

NEUROSCIENCE OF ERROR DETECTION

By

Alena Grigorievna Patsenka

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Psychology

2010

ABSTRACT

NEUROSCIENCE OF ERROR DETECTION

By

Alena Grigorievna Patsenka

Several accounts of anterior cingulate cortex (ACC) function currently exist – a conflict monitoring theory (e.g., Carter et al., 1998; Botvinick et al., 2001), a reinforcement learning theory (e.g., Holroyd & Coles, 2002), and an error-likelihood theory (Brown & Braver, 2005). The theories agree that the ACC is involved in error detection; however, the precise role of the ACC is still a subject of debate. In the conflict monitoring view, the ACC detects a conflict between intended and actual responses. In reinforcement learning view, the ACC is activated when events are worse-than-expected. In error-likelihood theory, the ACC stores information about error-likelihood. The results of the present research challenge the idea that the ACC performs conflict monitoring, and they are consistent with the idea that the ACC responds to worse-than-expected outcomes. The fMRI finding that the same region in the left dACC responds both to errors that people are aware of and to unexpected negative feedback is consistent with the reinforcement learning account and it is problematic for the conflict monitoring theory. The results of the present research also lead to reinterpretation of the findings taken to support the error-likelihood theory.

In the present study, I develop a paradigm in which a memory task and a multiple object tracking tasks are combined. Participants receive feedback on the accuracy of their performance only after both tasks are completed. Negative feedback is ambiguous in terms of what type of error occurred (memory or tracking). After receiving negative

feedback, participants are asked to interpret it. The behavioral results showed that most of the time the interpretations are correct (e.g. when participants make a memory error they indicate that they made a memory error); however, sometimes the interpretations are incorrect (e.g. when participants make a memory error they indicate that they made a tracking error). I assume here that participants interpret feedback correctly if they detect an error and they interpret feedback incorrectly if they miss an error.

Contrasting all errors and correct trials, I identified a region within the left dACC responding greater to errors. A region of interest analysis revealed an interaction between error detection and phase of the trial: The region was more active to detected errors than undetected errors during task performance, but during feedback the pattern was reversed, such that the region was more active to undetected errors than to detected errors. The pattern is problematic for the conflict monitoring theory, because there is no conflict during feedback processing. The results are consistent with the reinforcement learning theory, since both error detection and processing of unexpected negative feedback are worse-than-expected outcomes.

The imaging results also showed that an error related signal within the left dACC is content specific. Comparing the trials on which participants failed to retrieve correct information with the trials on which participants lost track of dots, I found that tracking errors lead to greater ACC activation. Several explanations are possible: One of them is that the ACC contains error-specific neurons similar to feature detectors in visual cortex. Alternatively, participants might have stronger expectations regarding the accuracy of their performance on tracking task than memory task; therefore, violations of stronger expectations result in greater ACC activation.

To my father, Grigori Stepanovich Patsenko, for teaching me to set high standards and achieve goals regardless of how difficult and impossible they might seem.

ACKNOWLEDGMENTS

Most importantly, I would like to thank my adviser Erik Altmann. I realize now that no matter what I write here, it will not express the depth of my gratitude to him. He not only made my dream possible, but insured that the road to it was enjoyable! Erik was an amazing advisor: He set a great example of a true scientist, by being intelligent, thinking critically about ideas, and ready to give up ideas in the face of conflicting evidence. Being always open to new ideas and giving me a lot of freedom in my research, he, at the same time, provided me with strong guidance. Importantly, Erik was very generous in giving praise and encouraging me to keep working. It is hard to overestimate the motivation he gave me when I was buried in doubts in my abilities. I was very happy to have Erik as an advisor, he forever set the highest expectations for people of science for me.

Also I would like to thank members of my Master's and Doctoral committees: Tom Carr, Fred Dyer, Zack Hambrick, Taosheng Lue, and Rose Zacks, for their critique of my work, their comments improved my work greatly and made me prepared for every possible question about it. I am also grateful to them for the creative work environment that they support in the psychology department and cognitive science program at Michigan State University. Special thanks also goes to Bruce Burns for initially inviting me to Michigan State and giving me the opportunity to start a scientific research career.

In addition to faculty members, I am also thankful to graduate students who made these years more enjoyable: John Dewey, Paul Engelhardt, Matt Husband, Graham Pierce, and Emily Swensen (Darowski). Special thanks to Paul Engelhardt, who was not

only an extremely helpful colleague, advising me in all possible aspects of graduate life, but also he became a close friend throughout these years. I am also thankful to my non-academic friends, Nadia Karatkevich, Marina Kaspirovich, Lena Luchaninova, Oksana Naumenko, Kathy and Sebastian Seifarth, Anna Senkina, Paul Tucci for diversifying my life with parties, wine-tasting, dancing, and conversing with me on “human” subjects.

Finally, I want to thank my family. I am grateful to my parents, Galina Nikolaevna and Grigori Stepanovich Patsenko, for teaching me how to think. They taught me to set high standards and achieve impossible goals. They believed that one could become whoever they want if the desire meets the effort. I am also grateful to my sister, Natasha Patsenko (Sidorova) for setting an example in every aspect of a nonacademic life. I give my special thanks to my aunt Angelina Stepanovna and uncle Isaac Aronovich Garbar for making this “trip” to the U.S. possible. But no one deserves more thanks than my boyfriend, Yuri Shalagin! I am so grateful to you for believing in me even when there was no basis for it, or when you could not possibly know if there was. Thank you for taking me on trips in the most inconvenient time, and thank you for going to parties alone when I absolutely could not make it. You are always in my heart!

TABLE OF CONTENTS

LIST OF TABLES.....	ix
LIST OF FIGURES.....	x
1 INTRODUCTION.....	1
1.1 The ACC Involvement in Error Detection.....	2
1.1.1 Mismatch (or comparator) theory of ACC function.....	2
1.1.2 Conflict monitoring theory of ACC function.....	4
1.1.3 Reinforcement learning theory of ACC function.....	8
1.2 Overview of Current Research.....	12
2 ERROR DETECTION IN COMPLEX TASKS.....	15
2.1 Experiment 1: Behavioral study of error detection.....	15
2.1.1 Method.....	16
2.1.2 Results and Discussion.....	19
2.2 Experiment 2: Imaging study of error detection.....	21
2.2.1 Method.....	22
2.2.2 Behavioral Results.....	29
2.2.3 fMRI Results.....	31
2.2.3.1 Whole brain analysis: all errors versus correct trials during Task phase.....	31
2.2.3.2 ROI analysis dACC during Task phase.....	33
2.2.3.3 ROI analysis dACC: Task versus Feedback.....	38
2.2.3.4 Whole-brain analysis: Interaction between Consistency and Time point.....	41
3 GENERAL DISCUSSION.....	44
3.1 The dACC in response to worse-than-expected outcomes.....	44
3.1.1 Evidence concerning the Conflict monitoring theory.....	46
3.1.2 Evidence concerning the Reinforcement learning theory.....	49
3.1.3 Evidence concerning the Error-likelihood learning.....	50
3.1.4 Integrating the finding from ERP and neuroimaging studies.....	53
3.2 Content-specific signal in the ACC.....	56
3.3 ACC activation in tasks without response conflict.....	59
4 FUTURE DIRECTIONS.....	60
4.1 Does the ACC respond to better-than-expected outcomes?.....	60
4.2 Violations of expectations in the context of anxiety.....	61
4.3 Negative emotions and rate of learning.....	62
4.4 Lingering effect of emotions.....	64
4.5 Lingering effect of emotions in a context of anxiety.....	65
5 CONCLUSIONS.....	67

APPENDIX A: Analysis of regions other than the left dACC.....	68
APPENDIX B: Correct memory phase versus correct tracking phase.....	79
REFERENCES.....	80

LIST OF TABLES

Table 1. Group analysis results for the contrast: “all errors > all correct”	32
Table 2. Group analysis results for the contrast: “Consistently interpreted errors > correct”	37
Table 3. Group analysis results for the interaction between Consistency and Task Phase.....	42
Table A1. ROI analysis: Percentage BOLD signal change observed during feedback processing for consistently and inconsistently interpreted tracking and memory errors in a small region in the dACC (a) and in the left SFG (b). Repeated measures ANOVAs showed neither the main effects (Consistency or Error Type) nor the interaction between factors.....	71
Table A2. Group analysis results for the contrast: “Consistently interpreted tracking errors > correct”	74
Table A3. Group analysis results for the contrast: “Inconsistently interpreted errors > Consistently interpreted errors” during the Feedback phase.....	76

LIST OF FIGURES

Figure 1. An example of a trial with a negative feedback in a modified multiple object tracking task.	19
Figure 2. Frequency of different types of interpretations as a function of error type in Experiment 1	21
Figure 3. An example of a trial with a negative feedback in a modified task used for imaging study.....	26
Figure 4. Frequency of different types of interpretations as a function of error type in Experiment 2.....	30
Figure 5. Error-related activation. Regions identified in a whole brain analysis contrasting all error trials and all correct trials ($Terr_t + Terr_m + Merr_t + Merr_m$) – 2 ($Tcor + Mcor$). The arrow points to a region in left dorsal ACC, which is more active for error trials.....	31
Figure 6. ROI analysis: the left dACC. The region shows Error Type effect: tracking errors activate the region greater than memory errors, consistent with the idea that the ACC represents information regarding types of errors people make. There is no effect of Interpretation Type. The interaction of Error Type and Interpretation Type approaches significance, suggesting an effect of Consistency of interpretation: the ACC responds greater to consistently interpreted errors ($Terr_t$ and $Merr_m$) compared to inconsistently interpreted errors ($Terr_m$ and $Merr_t$). The error bars show SEM.....	34
Figure 7. Estimated Hemodynamic Response Functions (HDR) for different types of error during the Task phase. The graph shows the effect of Error type – blue lines above red lines, and the effect of consistency – solid lines above dashed lines.....	35
Figure 8. ROI analysis: the left dACC. The region shows Consistency effect: Errors interpreted consistently lead to a greater dACC activation than errors interpreted inconsistently. There is also an effect of Error Type. The interaction between Consistency of Interpretation and Error Type is not significant. The error bars show SEM.....	36
Figure 9. Consistently interpreted errors. Regions identified in a whole brain analysis contrasting consistently interpreted errors and correct trials ($Terr_t + Merr_m$) – ($Tcor + Mcor$). The arrows point to the regions more active for consistently interpreted errors. They are two regions in the left dACC (big and small), the left MFG, and the left SFG.....	37
Figure 10. ROI analysis: the left dACC. The region shows Consistency by Time Point interaction: During task performance, errors interpreted consistently lead to a greater dACC activation than errors interpreted inconsistently. However, during feedback processing, errors interpreted inconsistently lead to greater dACC activation than errors	

