradigm used here was based on a task used extensively in our lab (e.g., Butterfield & Mangels, 2003; Mangels, Rodriguez, Ochakovskaya, & Guerra-Carrillo, 2017), in which participants provided responses to challenging general knowledge questions and were provided with both accuracy feedback and learning opportunities after each response. This initial semantic retrieval task was purposely made challenging by using
a computer adaptive titration of task accuracy to a ~35%. Differences between goals are generally more evident when tasks are difficult, but not when they are relatively easy (see Avery & Smillie, 2013; Murayama & Elliot, 2011). Later participants were retested on those items that they had initially answered incorrectly. They were not informed about the retest because the present dissertation was interested in differences in incidental learning as a function of goals that may speak more to intrinsic motivational states. This is in contrast to intentional learning that may instead motivate students to learn information primarily because of the presence of a retest. Whereas the accuracy feedback simply indicated whether their own response was correct or incorrect, the learning opportunity provided the correct answer to the question. Learning success was defined as the percentage of initial errors that were corrected on a subsequent retest which occurred immediately after the initial test (Study 1) and/or after a 1-week delay (Study 1 and 2).

Defining PGs and MGs

In this dissertation, PGs were defined in terms of normative comparison only, and not ability, to examine how a more beneficial PG compares to MGs. With respect to having multiple factors for each definition, MGs and PGs were defined in terms of the type of a goal that participants were assigned and also mentioned how participants under each goal would be evaluated (see Ikeda, Castel, & Murayama, 2015; Murayama & Elliot, 2011). Some achievement goal researchers have pushed to operationalize MGs and PGs only with respect to the competence-based goal where students are either focused on learning or on doing better than others, respectively (Elliot & Murayama, 2008; for a review see Elliot, 2005). This is in contrast to the operational definitions employed by earlier achievement goal researchers, also described
above in relation to classroom goal definitions, who defined achievement goals orientations that
does not only speak to a goal but also to the beliefs that underlie goal adoption such as the role of
effort, ability, and mistakes among other aspects (see Anderman & Patrick, 2012; Pintrich,
2000). Although orientations are more relevant to the classroom goal literature and may reflect a
more real-world representation of achievement goals than the competence-based definition, in
the present dissertation it was deemed that manipulating only a few aspects of a goal
environment (i.e. goal and evaluation criteria) would reveal whether such aspects are crucial
when it comes to learning.

Specifically, across two studies, participants under an MG induction were instructed to
develop their knowledge (i.e. goal) and that they would also be evaluated based on how well they
develop their knowledge (i.e. evaluation). Under a PG induction, participants were instructed to
do their best to perform better than others (i.e. goal) and that they would be evaluated based on
how well they compare to others (i.e. evaluation). Evaluation language was also included
because a previous study showed that students were more likely to personally endorse MGs
when informed that they would be evaluated in terms of individual progress and PGs when
evaluated in terms of social comparison (Pekrun, Cusack, Murayama, Elliot, & Thomas, 2014).
Lastly, an additional component was added to the MG induction, where participants were
instructed not to focus on doing better than others (i.e. not to endorse PG), because a previous
study showed that students value social comparison even under an MG induction unless they are
explicitly instructed to ignore such normative comparisons (Van Yperen & Leander, 2014).
Thus, these experiments examined differences in neural mechanisms that support learning of
semantic information following errors as a function of an MG induction that suppressed PGs,
versus a normative PG induction.
Converging Methods

**Study 1 (ERPs).** The first experiment employed ERPs examine neural mechanisms during presentation of the learning information following errors as well as mechanisms that underlie processing of feedback indicating answer accuracy (i.e. correct or incorrect) as a function of achievement goals. ERPs represent neural activity extracted from a continuous electroencephalography (EEG) recording that is time-locked to the onset of a stimulus (or response) and averaged as a function of task-relevant conditions. One major advantage of EEG is its temporal specificity, such that differences between conditions during specific time windows of these ERPs can help researchers make inferences about the timing of underlying cognitive processes associated with processing of discrete stimuli (see Hauk, 2016). Of relevance to the present aims, ERPs have been used extensively to study processing of both performance/accuracy feedback (e.g. Van Meel & Van Heijningen, 2010) and differences due to memory (i.e. Dm effects) (e.g., Butterfield & Mangels, 2003; Mangels et al., 2017). Dm effects (for a review see Paller & Wagner, 2002) refer to differences in neural responses at encoding that predict successful versus unsuccessful retrieval at a later test.

In Study 1, we examined ERPs during accuracy feedback that putatively indexed bottom-up (i.e. FRN and P3a) and top-down (i.e. P3b and LPP) attentional processes such as attention to unexpected events and arousal, respectively (for reviews see Hajcak, MacNamara, & Olvet, 2010; Polich, 2007; San Martín, 2012). It was expected that PGs would lead to enhancement of ERPs that underlie bottom-up and top-down attentional processing during errors in comparison to MGs. Alternatively, if normative PGs do not elicit ability concerns during errors, then errors may be perceived as similarly informative under both goals.
We also examined both early (200-300 ms) and late (500-1500 ms) ERPs during learning feedback over superior frontal regions and late ERPs over inferior frontal and inferior temporal regions. Superior frontal regions were selected for analysis given previous studies demonstrating sensitivity of early waveforms at these sites to attentional orienting (Blanchet, Gagnon, & Bastien, 2007), and later, sustained waveforms to elaborative processes that predict subsequent retrieval success (Fabiani, Karis, & Donchin, 1990; Y. Liu, Rosburg, Gao, Weber, & Guo, 2017; Mangels, Picton, & Craik, 2001). Sustained waveforms over inferior frontal and temporal regions were also examined given that they have also been shown to predict subsequent memory in this particular general knowledge paradigm as well (Butterfield & Mangels, 2003; Mangels et al., 2017).

Given that MGs are thought to facilitate deeper elaborative processes (Ikeda et al., 2015; Mangels et al., 2017; Murayama & Elliot, 2011b), it was expected that an MG compared to a PG induction would result in greater early attentional orienting to learning feedback, as well as evidence of sustained elaborative along with semantic processing. Learning was also examined as a function of retest-delay, where participants were retested on half of questions immediately after and the other half a week later. Since Mangels and colleagues (2017) did not show differences in learning using an immediate retest, the present study investigated whether benefits of MG over PG in learning might be more evident at a more delayed retest.

**Study 2 (HD-tDCS).** Although ERPs can provide precise information about the temporal sequence of processes engaged during processing of specific stimuli, the measurement of activity by electrodes at the scalp makes it challenging to determine which neural regions give rise to that activity. In addition, ERPs only show neural activity that is correlated with processing a given
stimulus or response and cannot speak to causality when it comes to behavioral outcomes. To address these issues, Study 2 employed HD-tDCS to establish causality between frontal and temporal regions and learning as a function of goals. Specifically, participants completed the same challenging general knowledge task under the same MG or PG inductions and received HD-tDCS over DLPFC, LTC or sham HD-tDCS. Unlike Study 1, participants were retested on the same questions only a week later in order to focus on directly on predictors of long-term learning.

HD-tDCS is a form of non-invasive brain stimulation that involves a low-level current applied over the scalp to target a neural region (for a review see Nitsche et al., 2008; Yavari, Jamil, Mosayebi Samani, Vidor, & Nitsche, 2018). The amount of current is not sufficient to cause neurons to fire but is thought to facilitate firing for neurons that are already near threshold. Thus, if a task relies on specific cognitive processes that enhances neuronal activity across a neural region, then applying HD-tDCS to that neural region should facilitate that cognitive process and as a result improve behavioral outcomes (see Miniussi, Harris, & Ruzzoli, 2013). It was predicted that application of HD-tDCS to lateral frontal and inferior temporal regions would facilitate encoding of the correct answer in our general knowledge task.

Specifically, it was reasoned that the DLPFC and LTC may boost learning following errors through top-down elaborative (for a review see Blumenfeld & Ranganath, 2007) and bottom-up semantic processing (Gold, Balota, Kirchhoff, & Buckner, 2005; Gold & Buckner, 2002), respectively. Additionally, although general benefits of MG compared to PG on learning following errors were expected, HD-tDCS over DLPFC and LTC were expected to benefit learning compared to sham HD-tDCS regardless of differences between achievement goals inductions. However, to the extent that MGs lead to deeper elaborative processing (Ikeda et al.,
2015; Mangels et al., 2017; Murayama & Elliot, 2011b), it was expected that stimulation of DLPFC to be particularly beneficial to memory when an MG is emphasized. In contrast, to the extent that stimulation of LTC is involved in more of a bottom-up role in basic word processing (McCandliss, Cohen, & Dehaene, 2003), it may not be differentially modulated by achievement goals.
Chapter 2: Mastery goals facilitate learning following errors across test-delay: An ERP Study

Introduction

When it comes to teaching students factual information, we may praise students’ accurate responses as indicative of learning to the point of avoiding situations where students may make mistakes. However, giving students opportunities test their knowledge in low-stakes retrieval situations where they can make mistakes and learn from them has been shown to provide better long-term benefits to learning compared to having students merely review that information passively (for a review see Metcalfe, 2017). Nonetheless, these learning benefits may depend on the how students engage with that feedback (for a review see Kluger & DeNisi, 1996), and there is growing evidence that the goals and mindsets with which they approach that feedback can impact the efficacy of errors as a learning opportunity (see DePasque Swanson & Tricomi, 2014; Mangels, Butterfield, Lamb, Good, & Dweck, 2006; Moser, Schroder, Heeter, Moran, & Lee, 2011). Yet, it is not only the personal goals that the student brings to the classroom that may impact feedback-based learning, but also the classroom goals emphasized by the instructor and peers. Despite the importance of the classroom environment in learning success, however, little is known about how these environments might influence the neurocognitive processes by which students learn from their mistakes.

The present study focuses on how the achievement goals that instructors set for their students (for a review see Ames & Archer, 1988; Anderman & Patrick, 2012) influence whether and how students engage with and learn from feedback following errors. Achievement goals are typically categorized into mastery or performance goals, but each goal can be operationalized in
various ways (for a meta-analysis on personal achievement goals see Hulleman, Schrager, Bodmann, & Harackiewicz, 2010). Generally, a classroom mastery goal (MG) is defined as one where instructors focus their students toward learning and mastering information, and evaluate students based on the knowledge and skills that students acquired in the class. A classroom performance goal (PG) is defined as one where instructors focus their students toward getting high grades and doing better than others, and evaluate students based on their performance compared to others. In the present study, we modeled a core aspect of classroom learning — learning to correct errors in general knowledge — and addressed whether and how emphasizing MG or PG may be differentially beneficial to this type of learning.

In our task, participants answered challenging general knowledge questions (see Butterfield & Mangels, 2003; Mangels, Rodriguez, Ochakovskaya, & Guerra-Carrillo, 2017) while they were assigned to an MG or PG induction. For each question, participants were sequentially provided with 2 types of feedback: the first informed them whether their answer was right or wrong (i.e. accuracy feedback) and the second informed them of the correct answer (i.e. learning feedback). We employed event-related potentials (ERPs), a noninvasive method of measuring neural activity derived from scalp-recorded electroencephalography (EEG), to examine mechanisms that underlie processing of accuracy and learning feedback as a function of goals. ERPs can provide millisecond-level resolution, where differences in ERPs within specific time frames and scalp distributions can assist researchers in making inferences regarding the cognitive processes that underlie processing of that information (see Hauk, 2016). In the present study, ERPs will allow us to investigate neural mechanisms as participants are presented with the performance and learning information in real-time, as well as help us make inferences about the putative cognitive processes that may underlie learning as a function of goals. Learning was
examined as a proportion of errors that were corrected on a later immediate and a week later surprise retest as we were interested in differences in retention as a function of goals as well.

An overarching framework for understanding differences in mechanisms between goals stems from goal-setting theory that proposes that goals direct attention and effort to goal-relevant information (see Locke & Latham, 2002, 2006). When it comes to making mistakes, Ames and Archer (1988) theorized that PG environments elicit concerns regarding ability, where errors are likely to be perceived as threatening in that they may indicate low ability. Using surveys, the authors showed greater perceptions of classroom PG were positively related to beliefs that a lack of ability was a cause of failure. MG were thought to diminish ability concern, and errors were likely to be perceived as part of learning. Thus, PG environments may direct more attention toward errors compared to MG possibly due to concerns regarding ability.

It is important to note that Ames and Archer (1988) defined PG as a goal where students are focused on demonstrating their ability. However, in a meta-analysis of personal goals Hulleman and colleagues (2010) showed that when PGs are defined only in normative terms (i.e. focus on doing better than others) they are linked to better grades, but when they are defined in terms of ability (i.e. focus on appearing smart) are linked to worse grades. Ames and Archer (1988) classroom PG definition contains both ability and normative terminology, where doing better than others is a means by which one can achieve higher ability. In the present study, we defined PGs only in normative terms and did not include any language regarding ability to examine if under a normative PG errors are perceived to be more threatening compared to an MG induction. It may be that normative PGs also foster concerns regarding ability during failure. Alternatively, it may be that errors are perceived similarly under a normative PG induction that does not speak to ability in comparison to an MG induction. We also defined MGs and PGs in an
approach manner where goals are framed toward success (e.g. do better than others) as opposed to an avoidance manner where goals are framed away from failure (e.g., do not do worse than others), because approach goals are typically associated with better outcomes compared to avoidance goals especially when it comes to PGs (Van Yperen, Blaga, & Postmes, 2015).

Given the focus of MG on learning, goal-setting theory would suggest that an MG environment should bias attention toward learning-relevant information. Specifically, to the extent that negative performance feedback identifies where knowledge needs to be developed, students under an MG environment should be more likely than those under a PG environment to be motivated to maintain attention and engage effort to process the correct answer. Indeed, research in the classroom shows that a greater perception of an MG classroom is linked to greater reports of intrinsic motivation (Murayama & Elliot, 2009), interest in course material (Barron & Harackiewicz, 2003), effort, persistence, and use of rehearsal and elaborative learning strategies (Wolters, 2004). In addition, perceptions of an MG classroom are linked to greater reports of effort as being a cause for success (Ames & Archer, 1988). In contrast, greater perceptions of a PG classroom are linked to reports of less effort (Wolters, 2004), intrinsic motivation (Murayama & Elliot, 2009), and interest in course material (Barron & Harackiewicz, 2003). Similarly, researchers using laboratory memory tasks suggest that MGs compared to PGs lead to deeper elaborative processing of learning material (Ikeda et al., 2015; Mangels et al., 2017; Murayama & Elliot, 2009). Elaborative processing is thought to facilitate learning following errors (Cyr & Anderson, 2015; Huelser & Metcalfe, 2012; Knight et al., 2012) and is also likely to support learning in our task where error correction requires that the correct answer is successfully associated with a particular general knowledge question, not just remembered in
isolation. Thus, we expected an MG compared to a PG induction to facilitate greater engagement with the learning feedback and facilitate learning as evident by greater error correction.

To investigate the neural mechanisms that underlie processing of errors and learning as a function of achievement goals we examine ERPs during the presentation of accuracy and learning feedback. In the following sections we discuss ERPs of interest and the cognitive processes that they are likely to index to help us makes inferences of the cognitive processes that may underlie processes under each goal. We end with a section on study overview and predictions.

Accuracy Feedback & ERPs

In the present study, we examined ERPs during accuracy feedback that informed participants whether their response was correct or incorrect. Although errors may carry different meaning under PG compared to MG environments, thus far, differences have only been observed when examining personal endorsement of achievement goals, as measured through surveys, and not when goals are experimentally manipulated. Specifically, Mangels, Butterfield, Lamb, Good, and Dweck (2006) showed that greater personal endorsement of performance goals were associated with enhanced orienting to negative feedback signaling a response was wrong, particularly when participants were initially confident that they were going to be correct, and thus, negative feedback would have been particularly unexpected. This enhanced orienting response evidenced by positive correlations between the amplitude of the P3a, an early midline positive ERP potential, and endorsement of performance goals that were defined using both normative and ability terms. The P3a is a positive-going ERP component that occurs between
250 and 500 ms post-stimulus that is thought to index attention to unexpected events (for a review see Polich, 2007). In a later study focusing on manipulating achievement goals in the context of the same task, however, Mangels and colleagues (2017), failed to show any goal-related modulation of the P3a. These inconsistencies in goal effects on the response to signals of failure may be related to differences in how achievement goals function in a state (induced) versus trait (individual difference) situation or the inclusion of both ability and normative language in the definition of personal PGs in the first study.

The feedback-related negativity (FRN) is another ERP component that is typically examined during accuracy feedback. The FRN precedes the P3a and is a negative-going ERP peaking around 250 ms post-stimulus over frontal electrodes during errors compared to correct responses (for a review San Martín, 2012). Although differences in the FRN have not been shown as a function of either induced (Mangels et al., 2017) or personal achievement goals (Mangels et al., 2006), Van Meel and Van Heijningen (2010) did show differences in the FRN between errors and corrects during a probabilistic learning task when participants were instructed to outperform others, but not when the trials were presented as practice. It seems that students may be more sensitive to accuracy feedback when placed in a competitive setting, which may reflect similar attentional processes under our normative PG induction.

The P3a may speak to bottom-up attentional processing in response to feedback (see Polich, 2007), however, we were also interested in the effects of achievement goals on the P3b and the Late Positive Potential (LPP), two ERPs that have been characterized as indexing aspects of top-down, voluntary attention. The P3b is a positive-going ERP over parietal electrodes in the range of 300 to 600 ms post-stimulus that is generally thought to index top-down processes involved in updating working memory in response to a task-relevant event (for a review see
Polich, 2007), although it is also enhanced for unexpected outcomes as long as those outcomes have motivational relevance (for a review San Martín, 2012). Van Meel and Van Heijningen (2010) did not show the presence of competition to differentially modulate the P3b as they did with the FRN in relation to accuracy feedback compared to a practice condition. However, participants in that study typically provided correct responses, where in the present study the majority of responses are likely to be errors. In the context of repeated failures, PGs may elicit concerns about ability when it comes to errors (Ames & Archer, 1988). Consequently, the FRN may index bottom-up processing to negative feedback but only when negative feedback is rare or unexpected (see also Sambrook & Goslin, 2015), whereas the P3b may be more reflective of top-down processing to errors when they are more prevalent in the context of failure. Alternatively, goals may only modulate the P3b but not the FRN since negative feedback is persistent in our task. Thus, a PG compared to an MG induction may lead to greater enhancements of P3b during errors indexing the use of working-memory processes that are directed toward this undesirable event.

The LPP is also a positive-going ERP that appears around 300 ms post-stimulus over posterior midline electrodes, however unlike the P3b it is may extend even past 1000 ms post-stimulus. It is thought to index sustained attention and arousal to motivationally relevant information (for a review see Hajcak, MacNamara, & Olvet, 2010; Lang & Bradley, 2010). Previously, (Mangels, Good, Whiteman, Maniscalco, & Dweck, 2012) found that an enhanced LPP to negative feedback during a math problem solving task lead to poorer learning from corrective feedback in women who were informed about the negative stereotypes regarding female’s abilities in math (i.e., under stereotype threat). More recently, Whiteman & Mangels (2016) found, using a similar task to the present study, that the LPP to negative feedback was
enhanced in individuals who had a greater trait tendency to ruminate on the negative consequences of failing to attain personal goals. Consequently, in the context of a task where participants repeatedly experience negative feedback, a PG compared to an MG induction may also enhance LPP during errors. This may be the case especially if normative PGs may foster concerns regarding ability and the repeated negative feedback threatens perceptions of that ability.

Learning Feedback & ERPs

In support of goal theory (see Locke & Latham, 2002, 2006), where an MG compared to a PG induction is expected to direct greater attention and effort to learning-relevant information, Mangels and colleagues (2017) showed MGs compared to PGs led to deeper processing of learning feedback. Here, MGs led to enhanced inferior fronto-temporal negativity from 400 to 800 ms during learning feedback when that learning feedback was later remembered compared to when it was not remembered on a retest (i.e., difference due to memory (Dm) effects; Paller & Wagner, 2002). As for PGs, Dm effects during learning feedback in the same time window were instead shown over parieto-occipital regions. In accordance with views on the progression of visual verbal information progression through the ventral stream (see Marinkovic et al., 2003), the authors proposed that parieto-occipital Dm effect under PGs was more indicative of basic visual processing of the learning feedback, whereas the inferior fronto-temporal Dm effect under MGs putatively indexed semantic processing of the learning feedback. Thus, an MG compared to a PG induction may facilitate deeper processing of the learning feedback.
The present study will attempt to replicate these findings by examining ERPs over inferior fronto-temporal and parieto-occipital regions, but additionally will examine ERPs over superior frontal regions putatively involved in attentional orienting and elaborative processing. ERPs over superior frontal regions are typically more positive going when tasks necessitate semantic processing or elaborative memory strategies, compared to when participants are focused on less demanding tasks such as those involved in detecting visual features of words or using rote rehearsal memory strategies (Fabiani, Karis, & Donchin, 1990; Guo, Zhu, Ding, Fan, & Paller, 2004; Wieser & Wieser, 2003). Superior frontal positivity from as early as 150 ms post-stimulus and up to 400 ms has been shown to be enhanced when participants were instructed to employ deeper semantic processing or elaborative memory strategies compared to shallow visual processing or when participants were not instructed or able to use specific memory strategies (Blanchet, Gagnon, & Bastien, 2007; Guo et al., 2004; Wieser & Wieser, 2003). For example, tasks that instructed participants to evaluate the meaning of an item (Guo et al., 2004) or a relationship between two items (Wieser & Wieser, 2003) led to early enhanced frontal positivity compared to tasks that instructed participants to evaluate perceptual aspects of stimuli (e.g., Is a stimuli in bold face? Are words of equal length?). In another study, greater frontal positivity was shown for words when participants may have been more likely to use organization memorization strategies compared to when participants were unlikely to employ any specific memorization strategy (Blanchet et al., 2007), suggesting that this early frontal positivity reflects enhancements in attention and effort toward processing the incoming information. Thus, an MG compared to a PG induction may lead to enhanced early frontal positivity that may reflect enhancements in attention and effort to the learning feedback following errors in our task as well.
Later sustained positivity over superior frontal regions starting around 300 ms post-stimulus has been shown predict later memory and is thought to reflect elaborative processes that contribute to encoding success (e.g., Fabiani, Karis, & Donchin, 1990; Guo et al., 2004; Kamp, Bader, & Mecklinger, 2017; Liu, Rosburg, Gao, Weber, & Guo, 2017; Mangels, Picton, & Craik, 2001). For example, Fabiani and colleagues (1990) found that sustained frontal positivity was enhanced for later remembered words compared to words that were not remembered (i.e. Dm effect) when participants were instructed to use an elaborative memorization strategy such as forming sentences between words, but differences in frontal positivity was not related to later memory outcomes when rote-rehearsal was employed. Similarly, following initial study of word pairs, Liu and colleagues (2017) showed frontal Dm effects when participants were instructed to employ an elaborative associative strategy when presented with the same word pairs again (i.e. generate a word that relates to both words for each word pair), but not when participants were instructed to only use a retrieval strategy where they were presented with the first word of a previously studied word pair and asked to try and remember the other word that was paired with it. Liu and colleagues (2017) concluded that frontal positivity may index semantic processing or working-memory based elaboration given that participants were asked to link and strengthen the association of words in each pair. Thus, in our task, an MG compared to a PG induction may lead to greater Dm effects over superior frontal regions that may index greater use elaborative processing that supports the strengthening of the association between the correct answer and question.

Study Predictions
In the present study, we examined whether MG or a PG inductions differentially affected learning after errors in a general knowledge task, both in terms of behavior as measured by error correction on both an immediate and 1-week delayed retest, and in terms of their influence on selected ERP waveforms associated with processing of accuracy and learning feedback. As mentioned above, both MGs and PGs inductions were defined in an approach manner in terms of developing knowledge and in normative terms (i.e. do better than others), respectively. We also measured personal incoming achievement goals at the onset of the study to ensure that pre-existing differences in endorsement of MG and PG between groups did not occur prior to the goal induction.