interpreted consistently. Neither of the main effects is significant. The error bars show SEM.....	39
Figure 11. ROI analysis: the left dACC. The region shows Consistency effect during feedback processing only for tracking errors: Errors interpreted inconsistently lead to greater dACC activation than errors interpreted consistently. The effect of consistency is not significant for memory errors. The error bars show SEM.....	40
Figure 12. Estimated HDR functions for all types of error during the Feedback phase...	41
Figure 13. Consistency by Time point interaction. Regions identified in a whole-brain analysis contrasting (Terr_t + FTer_m) and (Terr_m + FTer_t). The arrows point to the regions more active for (Terr_t + FTer_m). They are the left MFG and the left Caudate.....	42
Figure 14. Summary of the results from three studies: Holroyd & Coles, 2002 (A), present study (B), and Brown & Braver, 2005 (C).....	54
Figure A1. ROI analysis: a big region of the left dACC. During negative feedback processing, the region shows Consistency effect for tracking errors, but not for memory errors. The error bars show SEM.....	70
Figure A2. Consistently interpreted tracking errors. Regions identified in a whole brain analysis contrasting consistently interpreted tracking errors and correct trials (Terr_t – Tcor). The arrows point to the regions more active for consistently interpreted tracking errors. They are the dACC bilaterally, the left medial frontal gyrus (MFG), the left superior frontal gyrus (SFG), and the inferior parietal gyri (IPG) bilaterally.....	72
Figure A3. ROI analysis: consistently interpreted tracking errors versus inconsistently interpreted tracking errors during feedback processing in brain regions identified with the contrast (Terr_t – Tcor). Neither of the comparisons reaches significant level. Thus the bilateral IP, the left MFG, and the bilateral ACC do not show a consistency effect during feedback processing. (The region in the left ACC described here is more anterior and inferior compared to the regions in the left ACC identified in “all errors vs. correct trials” contrast and in “consistent interpretations vs. correct trials” contrast.).....	74
Figure A4. Feedback: Inconsistently interpreted errors. Contrasting inconsistently interpreted errors and consistently interpreted errors during Feedback (FTer_m + FMer_t) – (FTer_t + FMer_m), I find a single area in the left IPG responding greater to inconsistently interpreted errors.....	76
Figure A5. ROI analysis: the left dACC. The region shows an effect of the Trial Phase: correct tracking phase is associated with a greater dACC activation compared to correct memory retrieval phase.....	79

“There are only two mistakes one can make along the road to truth:
not starting and not going all the way.”

Buddha

1 INTRODUCTION

Human performance is far from being infallible. Errors pervade all our activities: while typing one might press a wrong key; while preparing a cup of tea one might pour water in a bowl instead of a cup; while solving a problem one might follow a wrong path. Detection of an error, or deviation from accuracy, allows people to adjust their behavior correspondingly. These adjustments could have a general nature, for example, people could slow down after a typing error. However, errors often require task specific adjustments such as narrowing attentional focus on a central letter in the Eriksen task or processing ink color instead of word reading in the Stroop task (the tasks will be described later). The question is how people know what adjustments are appropriate in a current situation. One possibility is that error itself is informative in terms of how behavior should be modified. The other is that an error signals only generically that something has failed and people rely on various heuristics to determine what went wrong and what adjustments need to be made.

In spite of the pervasiveness of errors in behavior and their role in learning and performance, there is relatively little behavioral research on the topic (e.g. Norman, 1981; Reason, 1990). Within the last decade, following the discovery of a neural signature for errors (Falkenstein, Hohnsbein, Hoorman, & Blanke, 1991; Gehring, Goss, Coles, Meyer, & Donchin, 1993), there has been an explosion in neuroscience research on error detection mechanisms (e.g. Botvinick, Braver, Barch, Carter, & Cohen, 2001; Carter, Braver, Barch, Botvinick, Noll, & Cohen, 1998; Holroyd & Coles, 2002; Miltner, Braun, & Coles, 1997). An area in the medial frontal cortex, the anterior cingulate cortex (ACC), has been found to be active during error commission (Carter et al., 1998). Understanding

the nature of the error signal that the ACC carries – whether or not the signal is content-specific – could provide insights into the question of how the system determines what went wrong and how to adjust behavior following an error.

1.1 The ACC Involvement in Error Detection

Research accumulated during the past decade agrees that the ACC is involved in error detection; however, the precise role of the ACC is still a subject of debate. Several theories of ACC function have been proposed; among them are a mismatch theory, a conflict monitoring theory, and a reinforcement learning theory. None of the theories addresses the question of whether the ACC carries content-specific information, the information about the type of an error.

1.1.1 Mismatch (or comparator) theory of ACC function.

In the mismatch view, the medial frontal cortex detects a mismatch between actual and intended responses. The theory was proposed by Gehring and colleagues (1993) who found a negative going deflection in event-related potentials (ERPs) originating in the medial frontal cortex following an error commission. The error related negativity, known as ERN, correlated with a number of measures of compensatory behavior.

In this study, participants performed a version of the Eriksen flanker task. On each trial, participants viewed an array of letters (e.g., HSHH) and their task was to respond to the central letter by squeezing a dynamometer in the corresponding hand (e.g., left hand for S and right hand for H). On congruent trials, all letters in the array were identical – HHHHH or SSSSS; on incongruent trials, the center letter was different from

the flankers – HSHH or SSHS. As in previous studies (Eriksen & Schultz, 1979), incongruent trials resulted in lower accuracy and higher reaction time (RT) compared with congruent trials. In this study, the authors found a sharp negative going deflection 100 ms after the onset of electromyographic (EMG) activity in the limb that was about to make an error. Dipole modeling suggested a medial frontal generator of the negativity, possibly the ACC (ERPs do not allow conclusions regarding precise spatial localization).

The relation of the ERN to error detection is also evident from the correlations of the amplitude of the signal with three measures of compensatory behavior. First, the authors found that large ERN was associated with greater probability of correcting the error on the current trial. Second, the amplitude of ERN positively correlated with error force (or amplitude of error squeeze measured in Kg). Finally, larger ERN was associated with higher RT on the next correct trial. The association of the frontal negativity with error trials and measures of compensatory behavior suggested that ERN was a neural signature of an error detection mechanism.

The idea that ERN reflects a mismatch between actual and intended response was largely driven by the nature of errors in the task used in the study. In the Eriksen flanker task, the majority of errors are caused by impulsive behavior, when a response is issued before the processing of the central stimulus is complete. The authors hypothesized that an error is detected when an efferent copy of the actual response is compared with the correct response that soon becomes available.

The effects of error detection, discussed within this work, have a general nature. The system adjusts the overall speed of processing. The trials following error trials have usually higher RTs and probably as a result of that higher accuracy.

1.1.2 Conflict monitoring theory of ACC function.

In the conflict monitoring view, the ACC detects a conflict or competition between several simultaneously activated alternatives, and errors are just a special case of conflict. The theory was proposed by Carter and colleagues (1998) who found that the same area within the ACC is active on both types of trials: error trials and correct trials that involve response conflict. The ACC is believed to detect a presence of conflict and transmit this information to the areas responsible for cognitive control, for example, prefrontal cortex (PFC) (Botvinick et al., 2001). Kerns and colleagues (2004) provided evidence for that idea showing that the ACC activation on a current trial was positively correlated with PFC activation on a subsequent trial.

The ERP study done by Gehring and colleagues (1993) suggested that the medial frontal cortex, possibly the ACC, responds to errors. A methodology that has a better spatial resolution than ERPs was required to confirm that the ACC was the locus of the activation on error trials. Carter and colleagues (1998) used fMRI to address the question. In their study, participants performed an AX-CPT task. In this task, the sequence of events was as follows: First, participants saw a letter A or a letter B, which could be followed by the letters X or Y. The participants' task was to press a button when the sequence AX was presented, otherwise the response should have been withheld. The authors found that an area within the ACC was more active on error trials (e.g., when

participants responded to AY, BX, or BY sequences) compared with correct trials. This finding confirmed the idea that the ACC responds to errors.

In addition to comparing error and correct trials, Carter and colleagues (1998) also examined the ACC activation on correct trials alone. They separated all correct trials into high conflict and low conflict trials. Conflict here is simultaneous activation of responding behavior and withholding a response. AY or BX trials were classified as high conflict trials, because, although the participants had to withhold response on these trials, letters A and X could activate responding at least partially. BY trials were classified as low conflict trials, because participants were to withhold their response on these trials and neither of the letters was associated with pressing a button. The authors found that the same area of the ACC that was active on error trials was also more active on high conflict correct trials compared to low conflict correct trials. Carter and colleagues (1998) concluded that the ACC performs a general conflict monitoring function and errors are another instance of conflict.

Botvinick and colleagues (2001, 2004) elaborated the idea that the ACC serves a performance monitoring function and proposed that conflict (or error) detection triggers top-down control mechanisms. Thus, episodes of high conflict engage cognitive control, whereas episodes of low conflict relax it. Botvinick and colleagues (2001, 2004) modified computational models of the tasks that usually show ACC engagement (e.g., Eriksen flanker task, the Stroop task) by including a conflict-monitoring unit. In their model, the conflict-monitoring unit computes level of conflict on a current trial and transmits this information to a control unit, which is biasing processing on the following trial.

Thus, in the Eriksen flanker task described earlier, the conflict arises on incongruent trials – HSHH or SSHS – when flankers and the central letter simultaneously activate responding with left and right hands. The conflict-monitoring unit registers the level of conflict and transfers the information to the control unit, which biases processing to the central stimuli on the next trial. In the Stroop task, participants are presented with words written in different ink colors and their task is to name the ink color and ignore the words. The conflict, in this task, arises when participants are presented, for example, with the word RED written in blue color. Although the task is to name the ink color – blue – word reading being automatic can partially activate an incorrect response – red. A conflict-monitoring unit registers the presence of the conflict and transmits this information to control units, which biases processing toward color naming on the next trial.

In the model proposed by Botvinick and colleagues (2001, 2004), the level of control is not fixed as in previous models, but is flexibly modified throughout the course of a session by the presence of conflict. This addition allows the model to accommodate a number of effects reported previously: for example, sequential adjustment effects in the Eriksen task and the effect of trial-type frequency in the Stroop task. One of the sequential adjustment effects in the Eriksen task is that an incongruent trial preceded by another incongruent trial (iI trial) has lower RT than an incongruent trial preceded by a congruent trial (cI trial). A current incongruent trial (I) seems to have a benefit of control being engaged by the preceding incongruent trial (i). The Stroop model simulates the effect of trial-type frequency on performance. The difference in performance (i.e. RTs and accuracy) between incongruent and congruent trials is larger when incongruent trials

are infrequent (20%) compared with when incongruent and congruent trials are more balanced (50%). Thus the level of control is not hard-coded by an experimenter in the models, but is a result of the amount of conflict accumulated throughout an experiment.

In the simulations, the conflict-monitoring unit influences the strength of the control and it does not provide information regarding the content of the adjustment that need to be made. Its influence is quantitative rather than qualitative. The nature of the adjustments is task specific and is determined a priori by the experimenter. The system is never faced with the question of what adjustments are appropriate in a current situation. In the Eriksen task, the control units bias processing toward the central letter, whereas in the Stroop task, the control units bias processing toward color naming. The assumption of the model is that the conflict monitoring units provide only non-specific information regarding the occurrence of conflict. However, the authors noted that “conflict monitoring need not be limited in this way” and that it can convey qualitative information regarding the content of conflict.

Botvinick and colleagues (2001, 2004) proposed that conflict monitoring and cognitive control are related based on the behavioral findings that episodes of conflict are usually followed by superior performance in terms of RTs and accuracy. Kerns and colleagues (2004) provided imaging evidence for the idea. They showed that the ACC activation on a current trial was positively correlated with behavioral adjustments and with PFC activation on the subsequent trial.

In this study, the authors found that iI trials in the Stroop task activated the ACC significantly less than cI trials (see also Botvinick et al., 1999), showing that the system

experiences less conflict on the incongruent trial (I) preceded by another incongruent trial (i). Kerns and colleagues (2004) further explored the effect of the ACC activation on the adjustments in control by dividing all iI trials into low- and high-adjustment trials. Trials that had RTs above the median RT for all iI trials were classified as low-adjustment trials, and trials that fell below the median were classified as high-adjustment trials. High-adjustment trials were associated with greater ACC activation on the previous trial. Thus, the more active the ACC is on the preceding trial, the greater the benefit on the current trial. Moreover, the authors showed that the increase in the ACC activation on the preceding trial was significantly correlated with the increase in PFC activation on the current trial.

These findings supported the idea of Botvinick and colleagues (2001, 2004) that the conflict monitoring function presumably performed by the ACC communicates the information about the occurrence of conflict to control mechanisms implemented in the PFC. However, the question of whether the ACC merely modulates the strength of the PFC or provides a content-specific signal remains to be addressed.

1.1.3 Reinforcement learning theory of ACC function.

In the reinforcement learning view, the ACC receives rather than generates an error signal (e.g. Holroyd & Coles, 2002). The errors are detected by mesencephalic dopamine system, which sends an error-related signal to the ACC to train it to select an appropriate motor controller (e.g., amygdala, dorsolateral prefrontal cortex – DLPFC cortex). Thus, in this view, the ACC serves as a motor control filter.

The reinforcement learning (RL) theory is an attempt to unify two bodies of literature: reinforcement learning and error detection (or conflict monitoring). According to the theory, errors are detected by basal ganglia that compute a difference between an expected value of the ongoing events and the outcome. When ongoing events are worse than expected basal ganglia issue a signal to the ACC. This signal trains the ACC to select the motor controller that has greater chances to be successful at carrying out a task. The motor controllers mentioned in the model are amygdala, DLPFC, and orbitofrontal cortex (OFC). Each of the motor controllers could solve a problem or perform a task in a different fashion. The ACC decides which of the motor controllers will have an influence on the motor system. In the model, the ACC does not necessarily have to choose one of the controllers over the others; it can also combine input of several controllers in a particular manner. Thus, in this view, the ACC performs an action selection function rather than a performance monitoring function.

Experimental support for the RL theory mainly comes from an ERN-like signal produced by the medial frontal cortex in response to error (or negative) feedback. The signal was first observed by Miltner and colleagues (1997) and received the name feedback error-related negativity (fERN). In this study, participants performed a time interval production task. They were asked to press a key one second after the onset of a trial. Feedback indicated whether the accuracy was in an acceptable range. The authors found a negative going deflection in ERPs after the onset of negative feedback at about 250 ms. Dipole modeling suggested a medial frontal generator for the feedback error-related negativity. Based on the similarity in the ERN and the fERN locus, Miltner and colleagues (1997) concluded that the same cognitive process underlies both signals.

Whereas the fERN represents a problem for the conflict monitoring theory as there is no competition at the feedback point, the RL theory accounts nicely for it. Moreover, it explains why the ERN propagates from feedback to response with learning as was shown by Holroyd and Coles (2002). In this study, participants were asked to press the left or right button in response to a stimulus. The stimulus-response mapping was not given to the participants and had to be inferred in a course of the experiment. There were six stimuli in total. One stimulus was mapped onto the left button and another was mapped onto the right button – 100% mapping condition. For the other two stimuli, the feedback was delivered independently of participants' performance: on half of the trials, they received a positive feedback and, on another half of the trials, they received a negative feedback – 50% mapping condition. On the trials when the fifth stimulus was presented, participants always received a positive feedback – always correct condition. Finally, on the trials when the sixth stimulus was presented, participants always received a negative feedback – always incorrect condition.

At the beginning of the experiment, the ERN was registered following the negative feedback for all conditions. However, as participants progressed through the session and learning occurred, the ERN moved from feedback to the point of response in certain conditions. Thus, on trials with 100% mapping, the amplitude of ERN was larger at the response than at the feedback after learning has occurred. As the system acquired the stimulus-response mapping, it began to rely more on its own representations of what the correct response should be rather than the accuracy feedback. In contrast, on trials with 50% mapping, the ERN continued to be elicited primarily by the feedback throughout the session. The results of the simulation were in agreement with ERP data.

The RL theory of ACC function (Holroyd & Coles, 2002) has advantages over the conflict monitoring and the mismatch theories as it accounts for both the ERN and the fERN. However, the theory does not model any explicit error-detection mechanism. Information about accuracy is provided to the model by the experimenter.

In the model, basal ganglia send a non-specific signal to the ACC about the occurrence of an error in the form of a phasic reduction of dopaminergic input to the area. The signal punishes the ACC for its choice of a motor controller on the current trial. It is not content specific itself, although it is translated into a preference for one motor controller – DLPFC – over another – amygdala. However, the signal is not informative in terms of what particular representation or task set should be strengthened within the DLPFC, for example.

* * *

To summarize, the models described above – the mismatch model, the conflict monitoring model, and the reinforcement learning model – differ in terms of a particular function the ACC supports – mismatch detection, conflict monitoring, or selection for action. However, all three models agree that the ACC issues a signal in response to an error occurrence. The error-related signal, as it is currently implemented in the models, is not task-specific. The present work aims to extend the models by addressing the question of whether the signal that the ACC issues is content specific. Answering this question could shed light on to the larger problem of how people decide what went wrong and what behavioral adjustments are appropriate in the current situation. If the ACC does

carry qualitative information, that would suggest that an error signal itself is informative in terms of the type of the failure that has occurred.

1.2 Overview of Current Research

The aim of the current work is to examine neural substrates of an error detection mechanism in the context of a complex task in which different types of errors can potentially occur and the system is faced with a decision regarding what went wrong and how to adjust behavior appropriately. In particular, I am interested to see whether an area in the medial frontal cortex often associated with error detection -- the ACC -- discriminates between different types of error and/or discriminates between the decisions people make regarding the type of an error. If the ACC discriminates between error types then it can potentially be one of the sources of information that higher-level areas (e.g., PFC) use to make decisions regarding what went wrong. However, if the ACC discriminates between decision types, or different interpretations of an external error signal, then it can potentially be the locus of such a decision process itself.

To address these questions I developed a paradigm in which a memory task and a multiple object tracking task are combined. This paradigm creates a situation where multiple things can go wrong -- memory or tracking failures. Participants receive feedback on the accuracy of their performance only after both tasks are completed. Negative feedback is ambiguous as it informs participants that they were inaccurate but does not specify what type of error has occurred (memory or tracking error). Participants are given a chance to correct a mistake if it occurs. The corrective actions suggest how

participants interpreted the negative feedback—whether they believed they made a memory or tracking error.

In this paradigm, there are two independent variables: error type and interpretation type. There are trials on which participants correctly interpret negative feedback. For example, they make a memory error and classify it as a memory error. There are also trials on which participants incorrectly interpret negative feedback. For example, they make a memory error, but classify it as a tracking error. The same applies to tracking errors. This dissociation between error type and interpretation type allows me to address the question regarding the place of the ACC in the error processing stream - the question of whether the computations that the ACC supports occur prior to the judgment about the type of error or after the judgment is made. If the ACC responds to the actual error type then it is likely to be one of the sources of information that higher-level areas (PFC) use to make the judgment about the error type. Other sources of information could be error history (e.g., memory errors are more frequent in this task), or various explicit heuristics (e.g., “classify an error as tracking if the tracking task was long”). However, if the ACC responds to interpretation type, then the computations that it supports occur either after the judgment is made (and the ACC potentially selects corrective actions) or it is one of the areas that actually make such a judgment.

In a behavioral experiment – Experiment 1 in this document – I test the paradigm. In particular, I address the question of whether people can make accurate judgments about the error type in the complex task I developed. In the imaging experiment – Experiment 2 in this document – I address two questions: The first concerns the neural source of the error signal that differentiates between error types. Does the ACC respond

differentially as a function of error type? The second question concerns the neural source of participants' interpretation of negative feedback. Does the ACC respond differentially as a function of interpretation type?

In the present research, I assume that there are potentially different kinds of errors that the mind can make. Hypothetically, there are as many types of error as basic cognitive and perceptual processes. I chose here memory and tracking errors, primarily because memory and tracking processes are well-studied. We know that relatively different neural networks support memory retrieval and tracking suggesting that these are two different processes (e.g. Jonides, Lewis, Nee, Lustig, Berman, et al., 2008 – for memory; Culham, Cavanagh, & Kanwisher, 2001 – for tracking). So, the assumption is that different processes produce qualitatively different errors. However, it is also possible that in spite of the variability in the functions that the mind performs, all errors are detected similarly, for example, as mismatches between two or more representations (e.g. Gehring, Goss, Coles, Meyer, Donchin, 1993). Therefore, the question of what is the universe of errors undoubtedly requires future research.