Examining neural mechanisms as a function of goals that support learning even when memory is tested a week later can provide insight into the type of cognitive processes that may boost retention and speak to cognitive processes that benefit learning across both retest-delays. Behaviorally, we expected that the MG induction would result in better learning outcomes than the PG induction, and that these effects would be magnified when examining memory over a longer delay. Thus, the use of an immediate and week later retest will allow us to examine when learning differences between goals may appear and whether differences in delay matter.

Regarding processing of accuracy feedback, we expected that compared to MGs, normative PGs would lead to greater involuntary and/or voluntary attention to negative feedback impugning ability, as evidenced by enhancement of fronto-centrally focused FRN and P3a components, and/or centro-parietally focused P3b, and LPP components, respectively. In relation to the FRN, we also measured the preceding P200 component to ensure that our FRN measure was not influenced by this preceding positivity, and for that reason along with the peak-picked
means we also conducted peak-to-peak measures from the P2 to the FRN, as well as from the FRN to the P3a.

Regarding processing of learning feedback, we first attempted to replicate Mangels and colleagues (2017)’s findings of differential Dm effects for MG and PG inductions over inferior fronto-temporal and parieto-occipital regions, respectively, and examine whether the addition of the greater delay would enhance these effects further, in addition to revealing behavioral effects that had not been apparent in that earlier study. Then, to the extent that MGs not only foster deeper, semantic processing of learning-relevant feedback, but also more elaborative, associative learning strategies in particular, we expected greater attentional enhancement and deeper elaborative processing of the learning feedback under an MG compared to a PG induction. These would be evident in enhancement of superior frontal positivity during both an early and later, more sustained period, respectively. Again, we expected that indices of these relational processes might be particularly evident when examining memory at the 1-week delay. While we also expected to replicate sensitivity of activity at inferior temporal sites to error correction success overall, regardless of delay (e.g., Butterfield & Mangels, 2003; see also Whiteman & Mangels, 2016; Mangels et al., 2018), we did not predict differences here as a function of MG and PG instructions (see also Mangels et al., 2017), given the proposed role of this activity in more bottom-up processes involved in the activation of pre-existing representations of the correct answer in semantic memory.

**Method**

Participants
We tested 80 eligible participants in the ERP study, with half of the sample randomly assigned to each condition (MG = 40 [18 female]; PG = 40 [19 female]). These students represent 31% of the 260 EEG-eligible participants identified through an online prescreen of 1182 Baruch college students. Participants were eligible for the study if they reported to be right-handed, have normal vision and hearing, no psychological or neurological disorders, to have learned English by age 6 and have all of their schooling in the United States. As for ethnicity, 21 identified as Asian (MG = 13), 5 identified as Black or African American (MG = 3), 13 identified as white (MG = 7), and 9 identified as more than one race (MG = 4). Of the 80 participants, fourteen participants identified as Hispanic (MG = 7), and 2 participants did not respond at all to this or the ethnicity question. Participants in the ERP study were provided with research credits and/or paid $10 per hour with an additional credit/$10 for returning to the second session that occurred a week later.

Procedure

Overview. Our basic research design used a computerized test-feedback-retest general knowledge paradigm similar to that used in previous studies (Butterfield & Mangels, 2003; Mangels et al., 2017). In the current study, after completing a questionnaire assessing their personal achievement goals (Achievement Goal Questionnaire-Revised; Elliot & Murayama, 2008)¹, participants were prepared for EEG recording, and began the initial test phase. During

---

¹ This measure consists of 12 items, with 3 items per goal (i.e., MG approach, MG avoidance, PG approach, PG avoidance) that participants rate on a scale of 1 (strongly disagree) to 5 (strongly agree). MG approach items focus on learning the potential maximum (e.g., My goal is to learn as much as possible), whereas MG avoidance items focused on avoidance of not learning the potential maximum (e.g., My aim is to avoid learning less than I possibly
this phase, participants first were presented with either MG or PG approach instructions (see Achievement Goal Induction section), then answered 200 general knowledge questions broken down into four equal blocks of 50 questions while EEG activity was recorded. Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA) was used to present goal instructions and the general knowledge questions.

For each question, participants typed in an answer and their confidence in the answer’s accuracy, then received veridical positive or negative feedback about their answer accuracy followed by the correct answer (see General Knowledge Test & Retests section). Overall initial test accuracy was titrated to ~35% correct in order to ensure a similar experience of challenge for all participants, as well as a similar (and large) number of errors to be corrected on the retest, regardless of any pre-existing differences in general knowledge. After completion of the initial test, participants answered manipulation check questions (see Manipulation Check section). The EEG apparatus was then removed and participants were asked to complete a surprise retest comprised of half of the questions they encountered at the first test of the general knowledge questions. A week later, they were asked to complete a second retest of the remaining half of the first-test questions. No mention of either upcoming retest was made during the initial test phase rather participants were told they would return to answer additional general questions. Both retests included both initially correct and incorrect questions.

could). PG approach items focused on attaining normative competence (e.g., I am striving to do well compared to other students) while PG avoidance items focused on avoidance of not achieving the lowest level of normative competence (e.g., My goal is to avoid performing poorly compared to others). Because this survey was originally developed to assess incoming personal goals within a class, minor updates were made to the original survey to gear students toward thinking about their goals more generally. The first item queried MG approach and asked students to respond to this item in relation to “in this class”, which was updated to “in my courses.” The other minor change for 2 other MG items was an update from “this/the course” to “my courses.” Scores for the incoming goals were computed by averaging the 3 items for each of the 4 goals.
Achievement Goal Induction. Prior to answering questions in the general knowledge task, participants were presented with a screen titled “Task Instructions.” This instruction geared students toward an MG or PG.

Participants in the MG condition were instructed with the following:

_You are here to explore a set of general knowledge questions in order to develop your knowledge about various topics. You will be given feedback about the accuracy of your answers. Sometimes people think that we are comparing their performance to that of other students who have completed the task. However, your focus should NOT be to do better than other students, but rather to develop your own knowledge._

_At the end of session 2, you will be provided with information you may find useful for understanding how your knowledge has developed over the course of this task._

Participants in the PG condition were instructed with the following:

_You are here to answer a set of general knowledge questions about various topics. You will be given feedback about the accuracy of your answers in order to give you a sense of how well you are performing. Your performance on this task will be compared to other Baruch students and you should do your best to perform better than others._
At the end of session 2, you will be provided with information about how well you did compared to other Baruch students.

This full instruction set was only presented prior to block 1, but at the onset of blocks 2, 3, and 4, participants were provided with brief goal reminder instructions. Specifically, in the MG condition, participants were simply reminded that they “are here to explore the general knowledge questions and to develop your knowledge.” In the PG condition, participants were reminded that they “are here to do your best to perform better than other students at Baruch.”

A brief version of the achievement goal instructions was also restated prior to both the immediate and delayed retest. Participants in the MG instruction were instructed to “keep in mind that you are here to explore the general knowledge questions and to develop your knowledge” and that “at the end of Session 2, you will be provided with information you may find useful for understanding how your knowledge has developed over the course of this task.” For the PG instruction, participants were instructed to “keep in mind that you are here to do your best to perform better than other students at Baruch” and that “at the end of session 2, you will be provided with information about how well you did compared to other Baruch students.”

General Knowledge Test & Retests. The 200 general knowledge questions that comprised the initial test for each participant were selected from a larger database of 416 difficulty-normed questions (http://www.mangelslab.org/bknorms). Questions were related to topics in history, geography, math, sciences, arts and humanities. The particular subset of questions presented varied across participants because it was based on an individualized titration algorithm designed
to stabilize their performance at ~35% accuracy across the initial test (see Mangels et al., 2017 for algorithm details).

On each individual trial, participants first were asked to type in their response within a 3-minute time limit and then select a confidence rating on a scale of 1 (sure the response is wrong) to 7 (sure the response is right). Then participants were presented with a central fixation cross for 1.5 seconds, followed by feedback signaling accuracy (accuracy feedback; negative feedback: low pitched tone paired with a red asterisk, positive feedback: high-pitched tone paired with a green asterisk. Positive feedback was shown if there was at least a 70% letter match (regardless of letter order) between a participant’s response and the correct answer. This algorithm allowed for participants to receive positive feedback for a slightly misspelled correct answer. Negative feedback was shown when the letter match between a participant’s response and the correct answer fell below 70%. The accuracy feedback was followed by another fixation cross for 1.5 seconds and then with the corrective learning feedback (i.e., correct answer) for 2 seconds (see Figure 1). The ITI was 500 ms.
Fig 1. Trial sequence of an incorrect response in the general knowledge test. Participants have a 3-minute time limit to type in a response to a question. After their response, participants are asked to make a confidence rating with a response of 1 if they are sure the answer is wrong, a 4 if they are unsure whether their answer is right or wrong, and a 7 if they are sure they are right. If the answer is incorrect, then the participant is presented with a red asterisk paired with a low pitch sound, but if they are correct then a green asterisk is presented with a high pitch sound. If a participant does not respond within 3 minutes, then they are immediately presented with the first fixation skipping the confidence rating and are presented with negative feedback (i.e. incorrect response). Learning feedback (i.e. correct answer) is presented regardless of response accuracy.

The immediate surprise retest was given on the same day as the initial test, following removal of the EEG cap and an opportunity for the subject to clean and refresh themselves (~15 mins). The delayed surprise retest took place 1 week later for all except one participant who
came back after 1 week and 1 day. To ensure that the two retests were well-matched to each other for difficulty, the split of first test questions across retests attempted to distribute questions equally as a function of question block, subject accuracy and subject confidence. Each of the 100 question retests were subdivided into 4 blocks of 25 questions. For both retests, question block was preserved (i.e., a question in block 1 of the initial test was presented in block 1 of the retest), however question order within a block was randomized. The trial structure of the retest questions was identical to the first test, with the exception that feedback combined accuracy and learning components into a single composite feedback (negative feedback: correct answer in red and a low pitch tone; positive feedback: correct answer in green and a high pitch tone).

Manipulation Check. In line with Murayama and Elliot (2011), participants completed 2 goal adoption questions after completion of the general knowledge task, but before the start of the immediate retest. Here, participants were asked to rate their adoption of mastery (MG adoption) and then of the performance goal (PG adoption) on a scale of 1 (not at all) to 6 (very much). For the MG adoption question, participants were asked to rate how hard they tried to develop their knowledge. For the PG adoption question, participants were asked to rate how hard they tried to do well in comparison to other Baruch college students.

EEG Recording. Continuous electroencephalography (EEG) was recorded during the first test only using Neuroscan Synamps 2 (Compumedics USA Charlotte, NC) system with a 64-channel Quick-Cap. This montage included 7 monopolar electrodes to capture eye movements (Nz, IO1/IO2, LO1/LO2, and an electrode midway between lateral and inferior sites under each eye).
EEG was sampled at 500 Hz with a bandpass of DC-100 Hz and referenced to Cz. Impedance was established at 5 kΩ or below for all electrodes before the start of the study.

After recording, we applied an offline average-reference and added Cz back in. Data was filtered from .15 Hz to 35 Hz and eye movements were corrected using a PCA-derived ocular correction algorithm in BESA 5.3 (Brain Electric Source Analysis, Gräfelfing, German) resulting in 6 ocular components for each participant. Prior to averaging, epochs underwent baseline correction (-100 ms prior to stimulus onset) and artifact rejection (manual review, and automatic scan set to reject epochs with amplitudes ±100 µv). Epochs of 1 s and 1.5 s were used in analyses of accuracy and learning feedback, respectively.

Data Analyses

Behavioral data. Of the 80 participants, 6 participants (MG = 4) were excluded because of missing retest data either due to running out of time to complete the immediate retest or not returning for the delayed retest. An additional participant was excluded in the PG condition for falling asleep and another in the MG condition for reporting to look up numerous answers prior to the completion of week-delayed retest. In addition, we excluded 5 participants who were outliers in initial test accuracy despite our use of titration to stabilize performance (boxplot outlier method: >1.5 inter-quartile range above or below median). Extreme differences in initial

---

2 Two participants in the MG condition reported looking up or rehearsing some items in the week intervening between the immediate and delayed retests. This post-encoding rehearsal may undermine the relationship between EEG activity recorded during encoding and memory performance on a retest. One participant could report the 17 questions/answer they rehearsed, and thus, we opted to remove these trials from their delayed retest analysis, but otherwise keep the participant in the analysis. The rehearsed trials were also kept in the immediate retest analysis. However, the other participant was removed from analysis because we did not have specific information on the trials they rehearsed. This participant was also an outlier (more than 1.5 IQR) in their delayed retest score, but not at the immediate retest, suggesting rehearsal had advantaged their delayed performance.
test performance could impact cognitive and emotional factors known to influence retest performance (i.e., memory load, pre-existing semantic knowledge, feelings of failure). There were 3 outliers in the MG condition (2 below, 1 above), and 2 in the PG condition (both below). This left 32 participants (13 females) in the MG condition and 35 participants (16 females) in the PG condition for behavioral analysis.

In considering how accuracy at the initial test and retests were calculated, we note that the letter-matching algorithm the program used to provide rapid, automatic feedback was occasionally imprecise and resulted in inaccurate or ambiguous feedback (1% of all trials). This necessitated manual review and confirmation of response accuracy by three raters. The majority determination by manual rating was used in calculating both first-test and retest accuracy, irrespective of the feedback provided by the program. Given the importance of accurate feedback as a learning signal, however, we opted to exclude questions from the calculation of retest error correction that had received inaccurate or ambiguous feedback at the initial test. In total, this resulted in the loss of an average of 2.9 behavioral trials for the MG and 3.5 trials for the PG condition, but this did not significantly differ between goal conditions, $t(58) = .777, p > 0.43$.

Although confidence ratings prior to accuracy feedback were measured, we did not subdivide trials further as a function of confidence levels because of concerns about meeting criteria for an adequate number of trials in all bins, especially given that encoding trials were already subdivided as a function of immediate and delayed retest. A 2 (Induced Goal: MG or PG) x 2 (Accuracy: Corrects or Errors) ANOVA on median confidence only showed a main

---

3 Incorrect positive feedback occurred when a correct and incorrect answer shared 70% or more letters (e.g., correct answer: “Uranus”; given answer: “Saturn”). Ambiguity regarding the appropriateness of negative feedback occurred in situations where their response was semantically similar (e.g., correct answer: “human”; given answer: “people”), was the first name instead of the last name of the correct answer (e.g., correct answer: “Charles”; given answer: “Dickens”), a misspelling close to a correct response but the match was less than 70% (e.g., correct answer: “ostrich”; given answer: “ostrid”).
effect of Accuracy, $F(1, 58) = 239.971, p < 0.001$. Median confidence for corrects ($M = 5.55, SEM = .18$) was greater than median confidence of errors ($M = 2.54, SEM = .14$). No other significant effects were shown ($ps > .51$). Furthermore, if we use this median for errors to split these trials into higher and lower confidence errors, we find that there are no goal differences in the frequency of either error type as evident by a lack of differences in the number of high confidence errors trials, $t(58) = .226, p > 0.82$ (MG: $M = 44.50, SEM = 2.56$; PG: $M = 43.60, SEM = 3.06$). Thus, although confidence in errors can influence the amplitude of ERPs to accuracy feedback and the likelihood of correcting errors on a subsequent retest (i.e., hypercorrection effect; see Butterfield & Mangels, 2003), given the lack of goal effects on confidence, we collapsed across confidence in order to focus on our primary aim of examining goal effects (for similar approaches see also Mangels, Hoxha, Lane, Jarvis, & Downey, 2018; Mangels et al., 2017; Whiteman & Mangels, 2016).

ERP Data Analyses

**ERP Sample.** From the 67 participants in the behavioral sample, it was necessary to exclude an additional 7 participants (2 from MG condition) from ERP analysis: 4 due to electrode bridging, and 3 due to excessive noise (i.e., alpha activity, muscle noise or issues with the reference electrode). The final ERP sample included 30 participants (13 females) in the MG condition and 30 participants (12 females) in the PG condition. Only trials with usable feedback were included in these analyses.
**Accuracy Feedback.** ERPs were averaged for positive and negative accuracy feedback separately. Trial counts did not differ as a function of conditions for either the positive feedback ($t(58) = .606, p > 0.45$; MG: $M = 67.8, SEM = 1.35$; PG: $M = 66.7, SEM = 1.29$) or negative feedback ($t(58) = .296, p > 0.76$; MG: $M = 117.2, SEM = 1.42$; PG: $M = 116.5, SEM = 1.87$).

Given their close temporal proximity, we measured FRN and P3a components by peak-picking at Fz and analyzing the amplitude of the 25 ms mean around each peak. In addition, we also used this strategy to select the large positive peak prior to the FRN (i.e., the P2) that then enabled us to analyze the FRN, which could sometimes be more difficult to isolate, with peak-to-peak measures from this prominent positive waveform. The peak of the P2 was identified as the first positive peak between 100 and 300 ms, the FRN was identified as the most negative peak following the P2, and the P3a was identified as the most positive peak following the FRN. Peak latencies of these did not differ as a function of goal, feedback valence or their interaction ($ps > .25$; see Supplemental Table A1 in Appendix A for means and SEMs). In addition to the analyzing the peak-picked mean amplitudes of each component, we also evaluated peak-to-peak amplitude differences between the P2 to the FRN and from the FRN to the P3a. As for the P3b and LPP components, mean average windows between 200 to 400 ms and 500 to 1000 ms were used, respectively, at electrode CPz.

**Learning Feedback.** Our analysis of the learning feedback focused on trials in which the participant made an error on the initial test, both in terms of the activity of all learning feedback following errors, and then after back-sorting as a function of whether that error was later corrected on a retest or not (i.e., difference due to memory [Dm] effects). We set a minimum of 8 usable trials in both corrected and uncorrected conditions for a participant to be included in the
analysis. Using this criterion, it was possible to include all participants if we collapsed over retest delay, but if we included the factor of test-delay, it reduced our sample to 22 participants in the MG condition and 26 participants in the PG condition. Thus, analyses both collapsing over test delay (30 subjects in each condition) and as a function of test delay (22 in MG, 26 in PG) are reported here.

Our replication of Mangels and colleagues (2017) focused on the sustained negative-going potentials observed over the inferior fronto-temporal (FT9, FT10, T7, T8) and parieto-occipital (PO3, O1, PO4, O2) regions that they had identified by as sensitive to achievement goal manipulation. As in that study, activity at fronto-temporal and temporal electrodes in each hemisphere were averaged to create a left (FT9/T7) and right (FT10/T8) fronto-temporal average, and activity at parieto-occipital and occipital electrodes in each hemisphere were averaged to create a left (PO3/O1) and right (PO4/O2) parieto-occipital average. Although Mangels et al. (2017) focused their analyses on the 400 - 800 ms period, in the current study we conducted analyses on the 500 – 1000 ms window for parsimony with our other analyses.

Our measurement of superior frontal activity focused on F3, Fz, and F4. Using average windows, we measured an early positivity to learning feedback from 200 to 300 ms, a mid-latency sustained positivity from 500 – 1000 ms, and a later sustained positivity from 1000 – 1500 ms. To examine the scalp distribution of activity during these time periods, these analyses also included central (C3, Cz, C4) and parietal (P3, Pz, P4) electrodes. Finally, we examined temporo-parietal negativity during the learning feedback using average windows from 500 – 1000 ms and 1000 – 1500 ms at two electrode sites within each hemisphere (TP7/TP8; TP9/TP10), similar to previous studies (Butterfield & Mangels, 2003; Whiteman & Mangels, 2016).
Although all significant (p < .05) and marginal (.10 > p > .05) main effects and interactions are reported, only effects involving either memory and/or goals are discussed in detail (i.e., other effects are described in footnotes). Post-hoc comparisons were conducted using Bonferroni adjustment for the relevant number of comparisons.

Results

Behavioral Results

Pre-task Measures. There were no differences in any pre-task measures between our goal conditions (see Table 1). A 2 (Induced Goal: MG or PG) x 2 (Goal Type: Mastery or Performance) x 2 (Valence: Approach or Avoidance) ANOVA on mean AGQ-R ratings did not show any significant effects of Induced Goals (ps > .66). In addition, because Shapiro-Wilk tests showed that the 4 personal goals (i.e. MAP, PAP, MAV, PAV) failed normality (ps < .04), we also ran non-parametric Mann-Whitney U tests and all 4 tests showed that differences at each personal goal did not differ between Goal Conditions (ps > .58). An independent t-test also showed that age did not differ between Goals (MG: M = 21.53, SEM = .83; PG: M = 20.30, SEM = .66), t(58) = 1.161, p > 0.24.

Table 1

Pre-task measures for the main ERP sample for each goal condition. Personal goal measures (i.e., MAP, PAP, MAV, PAV) reflect averages across 3 question items that were rated on a scale of 1
(strongly disagree) to 5 (strongly agree). Standard errors of the mean (SEM) are shown in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>MG</th>
<th>PG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>21.53 (.83)</td>
<td>20.30 (.66)</td>
</tr>
<tr>
<td>MAP</td>
<td>4.47 (.08)</td>
<td>4.42 (.13)</td>
</tr>
<tr>
<td>PAP</td>
<td>3.77 (.19)</td>
<td>3.82 (.22)</td>
</tr>
<tr>
<td>MAV</td>
<td>3.74 (.18)</td>
<td>3.74 (.21)</td>
</tr>
<tr>
<td>PAV</td>
<td>3.48 (.22)</td>
<td>3.62 (.20)</td>
</tr>
</tbody>
</table>

**Manipulation Check.** Related-Samples Wilcoxon Signed Rank Tests were conducted between the Mastery and Performance Goal Adoption ratings, for each Goal condition because data failed normality in all 4 Shapiro-Wilk tests ($p_s < .014$). Participants under MG had a higher median for Mastery Goal Adoption (Median = 4) compared to the Performance Goal Adoption (Median = 3), $Z = 3.115$, $p < .003$, whereas those under PG had similar medians for the Mastery Goal Adoption (Median = 4) and Performance Goal Adoption (Median = 4.5), $Z = 0.406$, $p > .68$ (see Fig 2). Thus, although under MG participants were focused more on mastery than performance goals, participants under PG were focused similarly on both goals.
Fig 2. Median ratings of Mastery and Performance Goal Adoption as a function of Goal Induction Condition.

*First-test Performance.* We calculated first test accuracy in two different ways to evaluate whether goals influenced first test performance: 1) based on accuracy assigned by our program via the letter matching algorithm (i.e., a measure of our titration algorithm’s effectiveness), and 2) after removing problematic trials where the wrong accuracy had been assigned or responses were ambiguous (i.e., characterizing those trials that were included in the ERP and retest analyses; see Data Analysis for further details).

Independent t-tests comparing the goal induction conditions showed that general knowledge accuracy did not differ across conditions regardless of how it was calculated (all $ps > .63$), and regardless of whether the full sample was included in analysis or only the subsample that met the criterion trials for ERP analyses involving the delay factor. These findings support the view that our titration was successful in equalizing performance. Interestingly, one-sample t-tests against the pre-set titration level of .35, showed MGs led to higher accuracy as determined
by our titration ($M = .36$, $SEM = .03$), and after removal of problematic trials ($M = .37$, $SEM = .03$; $ps < .04$). PGs only led to higher accuracy after problematic trials were removed ($M = .36$, $SEM = .03$), $t(29) = 2.140$, $p < 0.05$, but not when accuracy was determined only by our titration algorithm, ($M = .36$, $SEM = .03$), $t(29) = 1.479$, $p > 0.14$. The same pattern was found for the smaller ERP sample that involved the delay factor under MGs ($ps < .03$), but for PGs accuracy scores did not differ from .35 ($ps > .89$). Thus, even though goals did not differentially influence first test accuracy, MGs generally led to better accuracy in comparison to the pre-set titration level even before the removal of problematic trials increased overall test accuracy.