2 ERROR DETECTION IN COMPLEX TASKS

2.1 Experiment 1: Behavioral Study of Error Detection

In Experiment 1 I addressed the question of whether people are able to make informed decisions about what kind of error they make. For that purpose, I developed a paradigm in which a working memory task and a multiple object tracking task were combined to create a complex situation in which multiple things can go wrong (i.e. working memory or tracking failures). In this task, delayed error feedback indicates only generically that something went wrong. Participants had to decide on their own what kind of error had occurred and make corresponding behavioral adjustments.

If people have access to the information regarding the type of error they have made on each trial, then I would expect participants' interpretations of a negative feedback would correspond to the error type. That is, participants are expected to correct working memory errors as working memory errors and tracking errors as tracking errors.

If, however, people do not have access to the information regarding the type of error, then they might either guess what the error was or they might attempt to match probabilities. In the case of guessing, the distribution of interpretation types should not differ from random distribution. In the case of matching probabilities, the frequency of interpretation types are expected to match the frequency of error types regardless of what particular kind of error has occurred. For example, if working memory errors are more frequent than tracking errors, then participants might correct most working memory errors as working memory errors and, more importantly, most tracking errors as working memory errors.

2.1.1 Method

Participants. Fourteen undergraduate students from the Michigan State University Psychology Department subject pool participated in exchange for course credit. All participants had normal or corrected-to-normal vision.

Apparatus. Stimuli were presented on a 19" CRT display monitor with a resolution of 1024 x 768 pixels and a refresh rate of 117 Hz. Participants were seated at a viewing distance of 80 cm from the computer display, with their head positioned on a chin rest. A viewable screen subtended an angle of 28° by 21°. Stimulus presentation and response collection were controlled by Experiment Builder software (SR Research Ltd., Mississauga, ON, CA).

Materials. In each trial, nine dots were displayed on the screen. The dots consisted of blue outline rings with red interiors and were displayed on a white background. The dots were opaque so that when two dots intersected each other, one was shown to be in front of the other. The interior of a dot was 1.363 degrees of visual angle; the outline ring was 0.055 degrees thick. Four of the dots formed a square, three dots formed a triangle, and two dots were placed elsewhere on the screen (Figure 1). The dots within the square and the triangle were separated by approximately 4.226 degrees of visual angle. The location of the shapes – left or right from the center of the display – varied randomly from trial to trial.

During the tracking phase of a trial, each dot moved with a velocity and direction that were changed at random. The velocities of the dots ranged from 1 to 9 degrees per second. The directions were chosen from among eight equal divisions of the compass.

The rate of velocity and direction changes ranged from 150 to 300 ms. When a dot reached an edge of the screen, the x or y velocity vector reversed its value, so that the dot appeared to bounce off the edge. The dots' trajectories were generated in real time during each trial.

The accuracy feedback was indicated by a colored square (Figure 1, Feedback panel). For half of the participants, a blue square meant "correct", and a green square meant "incorrect"; for the other half of the participants, the meaning of the color feedback was reversed. The square appeared in the center of the screen for 2000 ms and subtended 1.772 degrees of visual angle.

Procedure. Participants were tested individually. At the start of the one-hour session, they read through a description of the task. Then they were given five practice trials. After the practice trials, the participants had an opportunity to ask questions about the task. Eighteen blocks of eight trials each were then presented, totaling 144 trials per participant.

Each block of eight trials started with an encoding phase, during which eight randomly chosen shapes – squares and triangles - were presented sequentially (Figure 1). Each shape was displayed for 500 ms. The participant's task was to remember the sequence. The sequence instructed participants what shape would be relevant on each of the eight trials of the block: the first shape in the series was relevant in the first trial, the second shape in the second trial, and so on.

The beginning of each trial was marked with a fixation cross, which was displayed for 2,500 ms. Then the participant was shown a display consisting of nine dots: three dots arranged in a triangle, four dots arranged in a square and two dots placed

elsewhere in the display. That was the memory retrieval phase. The participant's task was to retrieve the shape that was relevant on that trial, and attend to the dots that formed that shape. For example, if participants saw the sequence of shapes displayed in Figure 1, then on the first, second, fifth, seventh, and eighth trials, they had to attend to dots that formed a triangle; and on the third, fourth, and sixth trials, they had to attend to dots that formed a square. The participant had 4 s to choose the correct shape, after which all dots began moving randomly on the screen. During the tracking phase, the participant's task was to keep his/her attention only on the relevant dots. The tracking phase was randomly assigned a duration of 5.0 s, 7.5 s, or 10 s. After all dots stopped moving, the participant was asked to indicate with mouse clicks which dots they tracked. As the participant clicked on a dot, an "x" sign appeared, covering the selected dot. The sign stayed on display until the end of the trial, telling the participant which dots they had selected. A color feedback then appeared, indicating whether they were accurate. After a positive feedback, the trial ended. After a negative feedback, the participant had a chance to correct the mistake by choosing dots one more time.



- Before a block of 8 trials

Encoding task

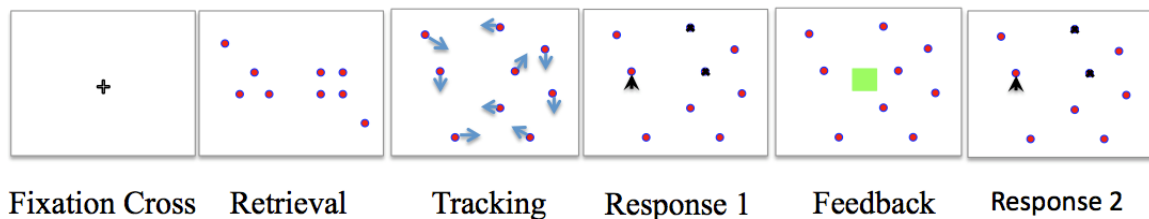


Figure 1. An example of a trial with a negative feedback in a modified multiple object tracking task. (For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.)

2.1.2 Results and Discussion

In this experiment, two types of error occurred: working memory and tracking errors. I classified a response as a working memory error if participants clicked on three dots during Response 1 (Figure 1) when a square was relevant on that trial; or when they clicked on four dots when a triangle was relevant on the trial. A response was classified as a tracking error when participants clicked on the correct number of dots, but one or more dots were not a part of the to-be-tracked subset. Participants were considered to have made both types of errors when they clicked on the wrong number of dots and at least one of those dots did not belong to the to-be-ignored shape.

The participants made errors on 33% of the trials. Memory errors were almost twice as frequent as tracking errors. Participants forgot what shape they had to attend to on 18% of the trials (middle bar in Figure 2). They failed to track dots accurately on 10% of the trials (left most bar in Figure 2). On five percent of the trials, participants made both types of error (right-most bar in Figure 2). First, they chose a wrong subset of dots

to track during the retrieval phase and then they lost a dot or several dots during the tracking phase.

Following negative feedback, participants had a chance to correct their errors by choosing dots one more time during Response 2 (Figure 1). The corrections participants made were informative in terms of the error interpretations they made. Two types of interpretations were possible: working memory interpretations and tracking interpretations. A working memory interpretation means that a participant realized they had tracked a wrong shape and decided to switch to a different shape. An interpretation was classified as working memory if the number of clicked dots differed between Response 2 and Response 1. In other words, a working memory interpretation was registered if, during Response 1, a participant chose three dots and during Response 2 they chose four dots and vice versa. A tracking interpretation occurred when a participant believed that he or she chose the correct shape to track, but lost one or more dots during the tracking task. An interpretation was classified as tracking if the number of clicked dots in Response 1 and Response 2 was the same.

The corrections that the participants performed following a negative feedback suggest that people make informed decisions regarding what kind of error has just occurred. Figure 2 shows that 72% of the memory errors were interpreted as memory errors (grey shading of the middle bar), and that 67% of the tracking errors were interpreted as tracking errors (white shading of the left most bar). A Chi-square test indicates that types of behavioral adjustments depended on types of error made $\chi^2(1) = 5.78, p < .05$. Not surprisingly, half of the trials on which participants made both types of

error, were interpreted as trials with memory errors and half of the trials were interpreted as trials with tracking errors.

In sum, the paradigm I developed created a complex situation where different types of error occasionally occurred (i.e. memory or tracking errors). Given delayed error feedback that indicated only generically that something went wrong, participants most often correctly identified the type of error they made. The results suggest that the system does detect and represent information about processing errors that occur in the course of performing a complex task.

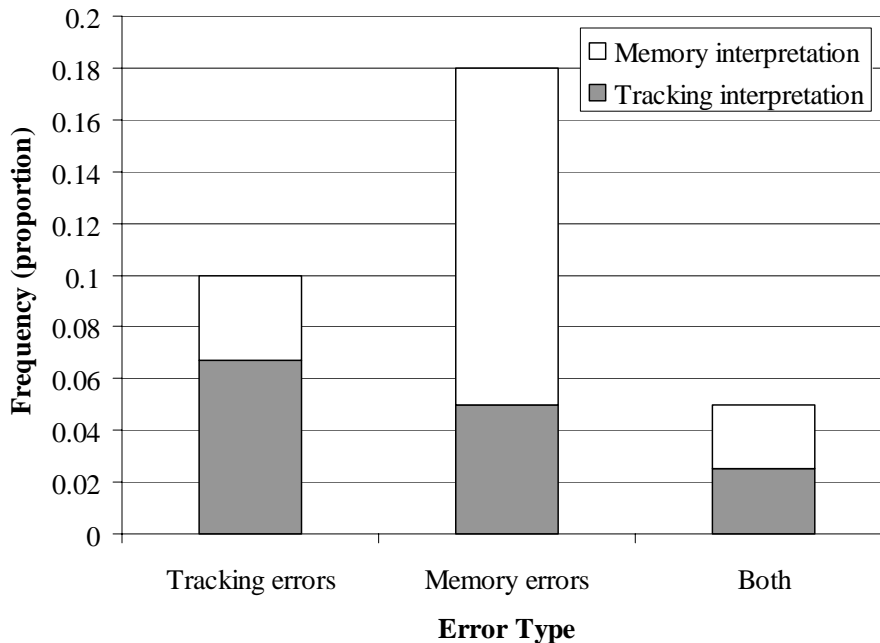


Figure 2. Frequency of different types of interpretations as a function of error type in Experiment 1.

2.2 Experiment 2: Imaging Study of Error Detection

In Experiment 2, I address two questions concerning the neural substrates of an error detection mechanism: The first question is whether the ACC represents information about an error type (i.e. memory versus tracking errors). In the behavioral experiment, I

determined that people do have access to the information regarding the type of error they have just committed—in more than 70% of cases they correctly identify the type of error that has just occurred. In the imaging study, I identify whether the ACC is the brain region that carries this information. The second question concerns participants' interpretation of the performance on a task—does the ACC represent participants' judgment regarding what went wrong? In the behavioral task, I found that, for example, some of the tracking errors are classified as tracking, whereas other tracking errors are classified as memory errors (the same is true for memory errors). In the imaging experiment, I explore whether the ACC codes participants' interpretation of an error regardless of an actual error type.

2.2.1 Method

Participants. Twenty-one right-handed, healthy undergraduate students from the Michigan State University Psychology Department subject pool participated in the experiment. All participants gave a written consent according to the study protocol that has been approved by the Institution Review Board at the Michigan State University. All participants had normal or corrected-to-normal vision. The participants were rewarded with either course credits or \$20 whichever they preferred.

Given that the study focused on errors and that participants had variable error rates (e.g. some participants did not produce enough error trials, others did not produce enough correct trials to fill the cell of the design), participants whose accuracy was higher than 60% or lower than 10% were excluded from the analysis. Seventeen participants out of twenty-one met this accuracy criterion (three participants were excluded because their

accuracy was higher than 60%, and one participant was excluded because his accuracy was lower than 10%).

Apparatus. Stimuli were presented on a rear projection screen, a 1024 x 768 32-inch LCD monitor (Salvacion Design, Sausalito, CA), located in the scanner bore. Participants viewed the screen via an angled mirror attached to the head coil. A viewable screen subtended a visual angle of 13° by 10°. Stimulus presentation and response collection were controlled by Experiment Builder software (SR Research Ltd., Mississauga, ON, CA).

Materials. In each trial, fourteen dots were displayed on the screen. The dots consisted of blue outline rings with light purple interiors and were displayed on a white background. The dots were opaque so that when two dots intersected each other, one was shown to be in front of the other. The interior of a dot was 0.786 degrees of visual angle; the outline ring was 0.026 degrees thick. Four of the dots formed a square, another four of the dots formed a diamond, three dots formed an upward pointing triangle, and other three dots formed a downward pointing triangle, see Figure 3. The dots in the square, the diamond, and the triangles were separated by approximately 1.96 degrees of visual angle. The location of the shapes – left, right, up, and down from the center of the display – varied randomly from trial to trial.

During the tracking phase of a trial, each dot moved with a velocity and direction that were changed at random. The velocities of the dots ranged from 2 to 3.5 degrees per second in the easy version of the task and from 2 to 4 degrees per second in the difficult version. The directions were chosen from among eight equal divisions of the compass. The rate of velocity and direction changes ranged from 500 to 1000 ms. When a dot

reached an edge of the screen, the x or y velocity vector reversed its value, so that the dot appeared to bounce off the edge. The dots' trajectories were generated in real time during each trial.

The accuracy feedback was indicated by a colored square (Figure 3, Feedback panel). For half of the participants, a red square meant "correct", and a green square meant "incorrect"; for the other half of the participants, the meaning of the color feedback was reversed. The square appeared in the center of the screen and subtended 0.886 degrees of visual angle.

Procedure. Participants were tested individually in sessions lasting about 150 min. At the start of the session, they read through a description of the task. The experimenter then presented three trials and demonstrated performance on a correct trial, a tracking error trial, and a memory error trial, describing what participants need to do during each phase of the task (e.g., make a judgment after negative feedback). Then participants were given two blocks of five practice trials, totaling 10 practice trials per participant. The practice trials were from an easy version of the task meaning that the rate of shape presentation and dots' velocity were lower. If participants made four or fewer errors during the practice block then they were presented with a difficult version of the task in the scanner. That allowed me to adjust the difficulty of the task for each participant individually. Seven out of seventeen participants performed an easy version of the task in the scanner.

The task I used for imaging was a modification of the task used in Experiment 1. The modifications were made with two aims: to increase the number of errors and decrease the likelihood of correctly guessing a shape during the memory retrieval phase

of the task. The modified task is presented in Figure 3. In this task, trials were presented in blocks of five trials each. Each block of five trials started with an encoding phase, during which five randomly chosen shapes – squares, diamonds, upward pointing triangles, downward pointing triangles – were presented sequentially (one shape was repeated to create a five-shape sequence). Each shape was displayed for 550 ms in the easy version of the task and for 500 ms in the difficult version of task. The participants' task was to remember the sequence. The sequence instructed participants what shape would be relevant on each of the five trials of the block as in the Experiment 1.

The beginning of each trial was marked with a fixation cross. Then the participant was shown a display consisting of fourteen dots: four dots were arranged in a square, four dots were arranged in a diamond, three dots were arranged in an upward pointing triangle, and three dots were arranged in a downward pointing triangle. This was the memory retrieval phase. The participant's task was to retrieve the shape that was relevant on that trial, and attend to the dots that formed that shape, as participants did in Experiment 1. After that, all dots began moving randomly on the screen. During the tracking phase, the participant's task was to keep their attention only on the relevant dots. After all dots stopped moving, the participant was asked to indicate with mouse clicks which dots they tracked. As the participant clicked on a dot, an "x" sign appeared, covering the selected dot. A color feedback then appeared, indicating whether the participant was accurate. After positive feedback, the trial ended. After negative feedback, the participant was asked to interpret the negative feedback and indicate what type of error they thought they had made. This was the judgment phase. The participants were asked to click with the mouse on the "Shape" button if they believed they had made

a memory retrieval error, and on the “Tracking” button if they believed they had made a tracking error. The locations of “Shape” and “Tracking” buttons were counterbalanced, such that for the half of the participants “Shape” button was displayed above “Tracking” button, and for the other half of the participants “Tracking” button was displayed above “Shape” button.

The length of all phases—Retrieval task, Tracking task, Response, Feedback, Judgment—was predetermined by the experimenter and varied to create jittering. The phases were randomly assigned to the following durations: Fixation cross—2.5 s, 5 s, or 7.5 s; Retrieval phase—2.5 s or 5 s; Tracking phase—5 s or 7.5 s; Response—7.5 s or 10 s; Feedback—2.5 s or 5 s; Judgment—always 5 s. Participants performed 18 blocks of five trials in the scanner, totaling 90 trials per participant.

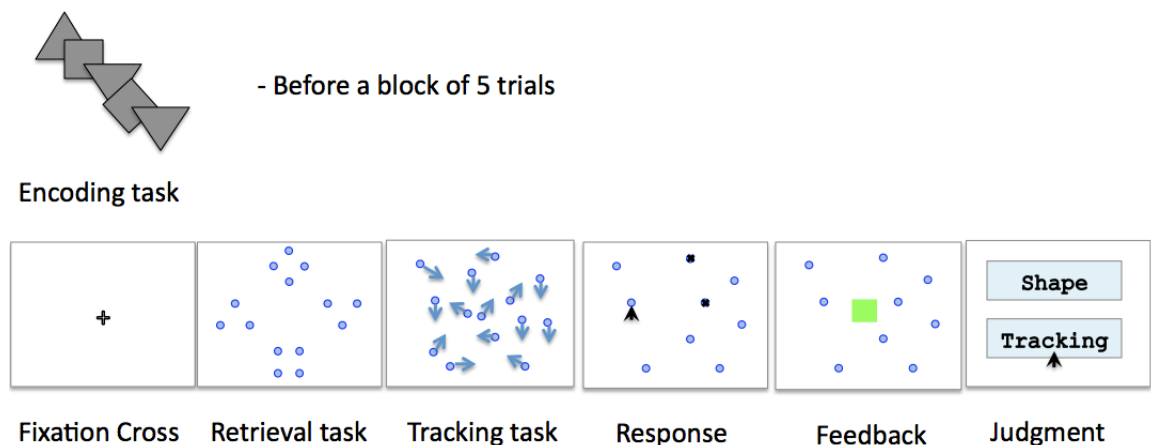


Figure 3. An example of a trial with a negative feedback in a modified task used for imaging study.

Image acquisition. Magnetic resonance imaging was conducted on a 3T Signa® HDx MR scanner (GE Healthcare, Waukesha, WI) with an eight-channel head coil. During each session, images were first acquired for the purpose of localization, and first and higher order shimming procedures were then carried out to improve magnetic field

homogeneity (Kim et al., 2002). T2*-weighted echo planar imaging was performed with the following parameters: 36 interleaved slices, slice thickness = 3 mm, TE = 27.7 ms, TR = 2,500 ms, flip angle = 80°, FOV = 22 cm, matrix size = 64 x 64, ramp sampling, the first four data points were discarded. Each volume of slices was acquired 206 times during each of the six functional runs, totaling 1236 volumes of images over the course of the entire experiment. After functional data acquisition, high-resolution volumetric T1-weighted fast spoiled gradient-recalled (FSPGR) images with cerebro-spinal fluid suppressed were obtained to cover the whole brain. 180 1-mm sagittal slices were acquired with flip angle of 8°, FOV = 256 mm, and matrix size of 256 x 256 with a voxel size of 1 mm x 1 mm x 1mm.

fMRI data pre-processing and analysis. All fMRI data pre-processing and analysis were conducted with AFNI software (Cox, 1996). For each subject, with the last volume of functional images as a reference, rigid-body motion correction was done in three translational and three rotational directions. The amount of motion in these directions was estimated and then the estimates were included as regressors in general linear model (GLM) during data analysis. For each subject, spatial blurring with a full width half maximum of 4 mm was applied to reduce random noise (Parrish et al., 2000), and also to reduce inter-subject anatomical variation and Talairach transformation variation during group analysis. For the group analysis, all images were converted to Talairach coordinate space (Talairach & Tournoux, 1988) with an interpolation to 1 mm³.

Individual subject fMRI analysis. For the data analysis of each individual subject, each phase of the task was modeled separately except Fixation cross displayed between trials in a block and the Memory retrieval task, which were included in the model

together as memory related regressors. Three regressors were included for the Memory retrieval phase: correct retrieval (Mcor), consistently interpreted memory error (Merr_m), and inconsistently interpreted memory error (Merr_t). Three regressors were included for the Tracking phase: correct tracking (Tcor), consistently interpreted tracking error (Terr_t), inconsistently interpreted tracking error (Terr_m). Six regressors were included for Feedback phase: feedback on correct trials (Fcor), feedback given Merr_m (FMer_m), feedback given Merr_t (FMer_t), feedback given Terr_t (FTer_t), feedback given Terr_m (FTer_m), and feedback give both types of errors (i.e., memory and tracking) were made (Fboth). One regressor for the Response phase, one regressor for the Judgment phase, one regressor for the Encoding phase were also included in the model. Therefore, 15 conditions were modeled, and the hemodynamic response (HDR) at each voxel with respect to each condition was resolved with multiple linear regression using the “3dDeconvolve” software in AFNI (Ward, 2000). The MRI signal model also included the subject motion estimations in three translational and three rotational directions, and the constant, linear, and quadratic trends for each of the six functional runs. The HDRs were resolved to seven points from zero to 15 s at the resolution of 2.5 s (TR). The BOLD signal change was calculated based on the area under the HDR curve. The equivalent BOLD percentage signal change relative to the baseline state was then calculated.