Finally, because our titration algorithm relied on the level of question difficulty to maintain performance ≈.35, we also examined differences in mean difficulty between goals using independent t-tests. Focusing on the usable trials, there were no differences across goal condition in the mean level of difficulty when it came to overall mean question difficulty, mean difficulty of questions answered correctly, or answered incorrectly, for either the whole sample or the delay-analysis subsample ($ps > .67$).

**Retest Performance.** When considering the full sample, a 2 (Induced Goal: MG or PG) x 2 (Retest-Delay: Immediate or Delayed) ANOVA on proportion of errors corrected showed a main effect of Goal, $F(1, 58) = 4.257$, $p < 0.05$, and a main effect of Test-Delay, $F(1, 58) = 786.964$, $p < 0.001$, but no Goal by Test-Delay interaction, $F(1, 58) = 1.240$, $p > 0.26$. As shown in Figure 3, participants under MG corrected more errors overall ($MEAN = .60$, $SEM = .02$) than participants under PG ($M = .55$, $SEM = .02$), and across goals, participants corrected more errors at the immediate ($M = .75$, $SEM = .01$) than the delayed retest ($M = .40$, $SEM = .01$). Thus, MG induction appeared to lead to benefits in error correction compared to PG, regardless of retest.
delay, and both groups similarly showed a reduction in memory performance as a function of the longer delay.

When confirming these findings with the subsample used in the ERP analysis with the delay factor, the main effect of delay remained robust, $F(1, 46) = 584.947, p < 0.001$, but the main effect of goal was no longer significant ($p > .13$), although it was in the same direction numerically (MG: $M = .58, SEM = .02$; PG: $M = .54, SEM = .01$). Nor did goal interact with delay ($p > .45$). While this may be the result of a general reduction in power, it may also be that excluding individuals who did not have enough uncorrected items in the immediate condition and/or corrected items in the delay condition had the consequence of constricting the range of participant performance values and reducing group differences. Regardless, this result means that the ERP analyses involving delay as a factor should be interpreted in light of a lack of goal effect on behavior.

![Graph showing proportion of errors corrected for immediate and delayed conditions for MG and PG groups.](image-url)
Fig 3. Proportion of errors corrected at retest as a function of induced goal and retest delay (immediately after first test vs. 1-week delay). Error bars show standard errors of the mean (SEM).

**Accuracy Feedback**

In order to test the hypothesis that induction of PG goals might enhance attention to performance feedback, we analyzed ERP waveforms maximal over frontal (FRN, P3a) and parietal sites (P3b, LPP) previously shown to be sensitive to feedback valence, target expectancy and/or motivation (see Figure 4). Mean amplitudes of each waveform were submitted to a 2 (Induced Goal: MG or PG) x 2 (Valence: Positive or Negative) ANOVA. For the FRN, P3a and P3b, these means were centered around a peak-picked latency. For the broader LPP, we analyzed the mean amplitude from 500-1000 ms.

**FRN.** The FRN exhibited only a main effect of feedback valence, \( F(1, 58) = 29.429, p < 0.001 \). No other effects were significant (\( ps > .27 \)). As shown in Figure 4A, the FRN was more negative-going for the negative feedback to errors (\( M = -1.19, SEM = .22 \)) than for the positive feedback to corrects (\( M = -1.3, SEM = .31 \)). However, when the FRN was analyzed as a function of amplitude differences from the P2 (i.e. peak-to-peak P2\(^4\) to FRN differences), no effects were significant (\( ps > .49 \)), including the effects of feedback valence. Thus, despite the appearance of

---

\(^{4}\) Although the P2 was not the focus of our analysis, 2 (Induced Goal: MG or PG) x 2 (Accuracy: Corrects or Errors) ANOVA on peak-picked mean P2 amplitude at electrode Fz showed a main effect of Accuracy, \( F(1, 58) = 58.405, p < 0.001 \), in that the P2 was enhanced for positive feedback (\( M = 2.83, SEM = .33 \)) compared to negative feedback (\( M = 1.60, SEM = .28 \)). There was no main effect of Goal, \( F(1, 58) = 1.378, p > 0.24 \), but the Goal by Accuracy interaction approached significance, \( F(1, 58) = 3.472, p < 0.07 \). Exploration of this interaction using 5 Bonferroni tests with a significance of \( p < .01 \) indicated that the interaction seemed to be driven by a marginally greater difference between corrects and errors under MG (\( M = 1.54, SEM = .24 \)) compared to PG (\( M = .93, SEM = .22 \), \( t(58) = 1.863, p < .07 \).
more negative-going FRN activity in the PG condition (see Figure 4A), goals did not reliably influence FRN amplitude.

**P3a.** A main effect of feedback valence was also found for this waveform, $F(1, 58) = 5.941, p < 0.02$. However, as shown in Figure 4A, peak-picked mean P3a amplitude was more positive-going for the relatively “rare” positive feedback to correct answers ($M = 1.26, SEM = .32$), than compared to the more frequent negative feedback to errors ($M = .75, SEM = .23$). All other effects were not significant ($ps > .54$). When considering this waveform in relation to the preceding FRN (i.e. FRN to P3a peak-to-peak amplitudes), the main effect of feedback valence remained, $F(1, 58) = 4.285, p < 0.05$, in that FRN-to-P3a mean amplitudes were smaller for positive feedback ($M = 1.39, SEM = .40$) than negative feedback ($M = 1.94, SEM = .29$). No other significant effects were found ($ps > .70$). Thus, goals did not appear to influence P3a amplitude.

**P3b.** When analyzing the P3b maximal over parietal sites, we also found only a main effect of feedback valence, $F(1, 58) = 51.767, p < 0.001$. As shown in Figure 4B, the P3b was more positive-going for positive feedback following corrects ($M = 3.03, SEM = .34$) than negative feedback following errors ($M = 1.89, SEM = .23$) (see Figure 4B). Nonetheless, there was no evidence that the goal manipulation influenced P3b amplitude. All other effects were not significant ($ps > .43$).
Fig 4. Grand mean ERP waveforms time-locked to accuracy feedback and averaged as a function of induced goal and feedback type (negative feedback to incorrect answers [errors], positive feedback to correct answers [corrects]). A) Waveforms at Fz, highlighting the P2, FRN and P3a. B) Waveforms at CPz, highlighting the P3b and LPP.

LPP. Finally, although the amplitude of the LPP appears numerically higher overall in the PG condition (see Figure 4B) neither the overall goal effect ($p > .19$), nor the interaction between goal and accuracy ($p > .53$) reached significance. There was only a main effect of Accuracy, $F(1, 58) = 11.643, p < 0.002$, in that amplitudes were higher for positive feedback to corrects ($M = 2.52, SEM = .21$) than negative feedback to errors ($M = 2.11, SEM = .16$).
Learning Feedback

In the analyses that follow, we evaluate the hypothesis that a MG induction will result in increased engagement of neural regions involved in the successful encoding of learning feedback. We focus on regions within the frontal and lateral temporal cortex, and assess effects of induced goal and subsequent memory performance across hemisphere and region first after collapsing over delay, and then after including Retest Delay as a factor.

We note that each approach to the delay variable has its advantages and disadvantages. By collapsing over delay, we ensure that all participants can be included in analysis with sufficient trial counts. Importantly, these participants represent the sample where a significant error correction advantage found for participants in the MG induction condition. Also, the MG advantage did not interact with delay, supporting the value of collapsing over this factor for the ERP analysis. However, we must also consider that when collapsing, a larger percentage of later corrected items come from the immediate retest, where as a larger percentage of later uncorrected items come from the delay condition. Therefore, we also follow-up with the analysis that includes delay as a factor, although by doing so, our sample size and statistical power are correspondingly reduced.
Fig 5. Selected grand mean waveforms over fronto-temporal (electrodes: FT9, T7, FT10, and T8) and parieto-occipital (electrodes: PO3, O1, PO4, O2) regions for the first-test learning feedback following errors, as a function of induced goal, and sorted as a function of whether the error was later corrected or not on a later retest. Waveforms are averaged over retest delay. The bracket over electrode FT9 signifies the time window (500 – 1000 ms) of interest.

*Inferior Fronto-Temporal vs Parieto-Occipital Negativity*

The primary goal of this analysis was to examine whether we would replicate the achievement goal-related double dissociation of subsequent memory effects across inferior
fronto-temporal and parieto-occipital regions that was originally reported by Mangels and colleagues (2017). Specifically, in that study, fronto-temporal regions were differentially involved in encoding under MGs, whereas parietal-occipital regions are differentially involved in encoding under PGs, although there were no goal-related differences in behavioral outcomes on the later retest.

The present study primarily differed from Mangels and colleagues (2017) in that it included both an immediate and delayed retest, rather than just an immediate test, and used a between-subject rather than within-subject manipulation of goals. There were also some subtle differences in goal wording. Finally, whereas in that study, differential involvement of region as a function of goal was found despite no goal differences in behavioral outcomes, here, we have the opportunity to evaluate these neural effects in the context of significant behavioral differences. To this aim, these analyses employ a series of 2 (Induced Goal: MG or PG) x 2 (Memory: Corrected or Not Corrected) x 3 (Hemisphere: Left or Right) x 2 (Region: Fronto-temporal or Parieto-occipital) ANOVAs. Mangels and colleagues (2017) found effects from 400-800 ms after the onset of the learning feedback. In the present analysis, we focus on the 500-1000 ms period in order to make these analyses more comparable to the analysis of other sites of interest to our study.

When collapsing over retest delay, we found a significant main effect of Memory (Dm effect), $F(1, 58) = 20.607, p < 0.001$, where errors that were later corrected ($M = -1.06, SEM = .13$) elicited more negative-going waveforms during presentation of the learning feedback that were compared to errors that were not corrected ($M = -0.82, SEM = .13$) (see Figure 5). Although there was a trend toward a Memory by Hemisphere by Region interaction, $F(1, 58) = 2.842, p < 0.10$, there were no significant or marginal interactions that involved both Goal and Memory ($ps$
> .35), which were of the greatest relevance to the current hypothesis. The only effect involving Goal was a marginal Goal by Hemisphere\(^5\) interaction, \(F(1, 58) = 3.995, p < 0.06\), that subsumed a significant main effect of Hemisphere, \(F(1, 58) = 6.624, p < 0.014\). Post hoc analysis of the interaction with goal indicated that the overall pattern of more negative-going right hemisphere activity was only significant in the MG condition (MG: left: \(M = -0.79, SEM = 0.21\); right \(M = -1.38, SEM = 0.19\); PG: left: \(M = -0.75, SEM = 0.21\); right: \(M = -0.83, SEM = 0.19\)). Thus, although these finding replicate the sensitivity of these waveforms to later error correction, as well as their greater magnitude over the right hemisphere, it appears that during this time frame, neural activity across both the more posterior and anterior aspects of the ventral stream contributed equally to successful encoding in each of the goal conditions.

\[\text{Amplitude (µV)}\]

\[\begin{align*}
\text{Fronto-Temporal} & \quad \text{Parieto-Occipital} \\
\text{Left} & \quad \text{Left} \\
\text{Right} & \quad \text{Right}
\end{align*}\]

\[\text{MG} \quad \text{PG}\]

\(^5\) Additionally, there was a significant Hemisphere by Region interaction, \(F(1, 58) = 6.734, p < 0.013\), which post hoc comparisons indicated resulted from the pattern of greater right- than left-hemisphere negativity being confined to the parieto-occipital region (parieto-occipital: right: \(M = -1.42, SEM = 0.22\); left: \(M = -0.78, SEM = 0.25\); fronto-temporal region: right: \(M = -0.78, SEM = 0.18\); left: \(M = -0.76, SEM = 0.17\)).
Fig 6. Subsequent memory (Dm) effects at fronto-temporal and parieto-occipital sites from 500-1000 ms. A) Dm effects (later corrected-not corrected) as a function of Goal, Region and Hemisphere B) Dm effects as a function of Goal, Delay and Region.

Nonetheless, given that this analysis was designed to replicate the previous results of Mangels et al (2017), we explored whether significant Dm amplitudes might be observed for each Goal as a function of Region collapsed over hemisphere (as were found in that study), using one-sample t-tests with zero as the comparison value (i.e., zero would represent no difference in amplitude between later corrected and not correct items). As illustrated in Figure 6A, these Dm difference values were significantly different from 0 over fronto-temporal regions under MG ($M = -.39$, $SEM = .13$), however after Bonferroni correction, no other Dm values reached significance.

Results from this analysis suggest that we partially replicated the results of Mangels and colleagues (2017) with Dm effects over the fronto-temporal region only being apparent in the
MG condition, but not the PG condition. Only a single dissociation was observed here, however, unlike the double dissociation found in Mangels et al. (2017). Indeed, over the parieto-occipital region, not only did the induction of PG did not lead to a differential Dm effect, but the magnitude of the Dm effect did not differ from zero for either condition.

When including retest-delay as a factor, which reduced our sample size (see Data Analyses), learning feedback that was later remembered at retest similarly elicited more negative-going waveforms ($M = -1.03, SEM = .14$) than those that were forgotten ($M = -.78, SEM = .15$), $F(1, 46) = 16.416, p < 0.001$ (see Figure 6B). There was also a trend toward an interaction of this memory effect with retest delay, $F(1, 46) = 2.995, p < 0.10$, and a significant 3-way interaction between Memory, Hemisphere and Delay, $F(1, 46) = 4.822, p < 0.04$. Post hoc tests of the higher-order interaction indicated that Dm effects appeared at all hemispheres and retest-delays, except over the left hemisphere at the delayed retest (errors corrected: $M = -.82, SEM = .18$; errors not corrected: $M = -.83, SEM = .18$).

As with the full sample, there were no significant interactions involving both Memory and Goal (all $ps > .35$), but unlike when the full sample was employed, targeted analysis the magnitude of the Dm effects using t-tests against zero found that none of the 8 conditions representing goal, region and delay survived Bonferroni correction. In terms of other effects involving goal, only a Goal by Region by Delay interaction approached significance, $F(1, 46) = 3.134, p < 0.09$, but post-hoc tests found no significant comparisons involving the factor of Goals.

---

6 Other effects involving Region were also observed, including a marginal main effect of Region, $F(1, 46) = 2.874, p < 0.10$, marginal Memory by Region interaction, $F(1, 46) = 2.862, p < 0.10$, and a significant Region by Hemisphere interaction, $F(1, 46) = 11.701, p < 0.002$. Focusing on the significant interaction, post hoc analyses confirmed what was observed in the full sample. Specifically, the overall greater negativity over the right hemisphere is limited to the parieto-occipital region (parieto-occipital: right: $M = -1.45, SEM = .26$; left: $M = -.83, SEM = .28$; fronto-temporal: left: $M = -.76, SEM = .18$; right: $M = -.51, SEM = .19$). In addition, differences between regions were not found over the left, but were found over the right hemisphere.
Fig 7. Selected grand mean waveforms over a superior frontal (Fz), central (Cz), and parietal (Pz) electrode for the first-test learning feedback following errors, as a function of induced goal, and sorted as a function of whether the error was later corrected or not on a later retest. Waveforms are averaged over retest delay. The brackets over electrode Fz signify the early (200 – 300 ms), middle (500 – 1000 ms), and late (1000 – 1500 ms) time periods of interest. Electrodes that are filled-in white in were also included in analyses but are not illustrated in this figure.

*Superior Positive-going Waveforms: Early, Middle and Later Latencies.*
In this section, we tested whether a series of positive-going waveforms maximal over the superior midline, were sensitive to memory and the manipulation of achievement goals. We were particularly interested in whether the MG induction would differentially modulate early and later sustained frontal activity that past studies suggest may index more effortful and elaborative encoding. To this aim, we conducted separate 2 (Induced Goal: MG or PG) x 2 (Memory: Corrected or Not Corrected) x 3 (Region: Frontal, Central, or Parietal) x 3 (Electrode: Left, Middle or Right) ANOVAs on an early positive deflection (200-300 ms), and both the midlatency and later portion of a sustained positive waveform (see Figure 7). We added the 2-level factor of Retest Delay for a secondary analysis for each of the sustained time periods.

**Early positivity (200-300 ms).** This early potential exhibited an overall effect of Memory, $F(1, 232) = 14.791, p < 0.001$, where errors that were corrected ($M = 1.22, SEM = .11$) elicited more positive-going activity compared to errors that were not corrected ($M = 1.05, SEM = .10$). There were no other effects involving Memory, indicating that similar memory effects could be observed across both anterior and posterior sites during this period.

Although there were no interactions involving Memory and Goal, we did find a Goal by Region interaction, $F(1.352, 213. 820) = 4.970, p < 0.02, \epsilon = 0.68$. Post hoc comparisons indicated that goal differences in goals were only significant at the frontal regions. As shown in Figure 8A, this early frontal positivity was greater for participants under the MG induction.

---

7 We did not analyze this period as a function of retest delay with the smaller subsample because it was reasoned that the strength of the memory trace and stability against memory decay would be influenced more by mid-latency and later sustained activity.

8 In addition to the interaction of Region and Goal, a significant Region by Electrode interaction was found, $F(4, 232) = 6.575, p < 0.0013$, that subsumed a significant main effect of Electrode, $F(2, 232) = 10.478, p < 0.001$, and marginal effect of Region, $F(1.352, 213. 820) = 3.288, p < 0.07, \epsilon = 0.68$. Post hoc tests of the interaction found no differences across frontal left, right and midline electrodes (left: $M = 1.01, SEM = .22$; midline: $M = 1.12, SEM = .28$; right: $M = 1.11, SEM = .23$), ($ps > .43$). However, at central sites, activity peaked over the midline region (left: $M = 1.11, SEM = .16$; midline: $M = 1.94, SEM = .25$; right: $M = 1.46, SEM = .15$), whereas at parietal sites activity was greatest over central and right hemisphere sites compared to left (left: $M = .30, SEM = .19$; midline: $M = 1.03, SEM = .19$; right: $M = 1.11, SEM = .20$).
compared to the PG induction. At parietal sites no effects of goal were observed. Thus, as expected, goals modulated the early positivity in the expected direction, although this modulation did not interact with subsequent memory effects.

A)

![Bar chart showing amplitude (µV) across frontal, central, and parietal sites for MG and PG conditions.]

B)

![Bar chart showing amplitude (µV) across frontal, central, and parietal sites for MG and PG conditions.]

C)
Fig 8. Average amplitude of superior positive-going ERP waveforms time-locked to learning feedback at frontal, central and parietal sites for the full sample (i.e., collapsed over retest delay) during the A) early (200-300 ms), B) middle (500-1000 ms), and C) later (1000-1500 ms) periods, as well as for (D) the retest-delay analysis subsample for the later period (1000-1500 ms). Waveforms are collapsed over later memory outcomes (i.e., later corrected and not corrected) given that no interactions were found between Goal and Memory, as well as across the three electrodes within each region. Error bars show the standard error of the mean (SEM).
Mid-Latency Positivities (500-1000 ms). In this time period, we found a significant effect of Memory, $F(1, 232) = 32.812, p < 0.001$, in the full sample. Errors that were later corrected elicited more positive-going activity in this time frame ($M = 1.33, SEM = .11$) compared to errors that were not corrected ($M = 1.07, SEM = .10$). We also found a Memory by Region interaction, $F(1.228, 202. 227) = 4.747, p < 0.03, \varepsilon = 0.61$, which post-hoc tests indicated arose from Dm effects occurring over frontal sites (errors corrected: $M = .18, SEM = .19$; errors not corrected: $M = -.27, SEM = .20$) and central sites (errors corrected: $M = 1.67, SEM = .15$; errors not corrected: $M = 1.41, SEM = .14$), but not parietal sites (errors corrected: $M = 2.13, SEM = .16$; errors not corrected: $M = 2.06, SEM = .15$; see also Figure 7).

Dm effects did not interact with any other variables, including Goal ($ps > .25$). Moreover, although the overall distribution of positivity also differed as a function of Region and Hemisphere in a pattern highly similar to that described for the early waveform\textsuperscript{9}, unlike the early positive deflection, Goal did not modulate the distribution of these effects (all $ps > .10$; see Figure 8B). Thus, induction of an MG goal did not lead to increased frontal activity during this period, at least when considering the full sample, even as the distribution of Dm effects shifted to a more frontal focus overall.

When adding the factor of retest delay to this analysis, we continued to find enhanced positivity for items later corrected ($M = 1.27, SEM = .12$), compared to those later not corrected ($M = .99, SEM = .11$), $F(1, 184) = 30.290, p < 0.001$, regardless of region, electrode, goal or Hemisphere. We found a Region by Electrode interaction, $F(3.138, 202. 227) = 5.158, p < 0.003, \varepsilon = 0.78$, that subsumed significant effects of Region, $F(1.445, 202. 227) = 48.888, p < 0.001, \varepsilon = 0.72$, and Electrode, $F(2, 232) = 5.003, p < 0.009$. Post hoc tests on the source of the interaction indicated that there were no electrode differences within the frontal region (left: $M = -.002, SEM = .17$; midline: $M = .01, SEM = .23$; right: $M = -.14, SEM = .22$, ($ps > .27$), or central regions (left: $M = 1.43, SEM = .16$; midline: $M = 1.72, SEM = .21$; right: $M = 1.46, SEM = .16$ ($ps > .09$), but over parietal region, waveforms at the right electrode were more positive-going than at the other sites, which did not differ from each other (left: $M = 2.40, SEM = .17$; midline: $M = 2.25, SEM = .18$; right: $M = 1.62, SEM = .20$).

\textsuperscript{9} We found a Region by Electrode interaction, $F(3.138, 202. 227) = 5.158, p < 0.003, \varepsilon = 0.78$, that subsumed significant effects of Region, $F(1.445, 202. 227) = 48.888, p < 0.001, \varepsilon = 0.72$, and Electrode, $F(2, 232) = 5.003, p < 0.009$. Post hoc tests on the source of the interaction indicated that there were no electrode differences within the frontal region (left: $M = -.002, SEM = .17$; midline: $M = .01, SEM = .23$; right: $M = -.14, SEM = .22$, ($ps > .27$), or central regions (left: $M = 1.43, SEM = .16$; midline: $M = 1.72, SEM = .21$; right: $M = 1.46, SEM = .16$ ($ps > .09$), but over parietal region, waveforms at the right electrode were more positive-going than at the other sites, which did not differ from each other (left: $M = 2.40, SEM = .17$; midline: $M = 2.25, SEM = .18$; right: $M = 1.62, SEM = .20$).
Additionally, we found a main effect of Delay, $F(1, 184) = 4.065, p < 0.05$, where positivity was greater overall under the delayed ($M = 1.18, SEM = .11$) compared to immediate retest ($M = 1.07, SEM = .11$), but no interactions between Delay and Memory that might suggest that the predictors of successful encoding during this time period differed across midline sites as a function of test delay. Although there was a trend for a Region by Delay interaction, $F(1.378, 163. 366) = 2.946, p < 0.08, \varepsilon = 0.69$, exploration of this interaction indicated that the pattern of distribution of this waveform was overall quite similar across retest delays, as well as to the general patterns observed when the full sample was included.\(^{11}\)

However, in the subsample used for this analysis, the Goal by Region interaction now approached significance, $F(1.375, 163. 366) = 2.745, p < 0.10, \varepsilon = 0.69$. Over frontal regions, MG induction elicited numerically greater positivity compared to the PG induction (MG: $M = .23, SEM = .32$; PG: $M = -.50, SEM = .29$), whereas over parietal sites the PG induction elicited numerically greater positivity over the parietal regions compared to MG (PG: $M = 2.24, SEM = .24$; MG: $M = 1.74, SEM = .26$), however these differences did not survive Bonferroni correction. No other effects of Goal met or approached significance ($ps > .20$). Thus, similar to the analyses across test-delay, the positive midline waveforms were sensitive to memory outcomes, but not significantly influenced by the manipulation of achievement goals.

\(^{10}\) Although there was a 5-way interaction that just met the criteria for significance, $F(4, 184) = 2.610, p < 0.05$, we opted not to explore this further because any goal effects that might emerge, but be specific only to single electrode, were considered less reliable than those that occurred across a broader region.

\(^{11}\) We found a Region by Electrode interaction, $F(2.971, 163. 366) = 4.409, p < 0.007, \varepsilon = 0.74$, that subsumed a main effect of Region, $F(1.375, 163. 366) = 35.906, p < 0.001$, and Electrode, $F(1.723, 163. 366) = 4.154, p < 0.03, \varepsilon = 0.86$. Post hoc evaluation of the interaction confirmed that there were no differences across the electrodes making up the frontal region (left: $M = -.10, SEM = .19$; midline: $M = -.10, SEM = .27$; right: $M = -.20, SEM = .25$) or central region (left: $M = 1.36, SEM = .17$; midline: $M = 1.75, SEM = .24$; right: $M = 1.46, SEM = .19$), but at the parietal region, the midline site was more positive than the right with no differences compared to the left site (left: $M = 2.22, SEM = .19$; midline: $M = 2.24, SEM = .20$; right: $M = 1.51, SEM = .23$).
Later sustained positive waveform (1000-1500 ms). An ANOVA with the full sample that focused on the later sustained portion of these waveforms confirmed presence of a significant Dm effect, $F(1, 232) = 21.049, p < 0.001$ (errors later corrected: $M = .86, SEM = .08$; errors later not corrected: $M = .69, SEM = .08$). However, in this time period, the overall Memory effect was qualified by a significant Memory by Region\textsuperscript{12} interaction, $F(1.359, 205. 355) = 6.646, p < 0.007, \epsilon = 0.68$, which post hoc comparisons indicated was driven by the presence of significant Dm effects at both frontal sites (errors corrected: $M = .84, SEM = .16$; errors not corrected: $M = .45, SEM = .15$) and central sites (errors corrected: $M = 1.33, SEM = .12$; errors not corrected: $M = 1.13, SEM = .12$), but not at parietal sites (errors corrected: $M = .41, SEM = .12$; errors not corrected: $M = .48, SEM = .11$; see also Figure 7).

Although there were no main effects of Goal or interactions involving Goal and Memory (ps > .29), we did find a significant Goal by Region interaction $F(1.343, 205. 355) = 4.610, p < 0.03, \epsilon = 0.57$. The mean amplitudes plotted in Figure 8C suggest that MGs led to greater positivity over the frontal regions compared to PG, whereas PGs led to greater positivity over the parietal regions compared to MG, however these differences were not significant after Bonferroni correction.

When including retest delay in the analysis, there was significant evidence of both the overall Dm effect, $F(1, 184) = 22.872, p < 0.001$, and Memory by Region\textsuperscript{13} interaction, $F(1.508,$

\textsuperscript{12} Region also interacted with Electrode, $F(3.513, 205. 355) = 4.318, p < 0.004, \epsilon = 0.88$, with interactions involving Region also subsuming a main effect of this factor, $F(1.343, 205. 355) = 11.038, p < 0.001, \epsilon = 0.57$. Post hoc tests confirmed that interactions along the electrode factor resulted from there being no amplitude differences across the electrodes making up the frontal region (left: $M = .57, SEM = .14$; midline: $M = .78, SEM = .18$; right: $M = .59, SEM = .18$), or parietal region (left: $M = .63, SEM = .13$; midline: $M = .46, SEM = .14$; right: $M = .25, SEM = .15$), but at the central region, amplitude of the sustained positivity being lower at the left compared to central site.

\textsuperscript{13} As with the analysis of the full sample, there was a significant effect of Region, $F(1.339, 177.064) = 9.573, p < 0.002, \epsilon = 0.67$, as well as a Region by Electrode interaction, $F(3.408, 177.064) = 2.865, p < 0.04, \epsilon = 0.85$. Post hoc analysis of this interaction indicated that there were no differences across left, midline, or right frontal electrodes (left: $M = .62, SEM = .17$; midline: $M = .71, SEM = .20$; right: $M = .53, SEM = .21$), or parietal electrodes (left: $M = \ldots$
177.064) = 9.507, \( p < 0.002, \varepsilon = 0.75, \) showing that Dm effects were focused over frontal and central (rather than parietal sites). No other significant effects involving Memory were found.\(^{14}\)

Interestingly, similar to the analyses with the full sample, in this retest-delay subsample, we found a Goal by Region interaction, \( F(1.339, 177.064) = 6.254, \ p < 0.01, \varepsilon = 0.67, \) that did not interact further with delay. However, unlike in the full sample, here MGs also led to significantly greater frontal positivity compared to PG, whereas PG led to significantly greater parietal positivity compared to MG (see Fig. 8D), even after Bonferroni correction for multiple post-hoc comparisons. No other significant effects involving Goal were found (\( ps > .19 \)).

**Inferior Temporo-Parietal Negativity.**

To examine the hypothesis that inferior temporo-parietal sites would be sensitive to subsequent memory (e.g., Butterfield & Mangels (2003), but not the manipulation of achievement goals, we conducted a 2 (Induced Goal: MG or PG) x 2 (Memory: Corrected or Not Corrected) x 3 (Hemisphere: Left or Right) x 2 (Electrode) ANOVA on mean amplitudes of mid-latency (500-1000 ms) and later (1000-1500 ms) sustained waveforms observed over in these regions (see Figure 9).

\[^{14}\] Although a Memory by Region by Electrode by Delay interaction approached significance, \( F(4, 184) = 2.148, \ p < 0.08, \) unpacking this 4-way interaction involved 18 comparisons, and after Bonferroni correction, the only remaining significant Dm effects were at Fz during the immediate retest, and at F3 during the delayed retest.
Fig 9. Selected grand mean waveforms over inferior temporo-parietal regions for the first-test learning feedback following errors, as a function of induced goal, and sorted as a function of whether the error was later corrected or not on a later retest. Waveforms are averaged over retest delay. The brackets over electrode TP7 signify the middle (500 – 1000 ms), and late (1000 – 1500 ms) time periods of interest.

500-1000 ms. As expected, a significant Dm effect emerged in this time window, $F(1, 58) = 27.676, p < 0.001$, such that waveforms were more negative for errors that were corrected compared to errors that were not corrected (later corrected: $M = -1.73, SEM = .16$; later not corrected: $M = -1.34, SEM = .16$; see also Figure 9). Although Memory also interacted with Electrode site, $F(1, 58) = 25.639, p < 0.001$, post hoc analyses indicated that Dm effects were significant at both sites, but simply larger at the more inferior pair of electrodes (TP9/10; see also
Most importantly, however, no effects involving the manipulation of achievement goal were observed (ps > .12), indicating that the contribution of this region to successful encoding was similar regardless of induced goal.

Including delay as a factor in this analysis did not influence either of the above effects involving memory (ps < .002), but with this subsample, the Goal by Memory interaction showed a weak trend toward significance, $F(1, 46) = 2.971, p < 0.10$. However, no differential effects of goal were found after Bonferroni correction for multiple comparisons (see Figure 10A). Additionally, although Memory and Retest Delay did not interact, indicating that this activity was predictive of subsequent memory regardless of whether memory was tested immediately or at a 1-week delay (see also Butterfield & Mangels, 2003), there was an overall effect of the Delay, $F(1, 46) = 6.095, p < 0.02$. Items tested on the delayed retest elicited more negative waveforms at these sites overall, regardless of subsequent memory performance (immediate retest: $M = -1.39$, $SEM = .19$; delayed retest: $M = -1.57$, $SEM = .18$).

There was also a significant main effect of Hemisphere, $F(1, 58) = 10.646, p < 0.003$, where mean amplitudes were more negative over to right hemisphere ($M = -2.00$, $SEM = .20$) compared to left ($M = -1.12$, $SEM = .20$) hemisphere, and a significant main effect of Electrode, $F(1, 58) = 127.267, p < 0.001$, where mean amplitudes across both regions were more negative over inferior temporo-parietal sites ($M = -2.41$, $SEM = .20$) compared to more superior temporo-parietal ($M = -1.86$, $SEM = .14$) sites.

A Memory by Hemisphere by Electrode interaction approached significance in this subsample, $F(1, 46) = 3.040, p < 0.09$, which post hoc comparisons indicated was the result of the Dm effects for the superior electrodes (TP7/8) being significant over the right hemisphere (TP8), but not the left (TP7). Additionally, there were significant main effects of Hemisphere, $F(1, 46) = 4.845, p < 0.04$, waveforms were more negative-going over the right ($M = -1.80$, $SEM = .23$) compared to left ($M = -1.16$, $SEM = .23$) hemisphere, and a significant main effect of Electrode, $F(1, 46) = 118.100, p < 0.001$. Mean amplitudes across both regions were more negative over inferior ($M = -2.34$, $SEM = .23$) compared to superior ($M = -1.62$, $SEM = .16$) electrode pairs.
A)

B)

C)
Fig 10. Average amplitude of inferior temporo-parietal ERP waveforms (collapsed over electrode and hemisphere) time-locked to learning feedback during (A) 500-1000 ms for the full sample (i.e., collapsed over retest delay), (B) 500-1000 ms for the delay analysis subsample, (C) 1000-1500 ms for the full sample (i.e., collapsed over retest delay), and (D) 1000-1500 ms for the delay analysis subsample. Error bars show the standard error of the mean (SEM).

Interestingly, as shown in Figure 10B, a post hoc analysis of a significant Goal by Delay interaction, $F(1, 46) = 5.257, p < 0.03$, indicated that this pattern of greater negativity for items tested at the delayed test was only significant in the MG condition ([MG: immediate retest: $M = -1.55$, $SEM = .27$; delayed retest: $M = -1.90$, $SEM = .27$]; [PG: immediate: $M = -1.23$, $SEM = .25$; delayed: $M = -1.24$, $SEM = .24$]), suggesting the greater overall engagement of these regions in the items that MG subjects retrieved during the delay. Since it was not clear why this interaction appeared, we explored whether significant Dm amplitudes might be observed for each Goal as a function of Delay using one-sample t-tests with zero as the comparison value (i.e., zero would represent no difference in amplitude between later corrected and not correct items). After
Bonferroni correction, we found Dm difference values were significantly different from 0 for both retest delays under MG, but only at the immediate retest under PG. Thus, inferior-temporal negativity may have generally supported learning under MG, but only supported learning given a shorter retest delay under PG. No other effects involving Goal were found (ps > .18).

1000 – 1500 ms. At this later time window, a significant Dm effect also emerged, $F(1, 58) = 18.453, p < 0.001$, in which errors that were later corrected elicited more negative-going waveforms in this region compared to errors that were not corrected (errors later corrected: $M = -1.35, SEM = .13$; errors later not corrected: $M = -1.04, SEM = .12$). Memory also interacted with Electrode site, $F(1, 58) = 21.127, p < 0.001$, however in this later time frame, only the inferior sites exhibited robust Dm effects (errors later corrected: $M = -2.03, SEM = .17$; errors later not corrected: $M = -1.56, SEM = .16$). The Dm effects at the superior sites were not significant (errors later corrected: $M = -.66, SEM = .12$; errors not corrected: $M = -.52, SEM = .11$).17

Interestingly, a marginally significant interaction of Goal by Memory by Electrode $F(1, 58) = 3.429, p < 0.07$, suggested that Goal might moderate these differential site effects. However, post hoc comparisons confirmed that the MG and PG conditions exhibited similar patterns of memory effects; both goal conditions demonstrated significant Dm effects over the more inferior sites, but not at the more superior sites. No other effects involving Goal were found (ps > .20).

The majority of these effects remained significant when including Retest Delay as a factor, including the significant overall effect of Memory, $F(1, 46) = 15.966, p < 0.001$ (errors later corrected: $M = -1.35, SEM = .15$; errors later not corrected: $M = -1.03, SEM = .15$), and the finding that these effects were only significant at the more inferior pair of temporo-parietal sites

---

17 There was also a main effect of Electrode, $F(1, 58) = 111.231, p < 0.001$. Mean amplitudes across both temporo-parietal regions were more negative over the more inferior pair of electrodes ($M = -1.79, SEM = .16$) compared to the more superior pair ($M = -0.60, SEM = .11$).
Additionally, in the delay analysis subsample, a significant Memory by Hemisphere interaction emerged, $F(1, 46) = 8.481, p < 0.007$, whereby only Dm effects over the right temporo-parietal region survived post hoc analysis ([Right: errors later corrected: $M = -1.46, SE = .18$; errors later not corrected $M = -1.14, SE = .18$]; [Left: errors corrected: $M = -1.21, SE = .20$; errors not corrected $M = -0.90, SE = .19$]). In this later window, we also found a main effect of Retest-Delay, $F(1, 46) = 3.511, p < 0.07$, and a marginal Goal by Retest-Delay interaction, $F(1, 46) = 3.663, p < 0.07$. As with that earlier analysis, waveforms associated with items tested on the delayed retest demonstrated overall more negative-going activity over this region compared to items tested immediately (immediate: $M = -1.11, SEM = .16$; delayed: $M = -1.27, SEM = .14$). When examining the marginal interaction with goals, however, post hoc comparisons did not support any significant differences of Goal conditions as a function of Delay or Delay as a function of Goal. No other effects involving the factor of Goal approached significance ($p > .10$).

**Discussion**

The present study examined the neural mechanisms that underlie learning following errors as a function of goals, and in particular, the influence of goals on error correction as a function of retest delay. In support of the benefits of MG induction on engagement with learning, memory for the learning feedback (i.e. error correction) was better in the MG condition compared to the PG condition. However, in contrast to our predictions, we found no differential

---

18 A significant effect of Electrode in the Retest Delay subsample, $F(1, 46) = 36.673, p < 0.001$, revealed that waveforms were more negative-going over the inferior sites ($M = -1.81, SEM = .18$) compared to superior sites ($M = -0.57, SEM = .12$) within the temporo-parietal regions. Additionally, a significant effect of Hemisphere in this subsample, $F(1, 46) = 33.733, p < 0.001$, revealed that waveforms were generally more negative-going over the right ($M = -1.31, SEM = .18$) compared to left ($M = -1.07, SEM = .18$) hemisphere in this region.
benefits for the MG induction in terms of error correction at the delayed as compared to immediate retest. In other words, a MG environment did not appear to provide any particular buffer against general declines of memory that occurred for both groups as a function of a 1-week delay.

Based on goal-setting theory (see Locke & Latham, 2002, 2006), we had proposed that PGs may enhance attention to negative feedback compared to MGs given that they are thought to provoke concerns about one’s own ability (Ames & Archer, 1988). As for MGs, we expected greater attention and effort compared to PGs to the learning feedback given their focus on learning. However, we did not find that the PG induction differentially enhanced attention to the accuracy feedback. ERP correlates of bottom-up (FRN, P3a) and top-down (P3b, LPP) attention exhibited expected effects of feedback valence, but none of these potentials were enhanced under inducement of a PG state. These null effects replicating the findings of a previous study of induced achievement goals (Mangels et al., 2017), but are in contrast to findings of motivation-based modulation of feedback-elicited waveforms by trait (personal) achievement goals (Mangels et al. 2006; see also Lee & Kim, 2014).

We did find evidence that MG differentially modulated the neural response to learning feedback, however. Specifically, inducing MGs led to enhanced early and late positive waveforms over superior frontal electrode sites compared to PGs. Activity at these sites was also associated with successful encoding of that feedback (i.e., enhanced when that learning feedback was later successfully retrieved on a retest compared to when it was not retrieved). In contrast, inducing PGs differentially enhanced a late superior parietal positivity, however unlike the frontal positivity, the amplitude of this parietal positivity was unrelated to memory outcomes on the retest.
Successful encoding of the corrective feedback also enhanced negative-going waveforms over multiple electrode sites proximal to the visual ventral stream, from parieto-occipital sites, putatively involved in more perceptual processes, to inferior temporal-parietal and more fronto-temporal sites that have previously exhibited sensitivity to semantic, conceptual processing. As expected, goal state did not influence the inferior temporal parietal activity, but somewhat surprisingly, the double dissociation between MG-modulation of fronto-temporal Dm activity and PG-modulation of parieto-occipital Dm effects, as observed in Mangels et al. (2017), was not clearly replicated here. Only limited evidence for a single dissociation was found, with MGs leading to significant Dm effects over the fronto-temporal region, but no difference as a function of goal at the parieto-occipital region.

In the following sections, we first discuss the benefits of MG on error correction in comparison to PG in relation to past research, including past successes and failures to find significant effects of achievement goal inductions on behavioral measures. Then, we will address cognitive processes that may be indexed by the superior frontal positivity and superior parietal positivity that exhibited a dissociation between MG and PG conditions. This will then be considered in light of the similar effects of both goal conditions on inferior tempororo-parietal waveforms. Finally, we will discuss possible reasons for the failure to fully replicate Mangels and colleagues (2017), and the general lack of differences between goals during accuracy feedback.

*Error Correction*
In the present study, inducing MGs led to benefits in error correction across both the immediate and week-delayed retest compared to inducing PGs. Thus, focusing participants on developing their knowledge and not on trying to outperform others boosted learning even when participants were retested a week later compared to only focusing participants on doing better than others. A recent meta-analysis also showed benefits of MGs on general task performance compared to PGs (Van Yperen et al., 2015) and other studies showed similar benefits of MG compared to PG on working-memory tasks (Avery & Smillie, 2013) suggest that PGs may impair task performance because concerns regarding performance may deplete working memory resources unlike MGs (Crouzevialle et al., 2015). These findings are also broadly consistent with studies showing that perceptions of an MG classroom environment are positively associated with greater persistence when faced with difficulty, whereas the opposite relationship was shown when considering perceptions of a PG environment (Wolters, 2004).

It is important to note that participants under both goal inductions reported to focus on mastery goals, but those under PGs were also focused to a greater extent on performance goals. The main difference in goal adoption between the MG and PG conditions was that, in the MG condition, adoption of performance goals was suppressed. Thus, even in a task context where one is making repeated errors, it may be possible to maintain a focus on learning goals regardless of whether the environment is oriented toward PG or MG, however to the extent that performance goals are suppressed, one may be more able to direct that orientation toward learning-relevant task information in the context of repeated negative performance outcomes. In other words, we cannot rule out the possibility that it was the suppression of performance goals under MGs compared to PGs that was the most critical in boosting learning for the MG group.
 Nonetheless, our findings of significant learning benefits in the context of a MG environment (compared to a PG environment) support the general consensus that giving students a mastery orientation is more effective for learning success in challenging situations where failure is experienced. However, not every study that manipulated achievement goals has found behavioral benefits of MG compared to PG (Avery et al., 2013; Mangels et al., 2017; Murayama & Elliot, 2011b). The extent to which MG inductions benefit memory processes may depend on subtle differences in task and instruction parameters. For example, Avery and colleagues (2013) showed benefits in math problem solving when participants also had to concurrently complete a working-memory task under PG compared to MG. However, participants under both PG and MG were informed throughout the task that they accomplished the PG or MG goal assigned to them, and this feeling of competence may have preferentially boosted performance under PG (Cianci, Klein, & Seijts, 2010; see also Crouzevialle et al., 2015).

Importantly, using the same task as the present study, Mangels and colleagues (2017) did not show differences between goals in error correction. Although both studies broadly defined PGs in terms of normative comparison (i.e. doing better than others), goals were not induced in the same manner and MGs were also not based on the same operational definitions. The present study not only included goal language in the MG and PG inductions but also included differences in evaluation that spoke to the type of information (i.e. how well knowledge was developed under MG or scores in comparison to others under PG) that participants in each goal condition would receive at the end of the task. In addition, the MG induction also included language to suppress adoption of PGs. As for operational definitions of MGs, the present study defined MGs in terms of developing knowledge, whereas Mangels and colleagues (2017) defined MGs in terms of learning information that was interesting and useful. Indeed, a meta-analyses on
differences in personal achievement goal definitions shows outcomes depend on how goals are defined (see Hulleman, Schrager, Bodmann, & Harackiewicz, 2010). However, differences in goal induction definitions have not been explored given that there are only few studies available within each type of definition (see Van Yperen, Blaga, & Postmes, 2015). The present study also manipulated goals at the retest, where the effects of goals due to differences in encoding versus retrieval cannot be easily disentangled and goals may have influenced both encoding and retrieval processes. Finally, the present study used a between-subject design where participants answered all questions either under a consistent goal (MG or PG), whereas Mangels and colleagues (2017) used a within-subject design where participants answered half the questions under MG and the other half under PG. The longer and more consistent induction period in the present study may also be responsible for its stronger effects. Further research is needed to examine the task and instructional variables that optimize learning outcomes for individuals in mastery (or performance) goal environments.

Although MGs led to benefits in error correction when we collapsed across retest-delay, there were no benefits in errors correction in the smaller sample when we included retest-delay as a factor. Examination of the source of this difference indicates that in the subsample, the exclusion of MG participants who had corrected most of their errors at the immediate retest (i.e., because of low trial counts in the uncorrected bin), reduced the overall advantage of this group. Further research is needed to examine the influence of goals on within-subject variance as opposed to only examining goals as a between-subject factor since MGs may not always lead to similar behavioral benefits for all individuals. Indeed, Urdan and Schoenfelder (2006) warn that MGs may not be sufficient to promote learning when it comes to students who rely on
maladaptive learning strategies. Thus, MGs may lead to the use of successful learning strategies if students are also aware which learning strategies are best.

_Learning Feedback: Superior Frontal & Parietal Positivity_

When it comes to processing of the learning feedback following errors, MGs led to greater superior frontal positivity during the early time window (200-300 ms) in comparison to PGs in the full sample. The view that this indexes enhanced orienting of attention in order to more deeply encode the correct answer is consistent with past research showing that early frontal positivity is enhanced when participants were instructed to make deeper, semantic judgments of an upcoming stimuli, compared to when they were asked to make shallower, perceptual judgments (e.g. word length) (Guo et al., 2004; Wieser & Wieser, 2003). Similarly, Blanchet and colleagues (2007) showed early enhancement over frontal regions when participants were presented with word lists that could be memorized using organizational strategies (i.e. grouping semantically similar words) reflecting use of self-initiated strategies or when participants were instructed to memorize words using a semantic organization strategy compared to when they were asked to memorize semantically unrelated words. The authors concluded that this early positivity may reflect the influence of top-down processing where attention is directed to semantic processing of upcoming stimuli. Thus, in the present study, focusing participants on developing their knowledge under the MG induction may have led to greater attentional orienting to the correct answer following an error compared to focusing participants on doing better than others in the PG induction.
MGs also led to greater superior frontal positivity during the later time window (1000-1500 ms) in comparison to PGs, although this effect was more evident for analysis of the smaller sample that involved retest delay as a factor. Nonetheless, differential enhancement of this potential under MGs approached significance in the larger sample when we collapsed across retest delay. Unlike early frontal positivity, later frontal positivities have been examined in relation to memory outcomes and are generally more positive for words that were successfully encoded using elaborative, associative strategies, as rather than more item-specific or rote-rehearsal strategies (e.g., Fabiani, Karis, & Donchin, 1990; Kamp, Bader, & Mecklinger, 2017; Kamp & Zimmer, 2015; Liu, Rosburg, Gao, Weber, & Guo, 2017; Mangels, Picton, & Craik, 2001). In the present study, later frontal positivity was similarly enhanced for errors that were later corrected across both the immediate and delayed-retest compared to errors that were not corrected, supporting the view that the integration of the corrective learning feedback with the question benefits from elaborative, associative processing as well. Taken together with the finding of overall greater positivity in this region in the MG condition, this suggests that MGs may have also led to the greater use of elaborative associative strategies compared to PGs.

Interestingly, the lack of an interaction between goal condition and subsequent memory at either the early or late positivity suggests that both MG and PG participants engaged in similar strategies and that the difference between goal conditions was more quantitative than qualitative in nature. Interestingly, Fabiani and colleagues (1990), in their seminal ERP study on the relationship between subsequent memory and encoding strategies (i.e. elaboration versus rote-rehearsal), also found overall greater frontal positivity under elaboration compared to rote-rehearsal for both later remembered and forgotten items. It may be that actively generating responses to a question engaged some elaborative processing of the correct answer all
participants, regardless of goal (see Metcalfe, 2017), however inducing MGs may have fostered overall deeper and more sustained elaborative processing, both within a trial, as suggested by the lack of robust goal differences in the mid-latency period (500-1000 ms), and perhaps throughout the task as a whole, compared to PGs.