Whole-brain analysis for fMRI. After the percentage signal change was estimated for each condition for each participant, an ANOVA was performed over the 17-subject dataset for group analysis with mixed-effect two-factor model. The stimulus condition was the first factor and was modeled as a fixed effect. Subject was the second factor and

was modeled as a random effect. Monte Carlo simulation of the effect of matrix and voxel size of the imaging volume, spatial blurring, voxel intensity thresholding, masking, and cluster identification was used to estimate the overall statistical significance with respect to the whole brain (Ward, 2000). Based on these estimations, the ANOVA results above were further corrected for multiple comparisons based on the following criteria: The active voxel selection criteria required that the voxels had voxel-based $p \leq 1 \times 10^{-3}$ and were nearest-neighbor and within a cluster size of 177 mm^3 . Based on application of these criteria to the whole brain, the voxel-based $p \leq 1 \times 10^{-3}$ was corrected to be an equivalent whole-brain corrected $p \leq 0.027$.

2.2.2 Behavioral Results

In this experiment, two types of errors occurred: memory retrieval and tracking errors. The participants made errors on 63% of the trials. Tracking errors were more frequent than memory errors. Participants failed to track dots accurately on 30% of the trials (Figure 4, the left most bar). They forgot what shape they had to attend to on 16% of the trials (Figure 4, the middle bar). On 18% of the trials, participant made both types of error (Figure 4, the right most bar). First, they chose a wrong subset of dots to track during the retrieval phase and then they lost a dot or several dots during the tracking phase.

Following negative feedback, participants were asked to interpret the feedback and to make a judgment regarding what type of error they thought they had made. Two types of interpretations were possible: memory interpretations and tracking interpretations. The observed judgments suggest that people make informed decisions regarding what kind of error has just occurred. Figure 4 shows that 83% of tracking errors

were interpreted as tracking errors (grey shading of the left most bar), and 75% of memory errors were interpreted as memory errors (white shading of the middle bar). Not surprising, half of the trials, on which participants made both types of error, were interpreted as trials with memory errors and half of the trials were interpreted as trials with tracking errors (the right most bar). A Chi-square test indicates that the distribution of types interpretations differed from 50% x 50% for both memory ($\chi^2(1) = 25, p < 0.001$) and tracking errors ($\chi^2(1) = 43.56, p < 0.001$). These results suggest that participants did not merely guess about type of error they had made. Thus, I replicated here the results that were obtained in Experiment 1. The system **does** represent the information regarding error type, since participants most often identify errors correctly.

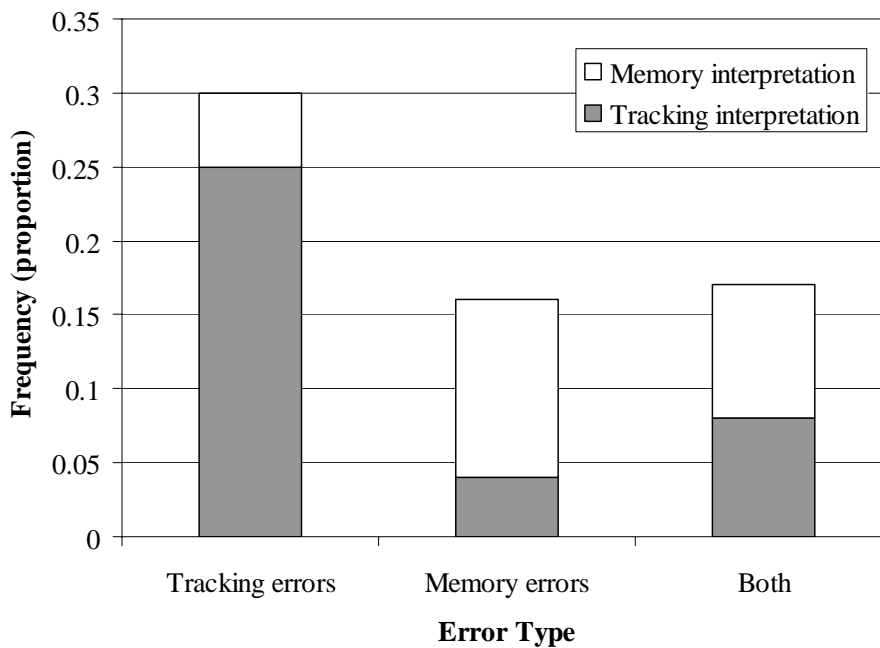


Figure 4. Frequency of different types of interpretations as a function of error type in Experiment 2.

2.2.3 fMRI Results

2.2.3.1 Whole brain analysis: all error versus correct trials during Task phase.

The tasks I used in the present study – memory retrieval and multiple object tracking tasks – are not commonly used in error detection research. Therefore, first, I wanted to determine if errors in these tasks activate dACC as they do in more traditional tasks such as the Eriksen, Stroop, or go/no go tasks. Contrasting activation on all error trials and correct trials (i.e. $(\text{Terr}_t + \text{Terr}_m + \text{Merr}_t + \text{Merr}_m) - 2(\text{Tcor} + \text{Mcor})$), I identified a single region in the left dACC, in BA 32, which is more active on error trials compared to correct trials, see Figure 5 and Table 1. The region with a significant effect extends from BA 32 dorsally into BA 8, which is consistent with previous studies of error detection and conflict monitoring (e.g., Brown, 2009, Carter et al., 1998). This finding suggests that errors in a memory retrieval task and a multiple object tracking task activate the dACC (the implications of this finding are considered in the General Discussion section).

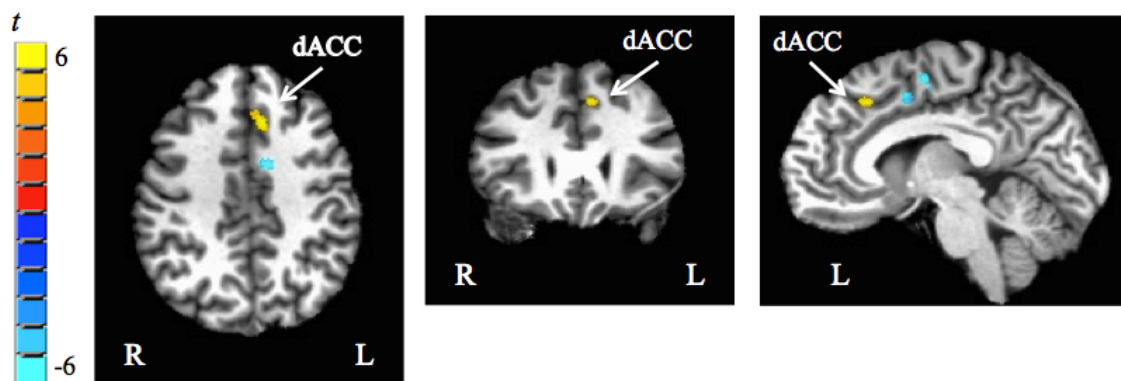


Figure 5. Error-related activation. Regions identified in a whole brain analysis contrasting all error trials and all correct trials $(\text{Terr}_t + \text{Terr}_m + \text{Merr}_t + \text{Merr}_m) - 2(\text{Tcor} + \text{Mcor})$. The arrow points to a region in left dorsal ACC, which is more active for error trials.

Table 1. Group analysis results for the contrast: “all errors > all correct”.

Cluster		L dACC
Active volume	mm3	255
Max <i>t</i> values		5.735
Max <i>t</i> value locations	RL, AP, IS	6,-24, 40
Merr_t		-0.03(0.07)
Merr_m		0.17(0.08)
Terr_t		0.59(0.10)
Terr_m		0.47(0.09)
Mcor		-0.08(0.10)
Tcor		0.23(0.06)

2.2.3.2 ROI analysis: dACC during Task phase.

Having identified a region in dACC responding to errors, I want to explore this region further and address the main questions of the present study: (1) Does the dACC respond differentially as a function of error type? and (2) Does the dACC respond differentially as a function of interpretation type?

To address these questions, I compared ACC activation during memory retrieval and tracking phases of a trial. I conducted a repeated-measures ANOVA (2 x 2) with Error Type (Tracking errors versus Memory errors) and Interpretation Type (Tracking interpretations versus Memory interpretations) as within subject factors, see Figure 6. The main effect of Error Type was significant, $F(1, 16) = 8.919, p = 0.009$. Thus, the left dACC region was more active on Tracking error trials compared to Memory error trials. The main effect of Interpretation Type was not significant, $F(1,16) = 0.295, p = 0.594$. Activation in the left dACC did not differentiate between types of interpretation. More importantly, the interaction of Error Type and Interpretation Type approached significance, $F(1,16) = 3.321, p = 0.087$. Thus, tracking errors that were interpreted as tracking errors (Terr_t) and memory errors that were interpreted as memory errors (Merr_m) activated the dACC more than tracking errors that were interpreted as memory errors (Terr_m) and memory errors that were interpreted as tracking errors (Merr_t). This result suggested an effect of *Consistency* of interpretations. Estimated hemodynamic response functions (HDR) are presented in Figure 7.