Unlike later frontal positivity, later parietal positivity that was enhanced under PG was not predictive of memory outcomes. Although some studies have found a mid-latency (i.e., ~400-800 ms) parietal positivity to be predictive of subsequent memory (for a review see Paller & Wagner, 2002), particularly for items that are encoded with rote-rehearsal (Fabiani et al., 1990), or retrieved on the basis of more item-specific processes (i.e., Mangels, Craik, & Picton, 2001), few studies have addressed encoding-specific effects at the later latency (1000-1500 ms) where current effects were found. In one recent study, however, Liu and colleagues (2017) showed greater late parietal positivity under an elaborative encoding condition that was more enhanced for later unsuccessfully retrieved items compared to successfully retrieved items. Thus, although other studies (e.g., Fabiani et al., 1990; Schott, Richardson-Klavehn, Heinze, & Düzel, 2002) have not explored or shown differences in later parietal positivity as a function of encoding strategies, greater parietal enhancement under PGs compared to MGs may reflect lower engagement of elaborative processing. Indeed, to the extent that this later positivity captures some of the processes associated with the mid-latency parietal activity described above, it may be that participants under PGs participants relied more on rote-rehearsal or other item-specific processing of the correct answer, as opposed to elaborative processing, which would be less diagnostic of error correction on this general knowledge task.

*Learning Feedback: Inferior Temporo-Parietal Negativity*
In line with Butterfield and Mangels (2003), the amplitude of a negative-going waveform over inferior temporo-parietal regions thought to reflect early semantic processing (Gold et al., 2005; Gold & Buckner, 2002) predicted learning for both the immediate and week-later retest delay. However, although goals did not influence either the magnitude of the memory effect, or negativity over temporo-parietal regions when including the retest delay as a factor, items tested at the delay compared to the immediate were associated with greater negativity during the mid-latency period (500-1000 ms) for the MG condition only. Because individuals did not know at the time of encoding what items would be tested immediately or at the delay, it is not immediately clear what might have driven this difference. However, when further investigating this effect through analyses of Dm effect magnitude, we found significant Dm effects for both retest delays under MG, but only at the immediate retest under PG. These results suggest that processes indexed by the inferior temporo-parietal negativity appear to have supported learning under MGs regardless of retest delay, but only supported learning under PGs at the immediate retest.

With regard to the processes that might underlie these differential effects, it is important to note that they were only evident during the middle latency period. This positions the timing of these effects as between the early frontal positivity putatively associated with attentional orienting and the later sustained frontal positivity putatively associated with elaborative, associative encoding. As such, it continues to be a good candidate for initial bottom-up semantic processing of the correct answer (Gold et al., 2005; Gold & Buckner, 2002), that then makes the answer available for further elaborative processes indexed by the late frontal positivity. Nonetheless, these results would also imply that these basic bottom-up processes are less predictive of participants’ ability to retain the correct answer over a 1-week period when
encoding had taken place in a PG goal environment. It may be that under PGs, encoding of a delay-resistant memory relies on other processes and corresponding potentials that were not measured in the current study.

Learning Feedback: Inferior Fronto-Temporal & Parieto-Occipital Negativity

Although Mangels and colleagues (2017) found goals to differentially modulate memory over fronto-temporal and parieto-occipital regions, the present study did not fully replicate these results. Mangels and colleagues (2017) had found that from 400-800 ms MGs, but not PGs, led to differences in memory over fronto-temporal regions thought to index greater semantic processing of the learning feedback under MGs. In the present study, a similar single dissociation was found when examining the magnitude of Dm effects in the full sample, from 500-1000 ms, although this difference was no longer apparent when retest delay was included as a factor. As for PGs, Mangels and colleagues (2017) showed differences in memory to appear over parieto-occipital regions, but not for MGs thought to reflect visual processing of the learning feedback. However, in the present study, analysis of Dm effect magnitude indicated that activity over the parieto-occipital region was not diagnostic of subsequent memory performance for either goal condition. Thus, not only did both goals lead to similar engagement of the processes indexed by these parieto-occipital potentials, but unlike Mangels and colleagues (2017), they did not appear to preferentially support successful learning in this study.

In summary, evidence of a partial replication of Mangels and colleagues (2017) was apparent in the present study suggesting that deeper semantic processing of the learning feedback may have generally supported learning across both retest delays under MGs, but not PGs.
Although differences between goals were not observed at the parieto-occipital sites, we cannot rule out the possibility that differences between goals might have emerged if stronger subsequent memory effects had been observed at these sites. Indeed, taken together, Mangels and colleagues (2017) and the present study converge in demonstrating that MG may lead to greater engagement over more anterior regions, whereas PGs may lead to greater engagement over posterior regions. The specificity of engagement within each region may depend on how goals are defined along with task demands. Thus, differences between the two studies may have contributed to differences in both behavioral and ERP effects.

**Accuracy Feedback: FRN, P3a, P3b, LPP**

When it comes to the accuracy feedback, it was reasoned that being informed an error was made would be more threatening under PGs compared to MGs (see Ames & Archer, 1988) and that this may be evident by attentional enhancements under negative feedback under PGs compared to MGs. However, there were no goal-related differences in the feedback-related ERPs measured in the present study, regardless of whether we assessed feedback processing in relation to earlier, more bottom-up aspects of attention to feedback (i.e. FRN and P3a) or later, more top-down aspects of attentional processing (i.e. P3b and LPP). Although some studies have found that personal achievement goals can influence the amplitude of the feedback-relevant ERPs (REFs), fewer studies have found positive support for the influence of induced goals. One of those studies used a probabilistic learning task, and found differences in the FRN between positive and negative feedback when participants were instructed to outperform others, but not when the task was framed as practice Van Meel and Van Heijningen (2010). However, in
contrast to the present study, negative feedback was rare in their task and the increase in salience may have modulated both its behavioral relevance and the sensitivity of the FRN to motivational effects (see Sambrook & Goslin, 2015).

One additional consideration for why our ERPs to accuracy feedback were not influenced by achievement goal condition may be that negative feedback is more threatening when ability goals, as opposed to purely normative goals, are activated. Indeed, Ames and Archer (1988) generally defined PGs structure as a classroom environment where students are focused on demonstrating their abilities and research on personal goals shows that when PGs are defined in terms of demonstrating ability they are linked to negative outcomes but when defined in normative terms they are linked to positive outcomes (Hulleman et al., 2010). Similarly, Mangels and colleagues (2006) showed greater P3a enhancement to negative feedback in relation to greater endorsement of PGs that were defined in both ability and normative terms. Interestingly, in that study, the relationship between P3a amplitude and PGs was strongest for high confidence errors, which are not only fairly rare, but represent both an error in accuracy and in the ability to gauge one’s own knowledge (i.e. ability) accurately. We necessarily sacrificed the power to subdivide trials along levels of confidence in favor of having the power to examine the manipulation of goals across both immediate and delayed tests. However, future studies could test the relationship of induced PGs on attention to accuracy feedback as a function of its relative frequency and salience to ability concerns.

Conclusion
The present study examined differences between MG and PG inductions on error correction as a function of retest delay and the neural mechanisms that underlie processes during learning following errors. The findings that an MG compared to a PG induction led to greater superior frontal positivity during the learning feedback following errors along with benefits in errors correction across retest delay supports the assertion that MGs unlike PGs facilitate adaptive approaches to learning in response to failure. Evidence of greater orienting to the learning feedback followed by sustained elaborative processing of that feedback confirms that MGs compared to PGs may potentially enhance learning through various cognitive processes. Thus, although instructors are already encouraged to provide students with opportunities to make mistakes and learn from them, they may also want to consider instilling an MG compared to a PG environment to help their students learn from errors in a more adaptive and successful manner.
Introduction

When trying to learn new semantic information, which is a fundamental aspect of school-based education, both instructors and students may undervalue instances when students provide the wrong answers to instructor-posed questions and place greater value on responses that are correct. Consequently, instructors may try to avoid situations where students are likely to make mistakes. However, research has consistently shown that learning is best when students first actively generate answers to questions, even if they are incorrect, and only after this generation exercise, are presented with the correct answer (for a review see Metcalfe, 2017). Compared to only having students review or study that information, allowing students to make errors ultimately seems to strengthen the association between the correct answer and question. Although various theories have begun to emerge regarding this associative process (Metcalf, 2017), less is known about the underlying neural substrates. To this aim, the present study used High-Definition transcranial Direct Current Stimulation (HD-tDCS) to investigate the causal role of two candidate neural regions for successfully encoding new general knowledge associations following initial retrieval errors: the lateral temporal cortex (LTC) and dorsolateral prefrontal cortex (DLPFC). Given our particular interest in the DLPFC, we also manipulated the achievement goals for the participants and whether instructions emphasized developing knowledge (mastery goals) or simply focused participants on answering questions correctly and doing better than others (performance goals).
Past research has implicated both the LTC and DLPFC in the process of updating of question-answer associations, at least when the correct answers already have a pre-existing representation in semantic memory. However, these regions may serve different roles. Regarding the LTC, in some of the earliest research into the neural correlates of general knowledge updating, which used event-related potentials (ERPs), Butterfield and Mangels (2003) showed that greater early negativity at electrodes over the left temporal cortex predicted learning of correct familiar, but not unfamiliar answers to general knowledge questions (familiarity was self-reported by the participant for each answer), regardless of whether memory was tested immediately or a week later. The authors reasoned this left temporal negativity may index activation of pre-existing semantic representations elicited upon presentation of the correct answer. Although ERPs have low spatial resolution so the exact source of the activity was unclear, a lateral temporal source would be consistent with fMRI findings localizing semantic processing of verbal information within the LTC (Gold et al., 2005; Gold & Buckner, 2002).

The view that areas of the LTC may be primarily involved in bottom-up processes involved in the activation of representations of the correct answer is further supported by the role of posterior portions of the LTC in basic word form processing (McCandliss et al., 2003). However, top-down attention to words can further enhance activity in these regions, and increases in LTC activity have been linked to benefits in memory when participants were explicitly instructed to memorize words (Ezzyat et al., 2018; Wittig et al., 2018). Thus, LTC may support early semantic processing of the corrective feedback following errors that is necessary, but not sufficiently for learning of the question-answer association. Updating semantic knowledge following errors additionally requires the association of the correct answer with a question to which it may not have a strong pre-existing semantic association, and that may not
even be concurrently presented. In other words, successful encoding is likely to also benefit from associative processing and the engagement of top-down strategic processes that serve to strengthen the integration of the correct answer and question (Cyr & Anderson, 2015; Huelser & Metcalfe, 2012; Knight et al., 2012).

The dorsolateral prefrontal cortex (DLPFC) is a candidate region for this top-down strategic association process, given its general implication in associative memory (for a review see Blumenfeld & Ranganath, 2007). Studies examining both activation of and brain stimulation over the DLPFC have shown that this region supports associative memory, particularly when associative demands are increased by the lack of pre-existing semantic relationships (i.e., unrelated word pairs: Blumenfeld, Parks, Yonelinas, & Ranganath, 2011; Sandrini, Cappa, Rossi, Rossini, & Miniussi, 2003; but see Hawco, Berlim, & Lepage, 2013); face and name associations: Leach, McCurdy, Trumbo, Matzen, & Leshikar, 2018). Similarly, brain stimulation (i.e. TMS, tDCS) studies have not found DLPFC activity to be linked to memory for items that already have strong semantic relationships (Lara, Knechtges, Paulus, & Antal, 2017; Sandrini et al., 2003), further supporting the view that DLPFC may be particularly engaged by the effortful strengthening of associations.

The DLPFC appears to support associative encoding even when participants are not explicitly instructed to memorize items, as long as task instructions emphasize associations, such as imagining items interacting together (Blumenfeld et al., 2011) or making judgments about whether items relate to each other (Leach et al., 2018; Sandrini et al., 2003). However, regions of the frontal cortex also have been implicated in the use of more strategic memory processes engaged in intentional learning tasks (Gershberg & Shimamura, 1995; Kirchhoff, 2009; Mangels, 1997). In general, through top-down processes, the DLPFC may facilitate the co-activation of
multiple representations in posterior cortex and as such, enhance the likelihood that inputs to the hippocampus will integrate into associative memories (Summerfield et al., 2006). Thus far, however, only a few studies have discussed the role of the DLPFC in associative memory pertaining to the correction of errors (X. L. Liu, Liang, Li, & Reder, 2014). Nonetheless, taken together, evidence is emerging that suggests the LTC as a candidate region for bottom-up processing of the correct answer when it is presented, whereas the DLPFC is a candidate region for the successful top-down integration of this answer with the previously presented question.

Although both the LTC and DLPFC may contribute to the successful encoding of corrective feedback, given the proposed sensitivity of the DLPFC to top-down goals, it may be possible to increase dependence on the DLPFC by manipulating the achievement goals emphasized by the task instruction. Specifically, we either directed students toward learning and mastery of new material for the sake of their own knowledge (mastery achievement goal) or on providing correct answers and how that compares to others knowledge (performance achievement goal) (see also Chapter 2). Manipulation of task goals in this manner draws upon both educational and social cognitive research (see Ames & Archer, 1988; Anderman & Patrick, 2012; Meece, Anderman, & Anderman, 2006), where the effects of promoting a mastery goal (MG) is typically found to enhance more semantic and elaborative processes compared to promoting a performance goal (PG) (Ikeda et al., 2015; Mangels et al., 2017; Murayama & Elliot, 2011a). Thus, compared to a PG emphasis, an MG emphasis may lead to greater use of elaborative associative memory processes and consequently, stimulation of DLPFC may lead to greater benefits in learning following errors under an MG induction.

In support of this hypothesis, previous ERP studies using the same task as the present study showed that, compared to a PG induction, an MG induction led to both greater error
correction and/or enhanced ERPs over frontal electrodes during presentation of the correct answer (Mangels et al., 2017; see also Chapter 2). In the first study of this dissertation, benefits under an MG induction persisted at a retest given one week later and were coupled with greater sustained frontal positivity at superior sites proximal to DLPFC but demonstrated no differences from PGs over temporal sites proximal to LTC. The superior frontal activity that was enhanced for MGs has been interpreted in past studies of memory encoding to index elaborative associative processing (Fabiani et al., 1990; Y. Liu et al., 2017; Mangels et al., 2001). Indirect support for the greater engagement of DLPFC by mastery-focused task goals also comes from a recent fMRI study finding that participants whose personal goals were more MG- than PG-focused exhibited greater DLPFC activity during feedback in a challenging rule-finding task (Lee & Kim, 2014). Thus, to the extent that an MG instruction similarly benefits error correction here, we expected that HD-tDCS stimulation over DLPFC would also disproportionately benefit students in the MG condition in contrast to LTC stimulation, which would equally benefit participants regardless of condition, compared to sham. Thus, the findings of this study will inform our understanding of the DLPFC and LTC to feedback-based learning of semantic knowledge following errors, as well as how this learning is modulated by MG and PG inductions.

Study Overview

In the present experiment, we used HD-tDCS to examine the causal effects of the left LTC and DLPFC on feedback-based learning after errors as a function of MG and PG inductions. We stimulated over left hemisphere sites only, given that memory benefits for verbal
material are most consistently found over the this hemisphere (Blumenfeld et al., 2011; Butterfield & Mangels, 2003; Ezzyat et al., 2018; Javadi & Walsh, 2012; Sandrini et al., 2003; Wittig et al., 2018). Stimulation of cortical sites was compared to a sham condition at the same sites as the active DLPFC stimulation.

We employed a general knowledge task where students first generated responses to challenging questions (e.g., What is the capital of Canada?), then were provided first with feedback about the accuracy of their response (correct, incorrect), then finally, with the correct answer (e.g., Ottawa) (e.g. Butterfield & Mangels, 2003). However, to increase the likelihood that presentation of the correct answer would activate pre-existing semantic representations in LTC regions, we also only used questions for which correct answers had been previously rated as familiar by at least 95% of a comparable student population (see Whiteman & Mangels, 2016). Finally, to increase the likelihood that DLPFC regions will also be engaged, we presented the correct answer after the question (4s) rather than concurrently, such that successful question-answer association may place greater demands on working memory processes. We note that it is not uncommon in classroom discussions for questions and answers to be presented sequentially rather than concurrently. Learning was examined by calculating the proportion of initial errors that were corrected on a surprise retest given after a one-week delay. We chose a week later retest and not an earlier time because we were interested in processes that supported learning that was retained over a longer period, which is also similar to the general goals of educators.

HD-tDCS is a type of brain stimulation during which a low-level current is applied to a targeted neural region (Miniussi, Harris, & Ruzzoli, 2013; Nitsche et al., 2008; Yavari et al., 2018). HD-tDCS is not thought to cause neurons to fire, but the addition of low-level current is thought to facilitate firing only for neurons that are near threshold. Stimulation is expected to
result in both “on-line” benefits as a result of actual stimulation and “off-line” benefits that persist even after stimulation is over (see Nitsche & Paulus, 2001). On-line effects may reflect changes in the membrane potential as evident by a lack of stimulation effects when sodium and calcium channels are blocked, whereas offline effects may reflect changes in plasticity as evident by lack of stimulation effects when NMDA receptors are blocked (for a review see Yavari et al., 2018). To examine both online and offline effects, the task was divided into 4 blocks. In the first two blocks, there was no stimulation, and thus, these blocks served to establish a baseline. Participants received HD-tDCS over the DLPFC, LTC, or sham HD-tDCS during the 3rd block of questions, which was off for the 4th block. The 4th block allowed us to examine “offline” stimulation effects.

It is important to mention that although HD-tDCS was applied during the 3rd block while participants were answering questions to examine differences in encoding of the correct answers, retrieval processes in relation to generating correct answers to these questions may also have been influenced by HD-tDCS over DLPFC and/or LTC. To ensure that the task was similarly challenging for all participants under all conditions, we used adaptive testing to bring all participants to a similar level of 35% correct. Thus, differences in correct responses between conditions were not expected.

Overall, both DLPFC and LTC stimulation were expected to facilitate error correction compared to sham. This would be evident in stimulation benefits compared to sham in either Block 3, during stimulation, or Block 4, in the post-stimulation period. The extent to which benefits to error correction carried over into block 4 (i.e. “off-line” benefits), would be particularly instructive about relationship between long-term plasticity in these regions and error correction on the delayed retest. Performance during the first two blocks would not necessarily
be expected to differ as a function of stimulation condition, and rather, serve as a baseline against which to evaluate the effects of stimulation.

We also predicted that the stimulation of DLPFC might infer greater benefits to error correction to participants who were engaged in the task under an induction that emphasized learning over performance compared to stimulation of the LTC and sham HD-tDCS. To this aim, at the outset of the task, and then before each block, participants received specific instructions designed to emphasize either an MG or PG (see also Chapter 2). Under PG participants were instructed to do better than others, whereas under MG instructions, participants were instructed to develop knowledge, but not to focus on doing better than others. The explicit direction to not focus on how others were doing in the MG condition was based on a previous study, which showed that even when MG goals are emphasized, participants may still engage in social comparison unless it is explicitly stated that this information is unimportant (Van Yperen & Leander, 2014). In both MG and PG instructions we also let participants know that at the end of the study they would be provided with information in relation to knowledge development or how they performance relative to others, respectively.

Benefits of an MG induction on error correction were expected to be observable as early as the first two blocks (i.e., pre-stimulation). However, to the extent that our MG instructions were successful in enhancing error correction compared to PG instruction in the pre-stimulation period, the application of DLPFC stimulation during Block 3 was predicted to amplify this effect further, given the proposed involvement of this area in the associative processes emphasized by the MG instruction. Stimulation of LTC was also expected to enhance error correction, equally for participants in MG and PG conditions, given that both instructional groups would be similarly engaged in bottom-up processing of the correct answer. The effects of goal inductions
seem to appear when the task is relatively difficult as opposed to easy (Avery & Smillie, 2013; Crouzevialle et al., 2015; Murayama & Elliot, 2011a) suggesting that the effects of stimulation as a function of goals may most influential later on in the task, such as during this third block, but may also extend to the 4th “off-line” block.

Method

Participants. Participants were eligible if they reported to be right-handed, had normal or corrected-to-normal vision and hearing, had no psychological or neurological disorders, learned English by age 5, had completed most of their education in the United States, and had no wounds or skin conditions on their scalp. 135 Brooklyn College students who met these eligibility requirements consented to participate in a manner approved by the Human Research Protection Program of the City University of New York (CUNY). They were compensated with research credit or $15 for each hour of participation.

A total of twenty participants were excluded from the analyses resulting in a final sample of 115 (ages 18-32; \( M = 20.47, \ SEM = .27 \); 78 Female/37 Male; see Table 2). Specifically, eight participants withdrew from the study due to: being unable to tolerate a HD-tDCS “pre-stim tickle” prior to the experiment (n=1), being unable to tolerate HD-tDCS during the experiment (n=1), time constraints (n=1), and failing to return for the retest (n=5). Additionally, twelve participants were withdrawn by the experimenter due to: issues with electrode impedance/contact quality from bridging or hair styles (n=4), the computer crashing during the stimulation period (n=1), using their cell phone during the experiment (n=1), rehearsing or discussing more than 15
items with others between session 1 and 2 (n=3; see Data Processing section), and due to poor performance on our task prior to administration of stimulation (n=3; see Data Analyses section).

Table 2

Pre-task measures for the final study sample. Participants reported their gender and age and prior to the start of the task completed the AGQ-R (Elliot & Murayama, 2008). Responses from the AGQ-R were averaged for the MAP (Mastery Approach) and PAP (Performance Approach) items that were rated on a scale of 1 (strongly disagree) to 5 (strongly agree).

<table>
<thead>
<tr>
<th></th>
<th>Gender</th>
<th>Age</th>
<th>MAP</th>
<th>PAP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean (SEM)</td>
<td>Mean (SEM)</td>
<td>Mean (SEM)</td>
</tr>
<tr>
<td>MG</td>
<td>DLPFC</td>
<td>18</td>
<td>6 M/ 12 F</td>
<td>19.17 (.41)</td>
</tr>
<tr>
<td></td>
<td>LTC</td>
<td>20</td>
<td>6 M/ 14 F</td>
<td>20.40 (.52)</td>
</tr>
<tr>
<td></td>
<td>Sham</td>
<td>19</td>
<td>6 M/ 13 F</td>
<td>20.63 (.68)</td>
</tr>
<tr>
<td>PG</td>
<td>DLPFC</td>
<td>20</td>
<td>7 M/ 13 F</td>
<td>21.20 (.88)</td>
</tr>
<tr>
<td></td>
<td>LTC</td>
<td>19</td>
<td>6 M/ 13 F</td>
<td>20.95 (.69)</td>
</tr>
<tr>
<td></td>
<td>Sham</td>
<td>19</td>
<td>6 M/ 13 F</td>
<td>20.37 (.66)</td>
</tr>
</tbody>
</table>

Participants were well-matched in terms of age, gender, and incoming personal mastery and achievement goals. As shown in Table 2, there were no differences between conditions in age (3 [Stimulation: DLPFC, LTC, Sham] x 2 [Induced Goal: MG or PG] ANOVA conducted on
age: $ps > .15$) or gender. We also measured incoming personal mastery (i.e. MAP) and performance (i.e. PAP) goals prior to inducing goals (see Materials section) to ensure participants in all conditions similarly adopted both types of personal goals. A 2 (Induced Goal: MG, PG) x 3 (HD-tDCS: DLPFC, LTC, Sham) x 2 (Goal Type: MAP, PAP) ANOVA on questionnaire means showed that personal mastery (i.e. MAP) and performance (i.e. PAP) means did not differ between goal and stimulation conditions ($ps > .30$).

**HD-tDCS.** Active and sham stimulation was administered using a 4x1 HD-tDCS adaptor connected to a 1x1 tDCS Low-Intensity Stimulator (Soterix Medical, USA) using 5 sintered Ag/AgCl electrodes. One central electrode was used as the stimulating electrode (i.e., the anode) and the other 4 electrodes were the returns (i.e., cathodes). The electrodes were fixed on an EEG cap with HD electrode holders and filled with SignaGel to ensure electroconductivity. Active stimulation consisted of a 2 mA current for 20-minutes. For each electrode, lead quality values above 0.10 were deemed acceptable ($M = 0.84, SEM = .04$), whereas values that were equal or below 0.10 were indicative of bridging and those participants were withdrawn from the study.

In the DLPFC group, the stimulating electrode was placed at F3, and the return electrodes at AF3, F5, F1, and FC3. In the LTC group, the stimulation electrode was placed at Tp7, and the return electrodes at P5, P7, P9 and CP5 (see Figure 1 for DLPFC and LTC montage and modelled current maps). The Sham group used the same electrode placements as the DLPFC group; during sham stimulation there was a 30 second ramp-up to 2 mA, then a ramp-down to .01 mA for the entire 20-minute period followed by a ramp-up to 2mA and a ramp down. This was a single-blind study.
Fig 11. Montage (left) and current mapping (right) of DLPFC and LTC stimulation. Sham stimulation entails the same placement as DLPFC. DLPFC/Sham anode at F3 (red electrode); cathodes at AF3, F5, F1, and FC3 (blue electrodes). LTC anode at TP7 (red electrode); cathodes at P5, P7, P9 and CP5 (blue electrodes).

After the end of session 1, participants were given a post-stimulation survey, in which they were asked to report the extent to which they experienced side effects of HD-tDCS (e.g., tingling, burning, etc.) (Villamar et al., 2013) and whether they thought they received active or sham stimulation. One participant could not decide whether they received sham or active HD-tDCS and was omitted from analyses related to subjective beliefs about stimulation. Side effects data are included in Appendix B.

Materials. After consenting to participate in the experiment, participants were asked to complete a 12-item achievement goal questionnaire (Achievement Goal Questionnaire-Revised (AGQ-R); Elliot & Murayama, 2008) measuring their endorsement of personal achievement goals. The questionnaire taps into 4 types of goals (i.e., mastery approach [MAP], mastery avoidance [MAV], performance approach [PAP], performance avoidance [PAV]) with 3 items per goal type that are rated on a scale of 1 (strongly disagree) to 5 (strongly agree). Minor modifications were
made to the original questions. The first item asked students to rate their goal endorsement “in this class”, which was updated to “in my courses” to query goal endorsement in a general manner. Similarly, 2 other items were updated from “this/the course” to “my courses.”

Both approach goals focus on goal attainment (e.g., MAP: My goal is to learn as much as possible; PAP: I am striving to do well compared to other students), whereas avoidance goals focus on avoiding failure to achieve the goal (e.g., MAV: My aim is to avoid learning less than I possibly could; PAV: My goal is to avoid performing poorly compared to others). For the purposes of the present experiment, we will focus only on approach goals given that they map more directly onto the MG and PG induction. Furthermore, many participants inquired about the meaning of a MAV item (i.e., My goal is to avoid learning less than it is possible to learn), and students often rephrased this item in terms of approach (i.e., MAP) when inquiring about its meaning, suggesting that the question was poorly understood, and participants may not have answered these items accurately. Scores for the personal achievement goals were computed by averaging ratings for the 3 items under MAP and PAP.

The “general knowledge” task was composed of 160 trivia questions (www.mangelslab.org/bknorms) in relation to history, geography, math, sciences, arts and humanities (Butterfield & Mangels, 2003; Mangels et al., 2006, 2017), and were presented across 4 blocks with 40 items per block. An example of a question that a participant may have been presented with is “What are the only birds able to fly backwards?” with “hummingbird” as a correct response. The correct answer for each question consisted of a single word and each answer was rated as being a familiar word to at least 95% by Baruch College undergraduates in a previously normed sample. During this norming study participants were not asked to rate familiarity of question and answer associations.
The 160 questions were selected for each individual participant from a pool of 416 questions that ranged in difficulty. Difficulty scores were calculated for each question based on the percentage of previous participants’ who correctly answered each question from both published (Mangels et al., 2018; Whiteman & Mangels, 2016) and unpublished studies conducted at Baruch College, with an average difficulty of 33% and a range of 0-95%. Note that difficulty is expressed in the percentage of previous participants who correctly answered the questions, which means that a higher percentage indicates an easier question. The range of difficulty allowed us to titrate question difficulty for each individual participant [see First-Test Titration (Session 1) below], which was important for 1) ensuring participants committed sufficient errors to allow for analysis of error correction, and 2) would experience failure. The experience of failure was further emphasized by framing these questions to participants as “general knowledge” suggesting that these items are typically known.

Procedure

After completing the AGQ-R survey, participants were set up for HD-tDCS and given a 30 second dose of 1 mA stimulation (i.e., a “pre-stim tickle”) to decide whether the sensations were tolerable or if they wanted to end their participation. Participants were then seated in front of a computer monitor and listened to recorded instructions that were read aloud to them through headphones. Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA) was used to present the task instructions and the general knowledge test.

Participants then completed the “First-Test” of general knowledge. First, to induce a specific goal, participants were presented with either MG or PG instructions before they began to
answer any general knowledge questions. Both MG and PG inductions were framed in an approach manner (i.e. toward success) as opposed to in an avoidance manner (i.e. toward avoiding failure) since approach goals are typically associated with better task performance (Van Yperen et al., 2015) and better achievement-related outcomes in the classroom, such as higher grades and greater interest in course material (Hulleman et al., 2010). The first paragraph of each induction focused students toward a goal while the second paragraph presented evaluative information in relation to that goal. Participants in the MG condition were instructed with the following:

*You are here to explore a set of general knowledge questions in order to develop your knowledge about various topics. You will be given feedback about the accuracy of your answers. Sometimes people think that we are comparing their performance to that of other students who have completed the task. However, your focus should NOT be to do better than other students, but rather to develop your own knowledge.*

*As the end of session 2, you will be provided with information you may find useful for understanding how your knowledge has developed over the course of this task.*

Participants in the PG condition were instructed with the following:

*You are here to answer a set of general knowledge questions about various topics. You will be given feedback about the accuracy of your answers in order to give you a sense of how well you
are performing. Your performance on this task will be compared to other Brooklyn College students and you should do your best to perform better than others.

At the end of session 2, you will be provided with information about how well you did compared to other Brooklyn College students.

In addition to these instructions, participants were also presented with shorter MG or PG reminder instructions at the onset of blocks 2 through 4. In the MG condition, participants were instructed, “you are here to explore the general knowledge questions and to develop your own knowledge”. In the PG condition, participants were instructed, “you are here to do your best to perform better than other students at Brooklyn College.”

A sample trial sequence is shown in Figure 12. For each general knowledge question, participants were instructed to type in their answer or their best guess if they did not know the answer within 3 minutes and to provide a confidence rating for their response on a scale of 1 (sure wrong) to 7 (sure right). This was followed by a 1.5 second fixation cross and then “accuracy feedback” that lasted for 1 second, which consisted of a green asterisk paired with a high pitch tone for correct responses or a red asterisk paired with a low pitch tone for errors. The program gave participants feedback indicating their response was correct (i.e., a green asterisk) if there was at least a 70% letter match between their response and the correct answer regardless of letter order, or the feedback indicated their response was incorrect (i.e., a red asterisk) if this letter match was below 70%. This type of coding allowed some spelling errors to be accounted for; however, it also resulted in some cases when participants were provided with inaccurate feedback. Manual corrections for feedback errors were done at the data analysis stage (see Data
Preprocessing). Accuracy feedback was followed by another 1.5 second fixation cross, and “corrective feedback” (i.e., the correct response) that lasted for 2 seconds. Note that corrective feedback was presented to participants regardless of their accuracy. There was an inter-trial interval (ITI) of 500 ms between questions.

Fig 12. Trial sequence of an incorrect response during the first-test. Participants have a 3-minute time limit to answer a question. After a response, participants are asked to make a confidence rating with a response of 1 signifying the participant is sure the answer is wrong, a 4 signifying the participant is unsure whether their answer is right or wrong, and a 7 signifying the participant is sure the answer is right. If the answer is marked incorrect, then the participant receives a red asterisk paired with a low pitch sound. If the answer is marked correct, a green asterisk is presented with a high pitch sound. If a participant does not respond within the 3-minute limit then they are pushed to the first fixation without being asked to make a confidence rating and are
marked incorrect. Learning feedback (i.e. correct answer) is presented regardless of response accuracy.

The 20-minute period of HD-tDCS occurred during the 3rd block only, and participants were told, during the consent process, that they would receive either sham or active stimulation. Block 3 was selected for stimulation to: 1) allow additional instances for participants to be presented with their goal induction, 2) become aware of difficult nature of the task in the blocks 1 and 2 and experience failure, and 3) to examine post-stimulation effects in block 4 (see Figure 13).

Fig 13. First-Test flow. Participants are either presented with MG or PG instructions in Block 1 (i.e. pre-stimulation block) before answering 40 general knowledge questions. Then they are presented with goal reminder instructions in Block 2 (i.e. pre-stimulation block) before answering the next 40 questions. In Block 3 (i.e. stimulation block), participants receive a 20-minute DLPFC, LTC, or Sham HD-tDCS while they are again presented with the same goal reminder instructions and another 40 questions. In Block 4 (post-stimulation block), participants do not receive any stimulation, but again receive the same goal reminder instructions and answer the final 40 questions.
Given that the task was largely self-paced, some participants completed the 3rd block of questions before the 20-minute sham or active stimulation session was over, whereas other participants took longer and might have even finished answering questions after the end of stimulation (range 12 - 41 minutes; $M = 20.28$, $SEM = .49$). Nonetheless, the average time to complete questions in Block 3 and the number of questions (range 15 – 40 questions; $M = 36.90$, $SEM = .50$) answered during active or sham stimulation did not differ between HD-tDCS and goal conditions (3 [Stimulation: DLPFC, LTC, Sham] x 2 [Induced Goal: MG or PG] ANOVA conducted on Block 3 Time: $ps > .70$; 3 [Stimulation: DLPFC, LTC, Sham] x 2 [Induced Goal: MG or PG] ANOVA conducted on number of questions answered during 20-minute active/sham stimulation: $ps > .79$).

Given that not all participants answered all 40 questions in the 3rd block during the 20-minute stimulation period, however, we examined effects of goals and stimulation using 15 trials per block since that was the minimum number of trials that each participant answered under sham and active stimulation. This allowed us to examine “on-line” effects during block 3 without including later items during that block that were essentially “off-line” for those who took longer than 20 minutes to answer all 40 questions in that block. Data from all 40 questions are included in Appendix B.

We examined participants’ subjective experiences of goal inductions among three types of ratings: goal adoption, importance of goal induction, and memory for goal induction instructions. After each block of questions, for a manipulation check, participants first answered questions about how well they adopted the induced goals on a scale of 1 (not at all) to 9 (extremely). For Mastery Goal adoption, participants were asked to rate “how hard did you try to develop your knowledge” and for Performance Goal adoption, participants were asked to rate
“how hard did you try to do better than other students” (see Murayama & Elliot, 2011). An ANOVA indicated that goal adoption did not interact with Block (2 [Goal Adoption: Mastery or Performance] x 4 [Block: 1-4] x 3 [Stimulation: DLPFC, LTC, Sham] x 2 [Induced Goal: MG or PG] ANOVA conducted on ratings: ps > .05). Only a block by stimulation interaction approached significance, $F(5.176, 321.268) = 2.171, p < 0.06, \varepsilon = 0.86$. We ran 18 pairwise post-hoc comparisons between each of the blocks in each stimulation condition to follow up on this interaction, and results were considered significant after Bonferroni correction ($p < .003$). However, none of the comparisons condition survived Bonferroni correction ($ps > .003$). Given that there were no block effects, our primary analyses of ratings for Mastery and Performance Goal adoption represent the average across the 4 blocks.

Participants also completed a post-task survey after completion of the goal adoption questions in block 4. They were asked to rate the extent that the goals were important to them on a scale of 1 (not at all) to 6 (very much) (i.e., PG: It was important for me to do better than other students; MG: It was important for me to develop my knowledge). Finally, they were asked to rate to what extent they agreed, on a scale of 1 (disagree) to 6 (agree), that their instructions were MG (i.e., develop your knowledge) and PG (i.e., perform better than other Brooklyn College students) focused. One participant did not answer goal importance questions and another participant did not answer the instruction questions. These two participants were omitted from the analyses in which they were missing data.

One week-later participants returned to the laboratory and were retested on the same general knowledge questions under a MG or PG induction, which we refer to as the “retest”. Participants in the MG instruction were instructed to “please keep in mind that you are here to explore the general knowledge questions and to develop your knowledge.” For the PG
instruction, participants were instructed to “please keep in mind that you are here to do your best to perform better than other students at Brooklyn College.” The same evaluative information was included at the retest as in the general knowledge test [see the second paragraph of the MG and PG inductions above (i.e. “At the end of session 2, . . .”)].

Retest questions were randomized within each block and participants were presented with all of the questions from the earlier block before they were presented with questions from the later blocks (e.g., questions from Block 1 preceded Block 2). Accuracy and corrective feedback were presented at the same time with the correct answer in green if the response was correct, or in red if the response was incorrect. One participant, who was included in the analyses, only was able to answer 31 out of the 40 questions in Block 4 of the first-test due to a computer crash and thus, was retested on 151 trials in total compared to 160 for the other participants. At the end of the retest, participants were asked to identify any questions they looked up, rehearsed, or discussed with others prior to coming to session 2.

First-Test Titration (Session 1). Because the main focus was error correction, performance was titrated during Session 1 of the first-test of general knowledge to ensure similar numbers of error trials across participants in each condition. In order to ensure that participants experienced challenge during the task, the titration target was ~35% accuracy in each block. Each question item was associated with a normed difficulty value (see Materials above) ranging from levels of 0 to 95% accuracy with 0% accuracy representing the most difficult questions and 95% accuracy representing the easiest questions. In each block, participants would start off with a question chosen at random in the 25% to 45% difficulty range. If overall block accuracy was
above 34% and below 36% (i.e., 35%), questions would again be randomly selected from the 25% to 45% difficulty range.

If overall block accuracy was less than 34% correct, then questions with an easier level of difficulty than the previous question would be selected, and this would continue until overall block accuracy was between 34% and 36%. For example, if a participant answered a question with a difficulty level of 60% correctly, but overall block accuracy was 30% then the next selected question would be randomly selected in the 61% to 95% difficulty range. If the easiest questions were exhausted, due to a participant answering multiple questions incorrectly, then the program would select the next available easy question, which may have been slightly more difficult (e.g., 59%) than the previous (e.g., 65%).

If the overall block accuracy was above 36% then a more difficult question would be selected in comparison to the difficulty level of the previous question until overall block accuracy was brought down to a minimum of 34%. For example, if a participant answered a question with a difficulty level of 60% correctly, but overall block accuracy was 40% then the next selected question would be randomly selected in the 59% to 0% difficulty range. If the most difficult questions were exhausted, due to a participant answering multiple questions correctly, then the program would select the next available difficult question, which may have been slightly easier (e.g., 15%) than the previous (e.g., 10%). Mean difficulty of all of the items presented in each block did not differ as a function of block order, stimulation condition and/or goal condition (4 [Block: 1-4] x 3 [Stimulation: DLPFC, LTC, Sham] x 2 [Induced Goal: MG or PG] ANOVA conducted on mean difficulty: \( ps > .20 \)).
Data Processing. Trials were removed from analyses if participants were provided with inaccurate feedback (i.e., feedback indicating an error was made when their response was correct, or feedback indicating a correct response when they made an error). Because “accuracy feedback” was based on a 70% letter match between a participant’s response and the correct answer regardless of letter order, occasionally the computer would give inaccurate feedback. Furthermore, this matching rule did not account for synonyms, other spelling errors, and two-word answers when the correct answer was always a one-word answer (e.g., Vatican City instead of Vatican when asked “What city is also the world’s smallest country?”) and would provide feedback indicating a response was incorrect. Thus, the data were checked manually for “accuracy feedback” errors after the experiment by at least two independent coders.

Discrepancies between raters were resolved through discussion, use of online tools, with the final decision being made by the first author (Y.O.). In some cases, an on-line thesaurus (www.thesaurus.com) was used to check for synonyms between a response and the correct answer. For example, for the question “What is the longest-living species of mammal?” a response of “people” was marked incorrect by the computer even though the official correct response in the database was “human.” Online spell checkers (i.e., www.jspell.com/html-spell-checker.html, www.gandjlawrence.co.uk/Werdz/ ) were used to check spelling if it was missed by the program. For example, for the question “What sailors’ disease resulted from a deficiency of vitamin C?” a response of “scurvies” was marked incorrect by the algorithm when the official correct response in the database was “scurvy.” At other times, participants were given correct feedback for unrelated words because the letter match reached above 70%. For example, for the question “Which one of the traditional 9 planets of our solar system has clouds of methane gas?”
a response of “Saturn” was marked correct by the algorithm when the correct answer was “Uranus.”

Trials were also removed if it appeared that participants may have known the correct answer, but were given the negative accuracy feedback. This included instances when participants provided a first name, but the correct answer was a last name, and instances when participants’ answers were ambiguous and it was unclear whether a participant did or did not know the correct answer at the first test (e.g., for the question “What stringed weapon fires a bolt?”, a response of “bow” was marked incorrect when the correct answer was “crossbow”). Additionally, some responses were removed because they failed computer spell-checking, but were similar to the correct answer. For example, for the question “What is the name of the cord that connect a mother and a fetus?”, a response of “abilical cord” was marked incorrect and was not flagged as a spelling error by computer spell checking when the correct answer was “umbilical.” The rationale for removing these trials was that the degree of learning for these “nearly-correct” responses would likely be significantly less than situation where someone had a clear miss from the correct answer.

The rules for removing trials for the retest differed from those for the general knowledge test because the retest was the last task and feedback inaccuracies were no longer deemed as problematic. If a question resulted in inaccurate feedback during the general knowledge test at session 1, then it was also removed from retest analyses at session 2 for that participant. If a question resulted in accurate feedback at the general knowledge test, but inaccurate feedback at the retest, then retest accuracy would simply be recoded based on the rules above. Specifically, a response would be recoded as correct if it was a synonym or passed an on-line spell checker, but would be recoded as an error if an unrelated word was generated. However, if first names instead
of last names were generated at the retest then these responses were left as errors since the last
name was clearly given as corrective feedback, and failure to use that information suggested it
was not learned successfully.

Finally, any items that participants answered incorrectly on the general knowledge test
that they reported to have looked up, rehearsed, or discussed with others prior to coming to
Session 2 were removed from the analyses. It was thought that this additional practice would
strengthen memory, making it difficult to attribute learning to Session 1 when more learning may
have occurred during this additional practice and less by the goal induction and/or stimulation
during the general knowledge test. The majority of participants (109 ; 91%) did not rehearse,
look up or discuss any information. However, nine participants reported rehearsing 4 items or
less and these trials were removed for these subjects. Additionally, three participants indicated
rehearsing more than 15 items (range 16-36), which would have resulted in 4 or more trials lost
for each block, and thus, we opted to exclude these participants from the analysis as a whole (see
also Participants section). Notably, for these three participants, these trials were mainly biased
toward initial errors, whereas for other participants items were typically lost due to both correct
and incorrect responses being marked inaccurately by the program at the general knowledge test.

The total loss of trials for the multiple reasons outlined above in the usable sample of 115
ranged from 0 to 11 \( (M = 2.15; \ SEM = .17) \), which did not differ or interact between conditions,
\( (3 \text{ (Stimulation: DLPFC, LTC, Sham)} \times 2 \text{ (Induced Goal: MG or PG)}) \) ANOVA conducted on
number of trials lost: \( ps > .44 \). As for the 15 trials per block analyses (see Results section), the
total loss of trials for the same usable sample ranged from 0 to 5 \( (M = .86; \ SEM = .10) \), which
did not differ or interact between conditions, \( (3 \text{ (Stimulation: DLPFC, LTC, Sham)} \times 2 \text{ (Induced
Goal: MG or PG)}) \) ANOVA conducted on number of trials lost: \( ps > .29 \).
Data Analyses

We first examined questions related to the success of inducing mastery and performance goals. Normality of each manipulation check item was examined using a Shapiro-Wilk test within each stimulation and goal condition, resulting in 6 Shapiro-Wilk tests in total for each item. A non-parametric test (i.e. related-samples Wilcoxon Signed Rank Tests) was employed if normality of any item was violated for at least one stimulation by goal condition. Greenhouse-Geisser corrections were used when sphericity, as determined by Mauchly’s Test, was violated.

Turning to the experimental task, accuracy on the first-test was calculated as a proportion of correct responses out of the total number of usable questions in each block. Participants who achieved a proportion correct score of less than .30 on either Blocks 1 or 2 were considered to have failed titration (i.e. initial poor performance). Although there were no participants who scored below .30 during Block 1, during Block 2 there were three participants whose proportion of corrects ranged from .23 to .26. After the removal of these three outliers, scores were within a range of .30 to .41 for Block 1 and within a range of .31 to .40 for Block 2. Titration issues were not examined during Blocks 3 and 4 since those blocks may also be influenced by stimulation and post-stimulation effects, respectively.

Even after removing outliers and achieving titration, accuracy around the titration target varied somewhat and may have been influenced by the goal induction and/or stimulation. Thus, a 4 (Block: 1-4) x 3 (Stimulation: DLPFC, LTC, Sham) x 2 (Induced Goal: MG or PG) ANOVA was conducted on proportion of correct responses at session 1 to examine differences in recall of general knowledge.
Turning to the retest, the dependent variable (DV) of interest was error correction, and scores were calculated as the proportion of errors corrected divided by the total number of errors made on the initial general knowledge test. In these analyses, because accuracy on the general knowledge test at Session 1 varied based on Block and Goal (see Results), the DV used in our analyses of error correction regressed out the proportion of correct responses on the general knowledge test. “Regressed error correction proportions” were recalculated for each participant in each block by taking the sum of the grand mean for the proportion of errors corrected in the block and the residual from a regression analysis that included the proportion of correct responses on the general knowledge test as the predictor and the proportion of errors corrected on the retest as the DV. To test for effects of block, goal, and stimulation on error correction, a 4 (Block: 1-4) x 3 (Stimulation: DLPFC, LTC, Sham) x 2 (Induced Goal: MG or PG) was conducted on regressed error correction proportions. Post-hoc tests were considered significant after Bonferroni adjustment for the relevant number of multiple comparisons.

Results

Manipulation Checks

Goal Instruction Memory. Our first manipulation check involved confirming that participants remembered their Goal Instruction (for the 114 participants that answered this question). Related-Samples Wilcoxon Signed Rank Tests were conducted between the Mastery and Performance Instruction ratings for each of the 6 Goal by Stimulation conditions (PG-DLPFC, PG-LTC, PG-Sham, MG-DLPFC, MG-LTC, MG-Sham) since this data failed
normality for all Shapiro-Wilk tests ($p < .05$). Participants in each condition remembered the instructions. As shown in Table 3, participants under MG had higher medians for the Mastery compared to the Performance Instruction item, whereas those under PG had higher medians for the Performance compared to the Mastery Instruction item ($p < .01$). This suggests that participants were aware of their goal manipulation instructions.

Table 3

Participant responses on the manipulation check items for each condition. Instruction and Importance were both measured after the first-test using single items rated on a scale of 1 (disagree/not at all, respectively) to 6 (agree/very much, respectively), with one item assessing Mastery and another item assessing a Performance focus. Adoption measured after each block during the first-test was calculated as an average across 4 items rated on a scale of 1 (not at all) to 9 (extremely) with 4 items for Mastery and another 4 items for Performance.