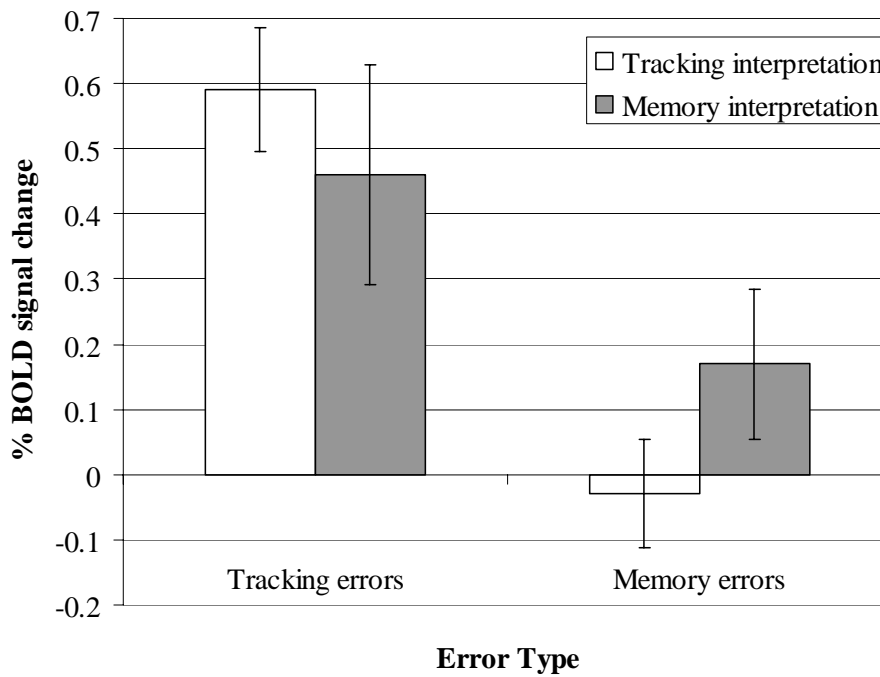


Figure 6. ROI analysis: the left dACC. The region shows an Error Type effect: tracking errors activate the region greater than memory errors, consistent with the idea that the ACC represents information regarding the types of errors people make. There is no effect of Interpretation Type. The interaction of Error Type and Interpretation Type approached significance, suggesting an effect of Consistency of interpretation: the ACC responds greater to consistently interpreted errors (Terr_t and Merr_m) compared to inconsistently interpreted errors (Terr_m and Merr_t). The error bars show SEM.

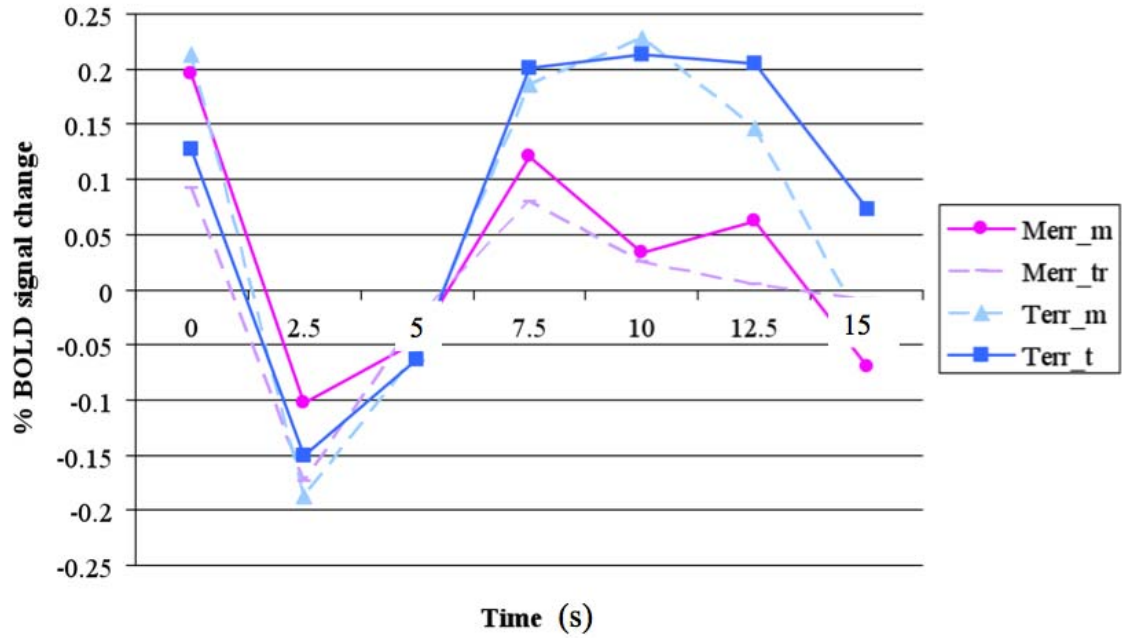


Figure 7. Estimated Hemodynamic Response Functions (HDR) for different types of error during the Task phase. The graph shows the effect of Error type – blue lines above red lines, and the effect of consistency – solid lines above dashed lines.

To explore the consistency effect further I re-represented the interpretation information as Consistent and Inconsistent instead of Tracking and Memory. Figure 8 shows the same data as Figure 6, but here, Terr_t and Merr_m are coded as consistently interpreted errors, and Terr_m and Merr_t are coded as inconsistently interpreted errors. To follow up the marginally significant Consistency effect, pair-wise two-tailed *t*-tests were conducted on Tracking and Memory errors separately (all *t*-tests presented in the dissertation are two-tailed). The Consistency effect was significant for Memory errors, $t(16) = 3.534, p = 0.003$, such that consistently interpreted Memory errors led to greater dACC activation compared to inconsistently interpreted Memory errors. The Consistency effect was not significant for Tracking errors, $t(16) = 0.779, p = 0.436$. After excluding one outlier (i.e. 2 SDs above the mean for T_inconsistent), who did show greater

activation for T_consistent compared to T_inconsistent, but had extremely high values in both conditions, which contributes quite a lot of variability, I found an approaching significance effect of Consistency, $t(15) = 1.916, p = 0.076$.

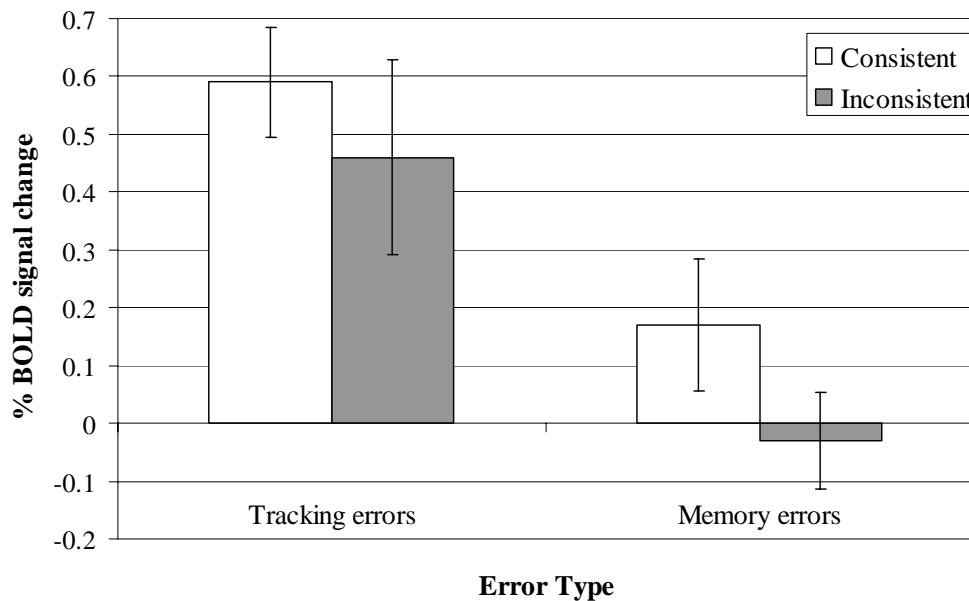


Figure 8. ROI analysis: the left dACC. The region shows Consistency effect: Errors interpreted consistently lead to a greater dACC activation than errors interpreted inconsistently. There is also an effect of Error Type. The interaction between Consistency of Interpretation and Error Type is not significant. The error bars show SEM.

A whole brain analysis supports the finding of the ROI analysis. The dACC shows greater activation for consistently interpreted errors compared to correct trials ($T_{err_t} + M_{err_m} - (T_{cor} + T_{cor})$), see Figure 9 and Table 2. At the same time, inconsistently interpreted errors do not show activity in this region. Consistently interpreted errors also activated the left medial frontal gyrus (MFG), and the left superior frontal gyrus (SFG). The analysis and the discussion of the activation in these regions are presented in Appendix A.

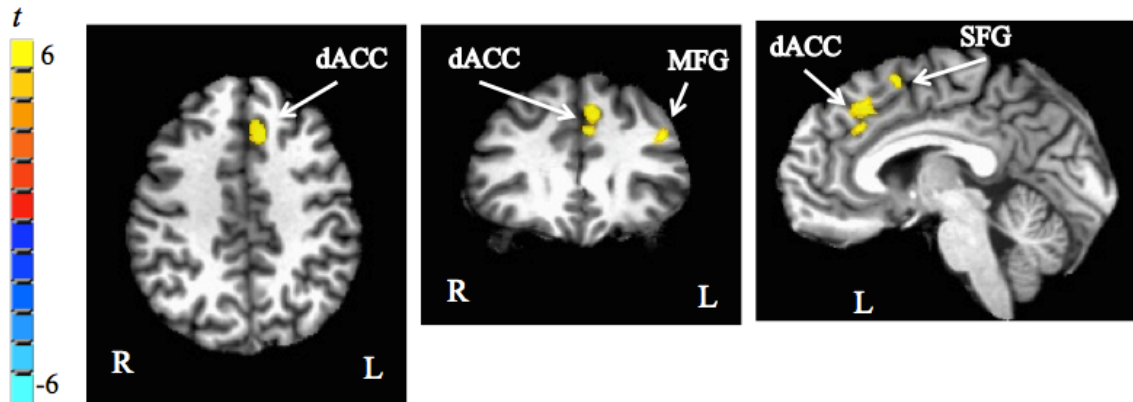


Figure 9. Consistently interpreted errors. Regions identified in a whole brain analysis contrasting consistently interpreted errors and correct trials ($Terr_t + Merr_m$) – ($Tcor + Mcor$). The arrows point to the regions more active for consistently interpreted errors. They are two regions in the left dACC (big and small), the left MFG, and the left SFG.

Table 2. Group analysis results for the contrast: “Consistently interpreted errors > correct”.

Cluster	Active volume	Max t values	Max t value locations				
	mm3		RL, AP, IS	Merr_m	Terr_t	Mcor	Tcor
L dACC (large)	625	7.334	6, -24, 41	.18(.08)	.54(.17)	-.03(.06)	.22(.09)
L dACC (small)	204	6.116	4, -29, 31	-.03(.05)	.30(.05)	-.14(.06)	.02(.08)
L & R SFG	274	7.123	-9, -13, 56	.04(.09)	.64(.18)	.02(.10)	.33(.15)
L MFG	283	6.359	40, -32, 30	.28(.07)	.39(.09)	.08(.05)	.10(.07)

Each trial in this study consisted of task performance (i.e., a memory retrieval task, and a tracking task) and feedback processing. So far I have considered dACC activation during task performance, the activation during a memory retrieval phase and a tracking phase. I found that dACC responded to errors that would be consistently interpreted later. It is reasonable to assume that consistently interpreted errors are the errors that people actually detect and/or are aware of, and that people are more likely to interpret an error inconsistently when they miss it. Making this assumption, I can conclude that dACC responds to errors that are actually detected as the system performs a task.