<table>
<thead>
<tr>
<th></th>
<th>Instruction</th>
<th>Importance</th>
<th>Adoption</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mastery</td>
<td>Performance</td>
<td>Mastery</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>Median</td>
<td>Median</td>
</tr>
<tr>
<td>MG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLPFC</td>
<td>5</td>
<td>1</td>
<td>4.5</td>
</tr>
<tr>
<td>LTC</td>
<td>6</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Sham</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>
Goal Importance. Our second manipulation check involved testing whether or not the MG group rated mastery goals as more important than performance goals, and the PG group rated performance goals as more important than mastery goals, and whether this differed by stimulation condition (for the 114 participants that answered these questions). Related-Samples Wilcoxon Signed Rank Tests were conducted between the Mastery and Performance Importance ratings for each of the 6 Goal by Stimulation conditions since data failed normality in 4 out of the 6 Shapiro-Wilk tests for the mastery goal item ($p < .03$) and in 3 out of the 6 tests for the performance goal item ($p < .03$).

The same pattern was found across LTC and sham stimulation conditions, but not for the DLPFC condition. For both LTC and sham, participants in both the MG and PG induction conditions had higher medians for the Mastery Importance Item (Medians: LTC = 4, Sham = 5) compared to the Performance Importance item ($p < .029$). In the DLPFC stimulation condition, however, participants under MG had significantly higher medians for the Mastery compared to the Performance Importance item ($p < .001$), whereas those under PG had similar medians for the Mastery and Performance Importance item ($p > .15$). This suggests that despite being aware of the goal manipulation instructions, participants generally valued mastery goals over performance goals in our task. Although the participants who experienced DLPFC stimulation
under PG indicated that both mastery and performance goals were valued, differences in goal importance ratings did not appear between stimulation conditions (see Appendix B).

**Goal Adoption.** Our third manipulation check involved testing whether or not the MG group adopted mastery goals more than performance goals, and the PG group adopted performance goals more than mastery goals, and whether this differed by stimulation condition. Related-Samples Wilcoxon Signed Rank Tests were conducted between the Mastery and Performance Adoption ratings, averaged across the 4 blocks, for each Goal and Stimulation condition because these data failed normality in 2 out of the 6 Shapiro-Wilk tests for the performance goal average ($p < .05$). Shapiro-Wilk tests for the mastery goal average did not fail normality ($p > .24$).

The same pattern was found across all stimulation conditions. Participants under MG had higher means for Mastery Adoption compared to the Performance Adoption ($p < .001$), whereas those under PG had more similar means for the Mastery and Performance Adoption questions ($p > .06$). The comparison under PG did approach significance for the LTC group ($p < .08$) but not for the other stimulation groups ($p < .34$) with higher means for Master Adoption compared to the Performance Adoption. This suggests that although under MG participants were focused more on mastery than performance goals, participants under PG were focused more similarly on both goals in our task.

**HD-tDCS & Subject Blinding.** As mentioned above, participants were asked to report whether they thought they received sham or active stimulation following completion of the general knowledge task. Table 4 shows the counts of participants in each of the conditions that
thought they received sham or active stimulation. To examine whether these beliefs of active and sham differed between stimulation conditions, we ran one logistic regression with DLPFC as a reference and another logistic regression with LTC as a reference. The main effect of stimulation conditions on stimulation perceptions approached significance (Wald Chi-Square = 5.179, \( p < .08 \)). Significantly more participants reported receiving active stimulation in the DLPFC compared to sham condition (Wald Chi-Square = 4.694, \( p < .031 \)) and the same pattern showed only a trend toward significance in the comparison between the LTC and sham condition (Wald Chi-Square = 2.967, \( p < .09 \)). However, there was no difference between DLPFC and LTC conditions (Wald Chi-Square = 0.243, \( p > .62 \)), suggesting that while blinding was not perfect between sham and the stimulation conditions, especially with the DLPFC condition, blinding was similar between the stimulation conditions.

Table 4
Number of participants who reported receiving active or sham stimulation in each condition.

<table>
<thead>
<tr>
<th></th>
<th>MG</th>
<th>PG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Active</td>
<td>Sham</td>
</tr>
<tr>
<td>DLPFC</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>LTC</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>Sham</td>
<td>8</td>
<td>11</td>
</tr>
</tbody>
</table>

*First-Test & Retest Performance*
“Online” trials/15 Trial analyses. As mentioned above, we examined the proportion of correct responses on the general knowledge test and error correction rate at retest as a function of Goal and Stimulation. Given that not everyone answered all 40 items during the 20-minute stimulation period, we focused analyses using the first 15 items, which was the minimum number of items that all participants answered during active or sham stimulation. Thus, we opted to focus our analyses of the effects of Stimulation and Goal during stimulation (i.e., “online” effects) on these “early items.” This cutoff was applied for each block (including baseline blocks and post-stimulation blocks) to account for any primacy effects that may have occurred within each block.

First-Test (Session 1): Figure 14 shows performance on the first-test as a function of block, stimulation and induced goal condition. A 4 (Block: 1-4) x 3 (Stimulation: DLPFC, LTC, Sham) x 2 (Induced Goal: MG or PG) ANOVA on the proportion of correct responses showed a main effect of Block, $F(2.73, 297.60) = 5.425, p < 0.01, \varepsilon = 0.91$. All other effects and interactions were not significant ($ps > .19$), but the main effect of Goal, $F(1, 109) = 3.546, p < 0.07$, and Goal by Block interaction, $F(2.73, 297.60) = 2.505, p < 0.07, \varepsilon = 0.91$, approached significance.
Fig 14. First-test accuracy across the first 15 items (early items) per block. Proportion of correct responses for early items in each block under Mastery (top) and Performance (bottom) Goals for each stimulation conditions. Participants were similarly titrated across goal and stimulation conditions, but accuracy was significantly worse in Block 4 compared to Block 1. Stimulation occurred during the 3rd block only. Error bars denote standard error of the mean.

Six pairwise post-hoc comparisons between each of the blocks were conducted to follow up on the main effect of Block, and results were considered significant after Bonferroni
correction (p < .008). The only comparison that was significant showed that participants answered more questions correctly in Block 1 ($M = .37; \ SEM = .01$) compared to Block 4 ($M = .34; \ SEM = .01$) (Bonferroni corrected $p < .002$). These results suggest that titration was less effective at maintaining performance at 35% correct in the final block, perhaps because most of the easier questions had been exhausted by then. Nonetheless, the lack of stimulation, induction condition or interactions indicates that participants were similarly titrated between goal and stimulation conditions across the early 15 trials.

Retest (Session 2): Figure 15 illustrates retest performance as a function of block, stimulation and induced goal. A 4 (Block: 1-4) x 3 (Stimulation: DLPFC, LTC, Sham) x 2 (Induced Goal: MG or PG) ANOVA on proportion of errors corrected with initial accuracy regressed out showed a main effect of block, $F(3, 327) = 2.718, p < 0.05$, a main effect of Stimulation, $F(2, 109) = 5.551, p < 0.01$, and Block by Stimulation interaction, $F(6, 327) = 3.064, p < 0.01$. All other effects and interactions were not significant ($ps > .18$).

Six pairwise post-hoc comparisons between each of the blocks were conducted to follow up the main effect of Block, and results were considered significant after Bonferroni correction ($p < .008$). The only comparison that was significant showed that error correction in Block 1 ($M = .41; \ SEM = .02$) was better than Block 4 ($M = .35; \ SEM = .02$) (Bonferroni corrected $p < .008$).
Fig 15. Retest accuracy across the first 15 items (early items) per block. Proportion of errors corrected with initial proportion correct regressed out for early items in each block under Mastery (top) and Performance (middle) Goals for each stimulation condition. The bottom graph represents the proportion of errors corrected with initial proportion correct regressed out for early items in each stimulation condition across goals and blocks. During Block 2, LTC led to significantly better error correction than DLPFC or sham. Across the 4 blocks, both DLPFC and LTC led to better error correction than sham. Error bars denote standard error of the mean.

As for the main effect of Stimulation, three pairwise post-hoc comparisons between stimulation conditions were conducted, and results were considered significant after Bonferroni correction (p < .016). Participants in the DLPFC condition (M = .40; SEM = .02) and LTC condition (M = .42; SEM = .02) corrected more errors than those in the Sham condition (M = .34; SEM = .02), (ps < .008).

For the Block by Stimulation interaction, twelve pairwise post-hoc comparisons between stimulations condition at each block were conducted, and results were considered significant after Bonferroni correction (p < .004). The only comparison that was significant showed that LTC stimulation (M = .49; SEM = .03) led to better error correction compared to Sham (M = .33; SEM = .03) and DLPFC (M = .35; SEM = .03) during block 2 only (Bonferroni corrected p < .002).

These results suggest that both stimulation conditions benefited error correction compared to sham overall, at least when considering the first 15 items of each block, and that LTC stimulation was especially beneficial to these items in the pre-stimulation block.
Discussion

The primary goal of the present study tested the roles of the DLPFC and LTC in updating of semantic knowledge following errors as a function of achievement goals. As predicted, HD-tDCS over both the DLPFC and LTC benefited subsequent memory on the week-delayed retest, compared to sham HD-tDCS, at least when focusing on the first 15 items of each block and when collapsing across all four blocks. Interestingly, however, memory benefits of stimulation did not appear to be specific to either the stimulation block (Block 3) or the post-stimulation block (Block 4), and rather, affected pre-stimulation blocks as well. Indeed, the effects of HD-tDCS over LTC was particularly beneficial for learning of Block 2 items, which preceded the stimulation period, compared to HD-tDCS over the DLPFC and sham.

Our secondary goal was to examine how stimulation effects might be modulated by achievement goals, with the main prediction being that DLPFC stimulation would be particularly beneficial under an MG induction, whereas LTC stimulation effects would be similar under PG and MG inductions. Although other studies have found that MGs lead to deeper elaborative processing of the learning information compared to PGs (Ikeda et al., 2015; Mangels et al., 2017; Murayama & Elliot, 2011a), there was no evidence of differences in error correction between goals in the present study, even during sham condition, suggesting that the manipulation did not influence encoding strategies overall. Not surprisingly, stimulation effects did not interact with the induced goals either.

In the following sections, we first consider how these results may inform our understanding of the role of the DLPFC and LTC in learning following errors and then why the achievement goals induction may not have been as effective in moderating these effects.
Effects of HD-tDCS over the DLPFC and LTC on Learning

Although HD-tDCS was applied only during Block 3, leading us to expect effects of HD-tDCS in Blocks 3 and 4, unexpectedly there were improvements in learning across all 4 blocks for DLPFC and LTC groups compared to sham HD-tDCS. One interpretation of the global benefits of stimulation on error correction across pre-stimulation, stimulation, and post-stimulation blocks is that both DLPFC and LTC stimulation similarly facilitated consolidation (for a review see Dudai, Karni, & Born, 2015) and/or reconsolidation (for a review see Besnard, Caboche, & Laroche, 2012) of question and answer associations activated during the first general knowledge test. Consolidation typically refers to the formation of novel memories, whereas reconsolidation refers to updating of previously established memories. With respect to reconsolidation, greater benefits in error correction are shown when there is already some degree of a pre-existing relationship between a question and answer compared to when a question and answer comprise a novel association (Butterfield & Mangels, 2003). However, it is not clear if participants were familiar with the question and answer relationships and if stimulation benefited learning through consolidation and/or reconsolidation processes. Consequently, stimulation over the DLPFC and LTC may have strengthened the association between a novel question and answer as well as the previously learned question and answer. Although there is a little research on the effects of tDCS on post-encoding processes, Rossi and colleagues (2011) showed that TMS over the left DLPFC disrupted memory for scenes only when applied 500 ms post-stimulus, but not when applied earlier from 100 to 400 ms post-stimulus, suggesting that the DLPFC may support post-encoding and/or consolidation processes. As for LTC, we are not
aware of any studies that examined the effects of LTC on post-perceptual processes or consolidation, and in our study HD-tDCS over LTC was especially beneficial to learning in the block that immediately preceded stimulation. It is not clear why benefits were greatest in this block, however, along with benefits in learning ERPs over LTC were also shown to be enhanced during negative feedback following high compared to low confidence endorsement (Butterfield & Mangels, 2003), and these types of memory mismatches more generally have been shown to benefit consolidation (Aberg, Müller, & Schwartz, 2017) and reconsolidation (Ecker, 2015). Thus, it may be that LTC stimulation may have been especially beneficial to consolidation and/or reconsolidation of question-answer associations following errors that were presented before the stimulation period putatively through mismatch-related bottom-up processes.

It is worth noting that brain stimulation-mediated consolidation effects are typically indexed by comparing memory performance before and after stimulation, where better later memory performance when compared to other stimulation conditions or baseline signifies greater effects of consolidation. The present study did not obtain memory measures prior to stimulation as a comparison to the week later retest and cannot speak to consolidation in terms of differences in memory performance before and after the stimulation. Further research is needed to examine these possible consolidation effects using memory measures before and after stimulation, as well as potentially obtaining more explicit measures of salience, arousal, or novelty of question-answer associations in order to determine if stimulation is particularly beneficial to these items, as might be predicted by a consolidation-based effect. Whereas bottom-up features such as stimulus novelty and semantic relatedness may mediate the post-encoding stimulation effects of LTC stimulation, it may be that top-down, goal-relevant variables such as stimulus interest or curiosity may mediate post-encoding effects of DLPFC stimulation.
We chose to stimulate over LTC and DLPFC regions because of their putative involvement in semantic and associative processes, respectively, both of which were hypothesized to be necessary for updating errors in semantic memory through sequentially presented feedback. The finding the stimulation of both regions facilitated error correction compared to sham supports the hypothesis that these areas contribute to this error correction process to a similar degree, and indeed may work together to support this associative process (i.e. Summerfield et al., 2006; see also Stagg et al., 2013). Nonetheless, a parsimonious explanation that aligns with past research would suggest that the LTC provides bottom-up semantic processing of the verbal stimuli (Gold et al., 2005; Gold & Buckner, 2002; McCandliss et al., 2003; Wittig et al., 2018), whereas the DLPFC primarily contributes top-down elaborative processes that also align with its more general role in working memory (Ranganath et al., 2005; for a review see Kane & Engle, 2002). Consequently, LTC-mediated bottom-up and DLPFC-mediated top-down processes may lead to similar behavioral outcomes when it comes to updating of factual information following errors for which there is some degree of a pre-existing association. Since LTC seems to support learning of weak pre-existing associations (Butterfield & Mangels, 2003) and DLPFC seems to support learning when elaborative associations are necessary (for a review see Blumenfeld & Ranganath, 2007), the roles between LTC and DLPFC in semantic learning may be more dissociable when it comes to learning novel information that may be more dependent on the DLPFC.

*Effects of MG vs. PG Inductions*
Based on previous EEG and fMRI work (Lee & Kim, 2014; Mangels et al., 2017; see also Chapter 2), we initially predicted that stimulation over the DLPFC would infer a greater benefit learning for those under MG compared to PG; however, the similar benefits of stimulation of both DLPFC and LTC to learning regardless of goals suggest that participants under both MG and PG may have engaged in similar cognitive processes. This is further supported by the finding that under sham conditions, MG and PG groups also performed similarly. In the sham condition, the pure effects of the induction could be observed in the absence of influence from the stimulation effects. We note that goal effects also failed to emerge even in the baseline Blocks 1 and 2, which was before any effects of expectations associated with sham stimulation might have influenced performance.

It is unclear why benefits of an MG induction failed to emerge. One possibility may lie in the nature of the instruction. A meta-analysis of personal achievement goals has noted that differences in wording can influence the size and direction of goal effects (Hulleman et al., 2010). For example, when PGs are defined in terms of demonstrating ability they are associated with negative outcomes, but when they are defined in normative terms as in the present study, they are associated with positive outcomes. However, the present study purposefully used the same instructions employed by the first study in this dissertation that showed a benefit in error correction under MG compared to PG. Thus, it would be difficult to attribute the lack of goal differences solely to instructions.

The effects may be influenced by other environmental factors, such as having the experimenter in the room in the present study. In the present study, it was necessary that the experimenter be in the same room during the entire session in order to administer HD-tDCS, unlike in the first EEG study of this dissertation where the experimenter could monitor the
recording outside of the room after the electrodes were applied. The presence of the experimenter combined with expectancy effects associated with the application of active stimulation (or sham) may have overshadowed more subtle effects of the goal manipulation.

Another consideration is the extent to which the groups adopted the goals they were instructed to, and the extent to which they felt these adopted goals were important to them. The lack of benefits of MG compared to PG on learning in the present study is unlikely to have been driven by reports of goal adoption. Similar to the first study in this dissertation, where an MG induction advantage for error correction was observed, participants reported greater adoption of mastery compared to performance goals under MG, but similar levels of mastery and performance goal adoption under PG. However, in the first study of this dissertation goal importance was not assessed, and for this measure both the LTC and sham groups indicated valuing mastery over performance, regardless of induction condition (i.e., even in the PG group when performance goals were emphasized). The PG induction was more effective at balancing the importance of PG and MG goals in the DLPFC group, however. For the DLPFC groups, the MG induction condition demonstrated the expected bias in importance toward MG goals, but under PG induction, both goals were valued similarly. Taken together, these findings suggest that when instructions emphasize PGs, students may still highly value learning, even if they attempt to adopt more of a performance goal than under MG instructions, and this may have neutralized differences between goal conditions.

Other goal manipulation studies have also failed to find quantitative differences in overall memory between goals (Crouzevialle et al., 2015; Mangels et al., 2017; Murayama & Elliot, 2011a), even though differences might emerge in the qualitative phenomenology of the memories. For example, Murayama and Elliot (2011) showed that memory for correctly
generated responses were more likely to be accompanied by contextual details (i.e. “remember” judgments) under PGs on an immediate test, but based more on familiarity (i.e. “know” judgments) under MGs. Additionally, it is interesting that in some studies, direct neural measures (e.g., fMRI, ERPs) of cognitive processes have also found significant neural differences between goals despite a lack of behavioral differences (Lee & Kim, 2014; Mangels et al., 2017). Taken together, these findings suggesting that behavioral measures of overall memory can be less sensitive to differences between goals than neural measures that might differentiate underlying processes that might nonetheless lead to similar behavioral outcomes, or behavioral measures that might differentiate memory based on more subtle aspects of memory phenomenology.

Finally, it is possible that there were neural differences between MG and PG inductions outside of the DLPFC and LTC areas that were stimulated. Indeed, Mangels and colleagues (2017) showed that despite a lack of behavioral differences, MG goals resulted in encoding-related activity over bilateral frontal regions that were more inferior than the sites stimulated in the present study, although that study did not analyze more superior frontal sites. Future research is needed to examine under which circumstances achievement goals result in differential influences on learning following errors.

Potential Limitations

We made the choice to examine only the first 15 trials from each block because the number of questions participants answered during the 20-minute sham or active stimulation session varied. We chose the minimum number of trials that every participant answered during that 20-minute period to examine more pure “on-line” and “off-line” effects of stimulation, and showed significant effects of DLPFC and LTC stimulation on error correction. Although we
believe that the analyses of 15 trials are cleanest, it is important to note that when analyzing all
40-trials (see Appendix B), the effects of stimulation were in the same direction, but less robust.
We can only speculate why this is the case. One possibility is that participants may have
experienced fatigue toward the end of each block, which may have interacted with stimulation
effects. Thus, one caveat to our findings is that they may only apply to early items but not to all
of the items in a challenging task.

Another potential limitation relates to participant blinding to the HD-tDCS condition.
Blinding in the sham condition was not perfect, where perceptions of receiving active was higher
for the stimulation conditions compared to the sham condition, especially for DLPFC. Although
these perceptions were not significantly different between active stimulation conditions, it is
possible that participants perceptions of receiving sham in the sham condition led to poorer error
correction compared to the active conditions. However, these perceptions in the sham condition
would then only be expected to influence error correction during the stimulation block and do
not explain the benefits shown for LTC compared to sham during the pre-stimulation block.

Although individual differences did not appear between stimulation conditions, it is
possible that benefits of stimulation over LTC on error correction compared to stimulation over
DLPFC and sham during the pre-stimulation block reflect a benefit of individual differences that
may have existed between LTC and these other groups that were not measured. Similarly, the
general benefits of stimulation over DLPFC and LTC across all blocks compared to sham may
also reflect individual differences between the active and sham groups. Additional research is
needed to examine whether HD-tDCS can indeed benefit memory for items that are presented
before the stimulation period as well as lead to general enhancement of consolidation and/or
reconsolidation processes.
Lastly, because current is applied over the scalp in an attempt to target a specific neural region of interest, other nearby regions under the electrodes or regions that are functionally connected to the target regions may receive stimulation (for a review see Yavari et al., 2018). Using fMRI, conventional tDCS over the DLPFC has been shown to influence perfusion in the primary sensory cortex, areas of the cingulate cortex, and parietal cortex (Stagg et al., 2013). Thus, it is plausible that other neural regions besides the target region were influenced by stimulation and may have led to benefits in learning. Similarly, the 20-minute stimulation that occurred throughout the block may have influenced other cognitive processes that underlie the targeted region. For example, researchers suggest that DLPFC indexes cognitive control (Koechlin, 2003; MacDonald, Cohen, Stenger, & Carter, 2000) and this cognitive process may have supported feedback-based learning as well. Consequently, we are making reverse inferences regarding the cognitive processes that underlie DLPFC and LTC regions and can only speak to the effects of stimulation over DLPFC and LTC regions, but not to the actual cognitive processes themselves.

Conclusion

The present study examined the influence of MG and PG inductions on feedback-based learning following errors as a function of HD-tDCS over the DLPFC, LTC, and sham HD-tDCS. Goals did not differentially influence learning, however, stimulation over both DLPFC and LTC led to learning benefits compared to sham stimulation. These benefits were shown for items presented across all four blocks even though stimulation only occurred during the 3rd block. Stimulation over the LTC was also shown to benefit learning for items presented during the 2nd
block compared to both DLPFC and sham stimulation. These findings suggest that stimulation over the DLPFC and LTC may facilitate feedback-based learning prior to, during, and after stimulation, and that LTC may be especially beneficial for learning of semantic information that precedes stimulation. Stimulation over both the DLPFC and LTC may reflect top-down elaborative associative and bottom-up semantic processes, respectively, that more generally support feedback-based learning following errors.
Chapter 4: General Discussion

The main goal of the studies in this dissertation was to examine whether and how achievement goals influenced error correction, with a primary focus on the neurocognitive roles of regions within the lateral prefrontal and temporal cortex. Both experiments manipulated mastery goals (MGs) and performance goals (PGs) in the same manner, with MGs defined in terms of developing knowledge and PGs defined in terms of doing better than others. The first experiment approached this question using event-related potentials (ERPs), a correlational method with high temporal resolution, but poor spatial resolution, and the second used high-density transcranial direct current stimulation (HD-tDCS), a causal method with better spatial resolution, but poorer temporal resolution. In this section, results of each experiment will be reviewed followed by a discussion of the inconsistencies between the two experiments in relation to the effects of goals on error correction. This section will end with a discussion on how findings from both studies contribute to the broad literature on achievement goals and the cognitive processes that support learning.

Overview of Results: Experiment 1

In this ERP experiment, although the MG induction was expected to boost error correction at the delayed retest compared to a PG induction, MGs were shown to benefit error correction equally across both the immediate and delayed retest compared to PGs. In terms of where parallel benefits for MGs were observed in the ERP responses to learning feedback, enhanced activity for the MG condition was found at superior frontal sites at both early (200-300
ms) and late (1000-1500 ms) time periods following the onset of learning feedback, both of
which were also predictive of subsequent memory (i.e., Dm effects). Importantly, although the
MG and PG groups both exhibited Dm effects at these sites, suggesting both groups engaged in
the processes associated with this activity, the overall greater amplitude of these waveforms
(regardless of subsequent memory) suggests a quantitative difference in the degree of
engagement between MG and PG groups. In addition, in the mid-latency period (500-1000 ms),
activity along both posterior (i.e., inferior temporo-parietal) and anterior (i.e., fronto-temporal)
sites proximal to the ventral visual stream were predictive of subsequent memory. While no
overall goal differences were observed at these sites, they exhibited evidence of greater
sensitivity to Dm effects in the MG group compared to the PG group, particularly in terms of
predicting memory on the delayed retest. Thus, it appears that MGs facilitate learning through
both qualitative differences in bottom-up processing of the correct answer within the visual
ventral stream and quantitative differences in involving both early and late frontal activity,
putatively associated with initial attention to the correct answer (Blanchet et al., 2007), and later
elaboration of its association with the question (see also Fabiani et al., 1990; Y. Liu et al., 2017).

In terms of processes specific to PGs, the only region where PGs appeared to result in
enhanced activity was over parietal sites during the later time period of learning-feedback
processing, which interestingly was one of the few regions examined that did not exhibit
sensitivity to encoding success. Indeed, late positive-going waveforms in this region have been
previously associated with poorer memory (Y. Liu et al., 2017), possibly indexing the presence
of less successful elaborative processing. In a similar vein, the prediction that successful
encoding under PGs might differentially engage early visual processing in parieto-occipital
regions (Mangels et al., 2017) was not supported, although in the present study this region was
also surprisingly less sensitive to encoding success overall. Moreover, PGs did not appear to result in enhanced attention to performance feedback, regardless of whether considering waveforms associated with more bottom-up (FRN, P3a) or top-down (P3b, LPP) aspects of attention. Thus, while PGs appeared to result in quantitatively less engagement of memory-related processes involving frontal regions, we did not find evidence of unique correlates of encoding for the PG group, unlike in Mangels and colleagues (2017), and no evidence for differential attention to accuracy feedback.

Taken together, these behavioral and ERP results provide support for the general idea that MGs benefit learning compared to PGs through greater elaboration of information (Ikeda et al., 2015; Mangels et al., 2017; Murayama & Elliot, 2011), but provide novel evidence that this benefit can extend to both the immediate and the more long-term updating of semantic knowledge following repeated failures, as well as for the time course and engagement of the mechanisms supporting this learning advantage.

Overview of Results: Experiment 2

In this second study, we targeted HD-tDCS stimulation on two of the regions that Experiment 1 had shown to be predictive of subsequent memory—dorsolateral prefrontal cortex (DLPFC) and lateral temporal cortex (LTC) — in order to test both their causal roles in encoding, and their possible modulation by achievement goals. In both cases, stimulation was applied to the left hemisphere, given the greater involvement of this hemisphere in processing of verbal materials such as those in the current studies, although we note that in Experiment 1, hemisphere effects were often not found, or in some cases favored the right hemisphere.
HD-tDCS over both regions of interest boosted error correction on a delayed retest relative to sham, however no effects of achievement goals were found either in the sham or stimulation conditions, thus failing to replicate the MG advantage found in Experiment 1. Also, somewhat surprisingly, even though HD-tDCS was administered during the 3rd block of questions, benefits of stimulation to left DLPFC and LTC were shown across all blocks, with stimulation of the LTC regions being especially beneficial to learning of items presented in the block that immediately preceded stimulation. The presence of stimulation effects on learning occurring before (and after) stimulation presents a challenge for hypotheses that stimulation only influences the encoding process, and suggests a role for post-encoding processes, including consolidation.

In summary, these findings contribute to the extensive literature supporting the role of the DLPFC and LTC in declarative memory and demonstrate a causal role in the updating of knowledge in semantic memory through successful encoding of feedback. Interestingly, they show that benefits may occur retroactively for items encoded before stimulation, suggesting a role for these regions in post-encoding processes, as well as encoding-specific plasticity. However, the findings from this experiment do not provide clear evidence on the relationship of these processes to achievement goal manipulations, given that the lack of goal effects should be considered in light of the lack of goal differences in the sham condition as well.

Inconsistencies in Behavioral Differences Between Studies

Goal differences on error correction that were found in the first experiment at the immediate and delayed retests were not replicated in the second experiment at the delayed retest.
One difference between the two studies was the absence of an immediate retest in the second study. Indeed, in the first study the difference in errors correction between goals was more apparent under the immediate compared to the delayed retest. Thus, in the second study, only measuring learning at the delayed retest may have led to a decrease in sensitivity to capture differences between goals, suggesting that the influence of goals on learning may be more transient. However, this may be specific to only brief learning opportunities since participants were only able to engage with the learning feedback for two seconds. Goals may still differentially benefit learning given a delayed retest when participants have more time to engage in deeper processing of the learning feedback.

Another factor that could relate to the failure to replicate is the potential differences in accuracy on the first test of general knowledge between groups in the second study; the MG group performed worse than the PG group. This was not the case in the first study, and indeed in that study, MGs even led to a boost in accuracy in relation to the titration level that was less apparent in the PG group. Even though first test accuracy was regressed out of error correction scores, this potentially worse initial accuracy under MG compared to PG may still have impacted error correction. Indeed, first test accuracy was a strong predictor of error correction in both studies ($r > .45$, $p < .001$) and thus, poorer first test accuracy for the students in the MG condition may have reflected a difference in cognitive processes necessary for error correction that could not be captured simply by regressing out these initial performance levels. For example, poorer first-test accuracy may impact or reflect lower levels of engagement, effort, or use of self-regulation strategies that may have impacted both first-test accuracy and later error correction. Perhaps if first-test accuracy had been better matched in the two groups, differences in retest accuracy may have emerged.
Differences in manipulation check ratings between the two studies may also explain why goal inductions were not successful in influencing error correction in the second study. As shown in Table 5, Independent-Samples Mann-Whitney U Tests indicated that reports of mastery adoption (i.e. trying hard to develop knowledge) were similar across goal conditions ($ps > .53$) and that reports of performance adoption (i.e. trying hard to do better than others) was higher under PG than MG ($ps < .002$) in both studies. Thus, difference in goal adoption between goal conditions was similar across studies. However, Related-Samples Wilcoxon Signed Rank Tests within goal conditions show a different pattern between the two studies. Here, reports of mastery and performance adoption was similar under PG in the first study ($p > .68$), but mastery adoption was greater compared to performance adoption under MG in the first study and under both goal conditions in second study ($ps < .04$). Thus, although performance adoption was suppressed under MG compared to PG in both studies, participants in the second study adopted mastery to a greater extent than performance regardless of the goal induction. Unlike the first study, PGs did not lead to similar rates of mastery and performance adoption and instead resulted in boosting mastery over performance. Consequently, greater mastery adoption compared to performance adoption in the second study under both goal conditions may explain why differences in error correction were not evident.

Table 5
Median Goal Adoption Ratings. Participants rated adoption of Mastery and Performance under each goal condition. In the first study, ratings were completed after the first-test on a scale of 1 (not at all) to 6 (very much). In the second study, ratings were completed after each block on a scale of 1 (not at all) to 9 (extremely) that then were averaged across 4 blocks.
<table>
<thead>
<tr>
<th></th>
<th>MG Mastery</th>
<th>MG Performance</th>
<th>PG Mastery</th>
<th>PG Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4.5</td>
</tr>
<tr>
<td>Study 2</td>
<td>5.5</td>
<td>2.5</td>
<td>5.8</td>
<td>5.4</td>
</tr>
</tbody>
</table>

It is important to note some of the possible methodological differences that may have led to inconsistencies in mastery ratings between the two studies. Having the experimenter being present in the room in the second study may have led to greater reports of mastery as students tend to provide higher ratings of mastery adoption to present themselves in a positive light to others (for a review see Daron, Dompnier, & Marijn Poortvliet, 2012). In the first study, the experiment did not have to be in the room since EEG activity had to be monitored outside of the room in which participants completed tasks and answered survey questions, but the experimenter had to be in the room during the second study to administer HD-tDCS while participants completed tasks and answered survey questions. Thus, participants in the second study may have focused more on mastery regardless of their goal induction because the experimenter was present in the room and consequently may have deemed mastery goals to be relevant to the study.

Differences in endorsement of personal goals that were measured prior to goal inductions may also speak to inconsistencies in error correction between studies. Related-Samples
Wilcoxon Signed Rank Tests showed that endorsement of personal mastery goals was greater than endorsement of performance goals for both studies ($ps < .002$) (see Elliot & Murayama, 2008). However, Independent-Samples Mann-Whitney U Tests indicated that endorsement of personal mastery was greater in the first (Median = 4.67) than in the second (Median = 4) experiment ($p < .001$), but there were no differences in endorsement of personal performance goals (First Study Median = 4, Second Study Median = 4, $p > .51$). It may be that the greater endorsement of mastery by participants in the first experiment amplified benefits of the MG compared to the PG induction and that these pre-existing differences between samples from both studies may have led to differential effects of goal inductions on error correction.

Finally, the way in which MGs and PGs were operationalized may not been sufficient to influence reliable learning differences in a long, focused experimental task. In the present study, MGs and PGs were operationalized by focusing students on a goal in relation to MG or PG (i.e. to develop knowledge versus doing better than others) and its evaluation (i.e. informed about how well knowledge was developed versus performance compared to others) at the beginning of each block, but there were no pervasive reminders of these goals while doing the task. Given that classroom definitions of MGs and PGs incorporate differences in the value of making mistakes, role of ability, and/or effort (Ames & Archer, 1988) it may be that operationalizing goal environments in relation to those features may mediate the relationship between goal environments as defined in terms of a goal and learning. For example, it may be that focusing students on interpreting errors as part of learning or as indicators of low ability may lead to adoption of MGs and PGs, respectively, and lead to differences in learning. Further research is needed to examine which aspects of MG and PG environments lead to greater and consistent differences in learning.
Converging/Diverging Evidence of Neural Mechanisms

Across both studies, activity over lateral frontal regions was consistently associated with error correction. ERPs over superior frontal regions and DLPFC activation have both been implicated in elaborative associative processing (e.g., Blumenfeld & Ranganath, 2007; Fabiani et al., 1990; Liu et al., 2017) and may reflect similar processes that are involved in the strengthening of the relationship between the correct answer and its related question. Although ERPs can only speak to an association between involvement of frontal regions and error correction, HD-tDCS over DLPFC demonstrated the causal influence of frontal regions on error correction. Taken together, the involvement of frontal regions on learning from errors supports assertions that this type of learning relies on elaborative processes (Cyr & Anderson, 2015; Huelser & Metcalfe, 2012; Knight et al., 2012). Even though these findings are not surprising, they confirm the role of the frontal regions in updating semantic knowledge. Previous work has primarily focused on associations of word pairs (e.g., Blumenfeld, Parks, Yonelinas, & Ranganath, 2011; X. L. Liu, Liang, Li, & Reder, 2014), whereas the present study establishes the role of frontal regions, along with its causal involvement, in learning of associations that are more complex (i.e. a general knowledge question and its related answer) and more relevant to education. Thus, this evidence can be used to inform students about the learning benefits of elaboration even when it comes to learning from errors.

There were inconsistencies between experiments, however, in how achievement goals influenced activity in this memory-related area. In the first experiment, ERPs were more positive going over lateral frontal regions during encoding of the correct answer when that answer was later corrected compared to not corrected on both the immediate and delayed retest. These same
ERPs were also more positive going under MG compared to PG across both retest delays. However, when applying HD-tDCS stimulation to a left hemisphere DLPFC region corresponding to the sites were these goal effects have been observed in the ERP study, there was no differential improvement in error correction for the MG condition compared to the PG condition. Thus, although both experiments implicated the importance of frontal regions in error correction that supported learning when either an immediate or 1-week delayed retest was employed, MGs were only associated with greater enhancement of frontal activity compared to PGs in the first experiment, where MGs also significantly benefitted behavioral measures of error correction.

One possibility for this inconsistency is that the ERP methods in Experiment 1 were able to detect neural differences specific to small time windows specifically within the presentation of the learning feedback, whereas the HD-tDCS used in Experiment 2 could only look at the overall behavioral effects of stimulation that had occurred throughout the entire test-feedback sequence. Using ERP and fMRI methods, other studies have also found neural differences between goals in the absence of behavioral differences (Lee & Kim, 2014; Mangels et al., 2017), and indeed, the majority of goal-related ERP findings in Experiment 1 persisted even when analyzing the subsample, which had not shown a significant behavioral effect of achievement goals. We also note that Experiment 1 had not found that superior frontal activity was more predictive of subsequent memory in MG compared to PG, but rather, the evidence suggested that these regions were simply more engaged overall in MG, regardless of memory outcomes. Thus, in Experiment 2, it may be that MGs led to greater engagement of frontal regions through greater engagement of elaborative processes compared to PGs, but that HD-tDCS over DLPFC led to similar boosts in these processes for both goals. Finally, the lack of goal-related differences in error correction
even in the sham HD-tDCS may explain why HD-tDCS over DLPFC did not lead to greater learning benefits for MG compared to PG. Further studies where benefits of a goal manipulation were achieved during sham would improve the ability to make conclusions about interaction of achievement goal contexts and the neural regions underlying successful learning.

Across both studies, activity over lateral temporal regions was consistently related to error correction as well. Although ERPs confirmed the presence of LTC in error correction similar to past ERP studies (Butterfield & Mangels, 2003; Mangels et al., 2018; Whiteman & Mangels, 2016), HD-tDCS over the LTC confirmed its causal influence on error correction. Thus, these regions, which other studies have shown to be engaged in semantic processing of verbal stimuli (Gold et al., 2005; Gold & Buckner, 2002), are also a component of the process of updating errors in semantic knowledge with the correct information. Interestingly, although Experiment 2 suggested that this area is similarly engaged in both MG and PG conditions, with the caveat of similar behavioral effects across induced goals, Experiment 1 suggested that this regions provided encoding-specific benefits under MGs regardless of delay, but only short-term benefits to the immediate retest under PGs.

Future Research and Implications

Interestingly, majority of research on environmental achievement goals has relied on surveys to measure both goals and outcomes (for a review see Anderman & Patrick, 2012) making it challenging to understand the causal effects of goals on objective measures. Few studies have examined the influence of goal inductions and even fewer have found goals to influence behavioral measures (e.g., Avery & Smillie, 2013; Avery et al., 2013; Murayama &
Elliot, 2011). Thus, even though both dissertation studies employed the same goal inductions and measured learning in the same manner, other factors (see also Van Yperen et al., 2015) may have influenced the effects of goals and further research is needed to examine which factors are relevant. Educators may need to be informed that the benefits associated with inducing MGs compared to PGs may not always be evident when it comes to behavioral measures such as grades (see Linnenbrink, 2005; but see Dishon-Berkovits, 2014), but that differences in cognitive processes may still exist (see also Lee & Kim, 2014; Mangels et al., 2017). Ultimately, understanding which behavioral and neural measures are sensitive to differences in cognitive processes may be essential to better understanding aspects of learning and how they are influenced by achievement goals.

When it comes to implications for education, the present findings suggest that instructors and institutions could benefit from instilling an MG compared to a PG environment for students. It is important to keep in mind that in the present dissertation participants were not provided with any normative feedback during the completion of the task, but under PG classroom environment students may eventually want to be informed about how they compare to others. Although focusing students on doing better than others may not necessarily impair a focus on learning, providing normative feedback that informs students that they are performing below average can be detrimental (see Cianci et al., 2010). As for learning from errors, the ways in which students engage with the learning information may influence the extent to which they learn from their mistakes. Thus, instructors may want to encourage and provide their students with the time to engage with the corrective information and use elaborative associative strategies to learn from their mistakes.
Appendix A

Supplemental Table

Table A1

P2, P3a, and FRN peak-picked mean latencies (ms) following onset of negative feedback to errors and positive feedback to corrects as a function of goals. (SEM) are shown in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>MG</th>
<th></th>
<th>PG</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Errors</td>
<td>Corrects</td>
<td>Errors</td>
<td>Corrects</td>
</tr>
<tr>
<td>P2</td>
<td>210.47 (7.42)</td>
<td>212.20 (6.26)</td>
<td>206.20 (6.58)</td>
<td>200.93 (7.27)</td>
</tr>
<tr>
<td>FRN</td>
<td>294.53 (9.59)</td>
<td>295.53 (9.22)</td>
<td>288.53 (7.42)</td>
<td>298.07 (9.06)</td>
</tr>
<tr>
<td>P3a</td>
<td>363.53 (10.50)</td>
<td>359.40 (10.08)</td>
<td>370.00 (9.36)</td>
<td>371.73 (9.00)</td>
</tr>
</tbody>
</table>
Appendix B

Supplemental Analyses

Manipulation Check. Although we were primarily interested in whether manipulation check ratings differed between MG and PG, we also examined whether these ratings differed between stimulation groups. Including only stimulation as a factor in an Independent-Samples Kruskal Wallis test revealed no differences between any of the 6 manipulation check (i.e. instruction, importance, goal adoption) items ($p > .09$). A Wilcoxon Signed Rank Test within each Stimulation condition revealed greater ratings of Mastery importance and Mastery adoption compared to Performance importance and Performance adoption, respectively ($p < .001$). As for instruction items, differences between MG and PG were not found within each stimulation condition ($p > .38$). We also included stimulation and goal as factors in an ANOVA to examine presence of Stimulation and Goal interactions on instruction, importance, and goal adoption items. The 3 ANOVAs did not reveal any effects of stimulation ($p > .13$).

HD-tDCS Side Effects. A chi-square test for independence was conducted on each of the 9 side effects (see Table B1) between stimulation conditions. Side effect responses were recoded into a dichotomous variable with sensations being either absent or present (i.e., mild, moderate or severe). Differences were not found between stimulation conditions ($p > .12$).
**Table B1**

Number of participants in each condition who reported side effects during the stimulation period (i.e. Block 3).

<table>
<thead>
<tr>
<th></th>
<th>MG</th>
<th></th>
<th></th>
<th>PG</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DLPFC</td>
<td>LTC</td>
<td>Sham</td>
<td>DLPFC</td>
<td>LTC</td>
<td>Sham</td>
</tr>
<tr>
<td>Headache</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Neck Pain</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Scalp Pain</td>
<td>10</td>
<td>11</td>
<td>8</td>
<td>8</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Scalp Burns</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Tingling</td>
<td>15</td>
<td>16</td>
<td>13</td>
<td>18</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>Skin Redness</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Sleepiness</td>
<td>10</td>
<td>12</td>
<td>9</td>
<td>9</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>Trouble Concentrating</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>13</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>Acute Mood Changes</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>8</td>
<td>8</td>
<td>5</td>
</tr>
</tbody>
</table>

**40 Trial analyses.** We subsequently ran analyses for all 40 trials within each block, however, it is important to note that during the 3rd stimulation block, for 46 participants (22 in MG), at least some items were answered after the 20-minute stimulation was over. Thus, for these participants effects of stimulation during the 3rd block would also capture “offline” effects in addition to the “online” effects. This suggests that for some participants block 3 would speak to both “online” and “offline” influences of stimulation, whereas for the remaining participants who answered all 40 items within the 20-minute stimulation period would only speak to “online” stimulation.
effects. Similar patterns were found for the analyses using 40 compared to 15 trials for each block.

**First-Test:** A 4 (Block: 1-4) x 3 (Stimulation: DLPFC, LTC, Sham) x 2 (Induced Goal: MG or PG) ANOVA on proportion correct showed a main effect of block, $F(1.43, 155.84) = 8.659, p < 0.01, \epsilon = 0.48$, and Goals, $F(1, 109) = 4.781, p < 0.05$. All other effects and interactions were not significant ($ps > .39$) and a Goal by Block interaction approached significance, $F(1.43, 155.84) = 3.348, p < 0.06, \epsilon = 0.48$.

Six pairwise post-hoc comparisons between each of the blocks were conducted to follow up on the main effect of Block, and results were considered significant after Bonferroni correction ($p < .008$). All 3 blocks differed from block 4. Participants answered more questions correctly in Block 1 ($M = .36; SEM = .002$), in Block 2 ($M = .35; SEM = .002$) and in Block 3 ($M = .35; SEM = .004$) compared to Block 4 ($M = .33; SEM = .008$) (Bonferroni corrected $ps < .006$). In addition, participants under MG performed worse overall ($M = .34; SEM = .01$) compared to those under PG ($M = .35; SEM = .01$) (see Figure B1).

These results suggest that although our titration resulted in proportion of correct that was around .35 when examining accuracy across all items, our titration was not able to titrate participants equally under each block and each goal condition. There were no differences in titration between stimulation conditions. Thus, we decided to regress out proportion of correct from proportion of errors corrected for each block.
Fig B1. First-test accuracy across all items. Proportion of correct responses for all question items in each block under Mastery (top) and Performance (bottom) Goals for each stimulation conditions. Stimulation occurred during the 3rd block only. Proportion of correct was significantly worse in Block 4 than the other 3 blocks and overall proportion of accuracy was worse under MG compared to PG. Error bars denote standard error of the mean.
Retest: A 4 (Block: 1-4) x 3 (Stimulation: DLPFC, LTC, Sham) x 2 (Induced Goal: MG or PG) ANOVA on proportion of errors corrected with initial accuracy regressed out revealed a main effect of block, $F(3, 327) = 5.903, p < 0.01$. All other effects and interactions were not significant ($ps > .25$) and the main effect of stimulation approached significance, $F(1, 109) = .898, p < 0.07$.

Six pairwise post-hoc comparisons between each of the blocks were conducted to follow up on the main effect of Block, and results were considered significant after Bonferroni correction ($p < .008$). All 3 blocks differed from block 1 (Bonferroni corrected $p < .008$). Error correction in Block 1 ($M = .44; SEM = .01$) was better than Block 2 ($MEAN = .40; SEM = .01$), Block 3 ($M = .40; SEM = .01$), and Block 4 ($M = .39; SEM = .01$) suggesting that participants benefited from a primacy effect (see Figure B2).

These results suggest that there were benefits in memory for Block 1, but that there were no differences in error correction based on goal. Although the effects of stimulation were significant when examining the 15 items, they merely approached significance for the 40 items. This marginal effect of stimulation showed benefits of DLPFC compared to Sham, (Bonferroni corrected $p < .02$), but LTC showed only a trend toward significance compared to Sham, (Bonferroni corrected $p < .09$). These comparisons did not survive Bonferroni correction ($p < .016$) for the 3 comparisons between stimulation conditions. However, it is worth noting that there was a similar pattern of overall stimulation benefits for the DLPFC condition compared to Sham across both the early 15 and the full 40 item analyses.
Fig B2. Retest accuracy across all items. Proportion of errors corrected with initial proportion correct regressed out for all question items in each block under Mastery (top) and Performance (middle) Goals for each stimulation condition. The bottom graph represents the proportion of errors corrected with initial proportion correct regressed out for all items in each stimulation condition across goals and blocks. Error correction in Block 1 was significantly greater than the other 3 blocks, regardless of stimulation condition. Main effects of stimulation indicated only marginally greater error correction in the DLPFC compared to sham condition. Error bars denote standard error of the mean.

*Block 2 Error Correction:* We also examined whether the LTC benefit during Block 2 on error correction shown for the 15 items existed for the 40 trials by examining block 2 error correction as a function of stimulation. There was a significant effect of stimulation, $F(2, 109) = 3.209$, $p < 0.05$. Three pairwise post-hoc comparisons between stimulation conditions were conducted to follow up on this main effect, and results were considered significant after Bonferroni correction ($p < .0167$). Only LTC in comparison to sham resulted in a significant benefit on error correction (Bonferroni corrected $p < .0162$). This suggests that the 40 item analyses may have masked some of the benefits found in the 15 items analyses.
Bibliography


Students with High Working Memory Capacity. *PLOS ONE, 10*(9), e0137629.  
https://doi.org/10.1371/journal.pone.0137629


https://doi.org/10.1111/j.1751-9004.2012.00457.x


https://doi.org/10.1177/0956797611419520


Murayama, K., & Elliot, A. J. (2011a). 37(10) 1339–1348 © 2011 by the Society for Personality and Social Psychology, Inc Reprints and permission:
sagepub.com/journalsPermissions.nav DOI: 10.1177/0146167211410575


https://doi.org/10.1177/0146167211410575


https://doi.org/10.1177/0146167214523475

https://doi.org/10.3791/50309

https://doi.org/10.3390/brainsci6010007

https://doi.org/10.1016/S1388-2457(03)00044-0

*Nature Neuroscience, 1.*


Neuroscience & Biobehavioral Reviews, 85, 81–92.

https://doi.org/10.1016/j.neubiorev.2017.06.015