2.2.3.3 ROI analysis dACC: Task phase versus Feedback phase.

Next, I consider the activation in the left dACC during the feedback phase of a trial and compare it to the activation during task performance. Figure 10 shows the activation in the left dACC as function of consistency of interpretation and the phase of the task when activation is measured – during task performance (Task) or during feedback processing (Feedback). A repeated measures ANOVA (2 x 2) with Consistency (Consistent versus Inconsistent) and Time Point (Task versus Feedback) as within subject factors showed a significant interaction between factors, $F(1, 16) = 6.021, p = 0.026$. Neither of the main effects were significant, $p > 0.4$. To follow up the interaction, pairwise t-tests were conducted. As expected, during Task, consistently interpreted errors led to greater dACC activation compared to inconsistently interpreted errors, $t(16) = 4.001, p = 0.001$. However, during Feedback, the pattern was reversed such that the dACC was more active on inconsistently interpreted error trials than on consistently interpreted error trials, the effect was approaching significance, $t(16) = 1.703, p = 0.108$.

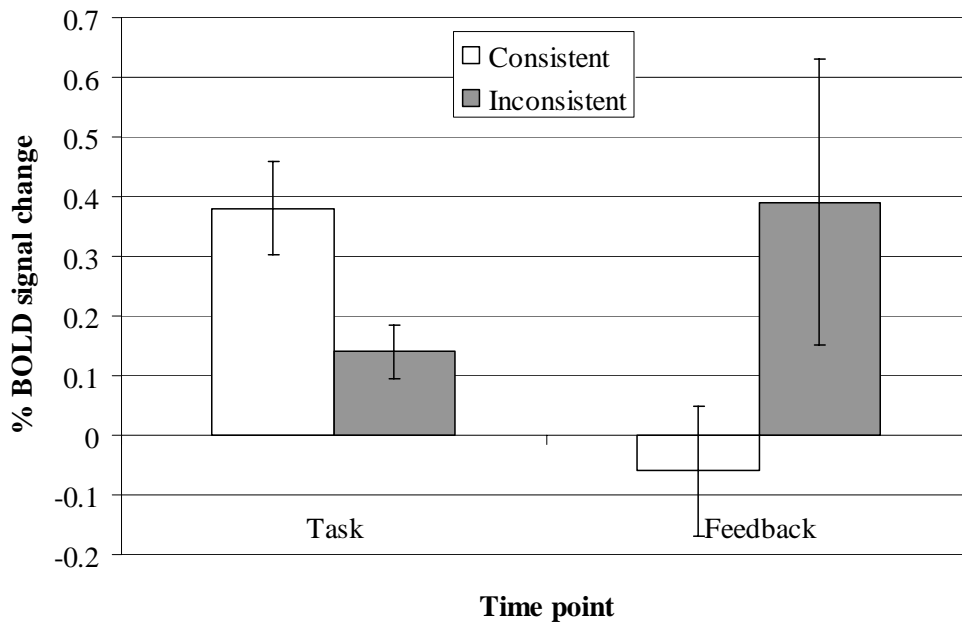


Figure 10. ROI analysis: the left dACC. The region shows Consistency by Time Point interaction: During task performance, errors interpreted consistently lead to a greater dACC activation than errors interpreted inconsistently. However, during feedback processing, errors interpreted inconsistently lead to greater dACC activation than errors interpreted consistently. Neither of the main effects is significant. The error bars show SEM.

To follow up the approaching significance consistency effect during feedback processing, I consider tracking and memory errors separately, see Figure 11. The consistency effect was not significant for memory errors, $t(16) = 0.687, p = 0.502$. Tracking errors showed a marginally significant consistency effect, $t(16) = 1.983, p = 0.065$, meaning that inconsistently interpreted tracking errors (FTer_m) activate the dACC region greater than consistently interpreted tracking errors (FTer_t). The lack of reliability in this test could be explained by high variance in FTer_m condition. The high variability in the FTer_m is probably due to the fact that the parameter was estimated based on a few trials (i.e. $M = 4.8$ $SD = 2.8$). To explore the effect of consistency I examined outliers. After excluding two outliers – one scoring 2 SDs above the mean,

another scoring 2 SDs below the mean on FTer_m – the marginal effect of consistency became significant, $t(14) = 2.561, p = 0.023$. I interpret this result as evidence for an effect of consistency during the Feedback phase. However, at the time of writing I am also collecting additional data to test whether the effect reaches conventional levels of significance. The HDR functions estimated based on the data from all 17 participants are shown in Figure 12.

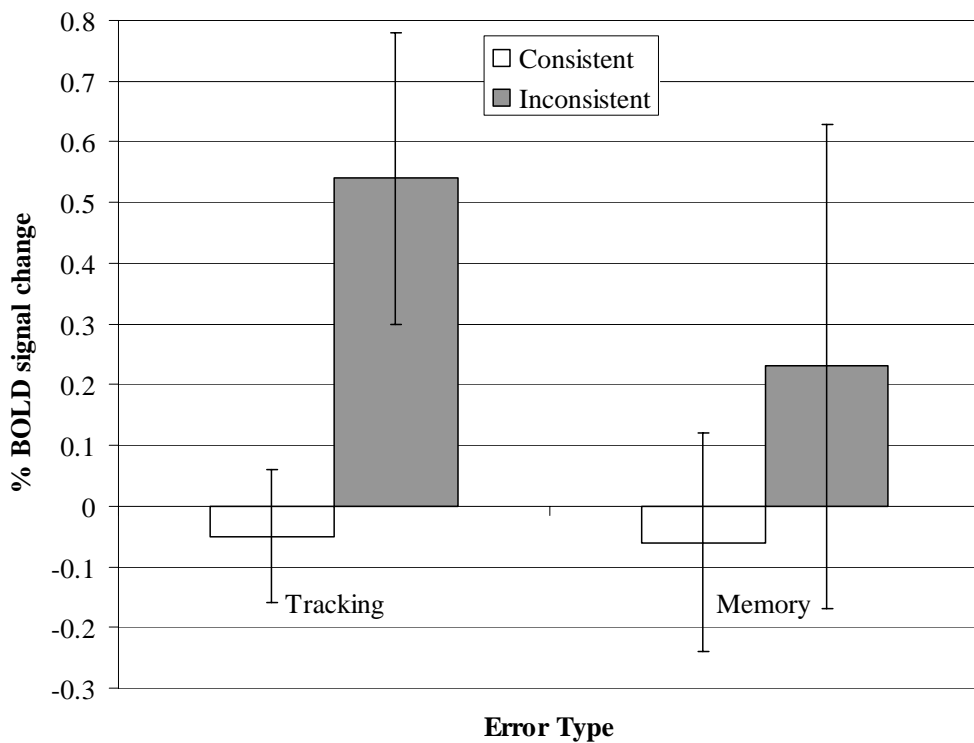


Figure 11. ROI analysis: the left dACC. The region shows Consistency effect during feedback processing only for tracking errors: Errors interpreted inconsistently lead to greater dACC activation than errors interpreted consistently. The effect of consistency is not significant for memory errors. The error bars show SEM.

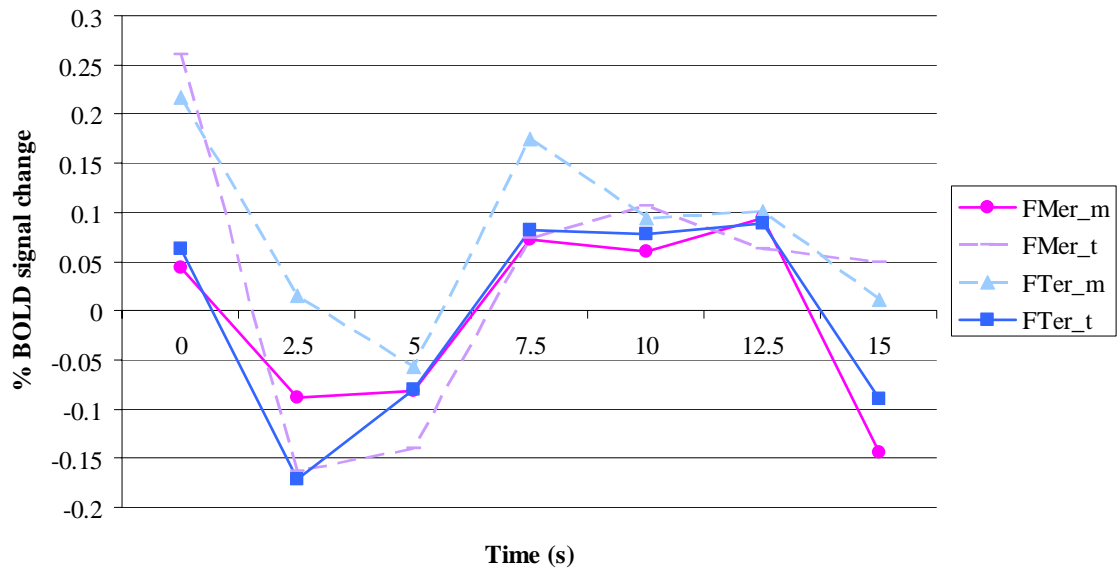


Figure 12. Estimated HDR functions for all types of error during the Feedback phase.

2.2.3.4 Whole-brain analysis: Interaction between Consistency and Time point (only for tracking errors).

A whole-brain analysis of the interaction between Consistency of interpretation (Consistent or Inconsistent) and Time point (Task or Feedback) for tracking errors revealed two areas – left MFG and left Caudate – that were more active during consistently interpreted errors compared to inconsistently interpreted errors during Task, but showed a reversed pattern during Feedback – the areas were more active during inconsistently interpreted errors compared to consistently interpreted errors during Feedback, see Figure 13 and Table 3. Activity in the dACC appears in the whole-brain analysis only at lower thresholds, $p = 0.003$ (voxel-wise uncorrected). Thus the whole-brain analysis does not seem to be sensitive enough to detect dACC activation when testing for the interaction of Consistency and Time point. Whole-brain analyses of the

interaction for memory errors and memory errors combined with tracking errors did not reveal any areas of activation.

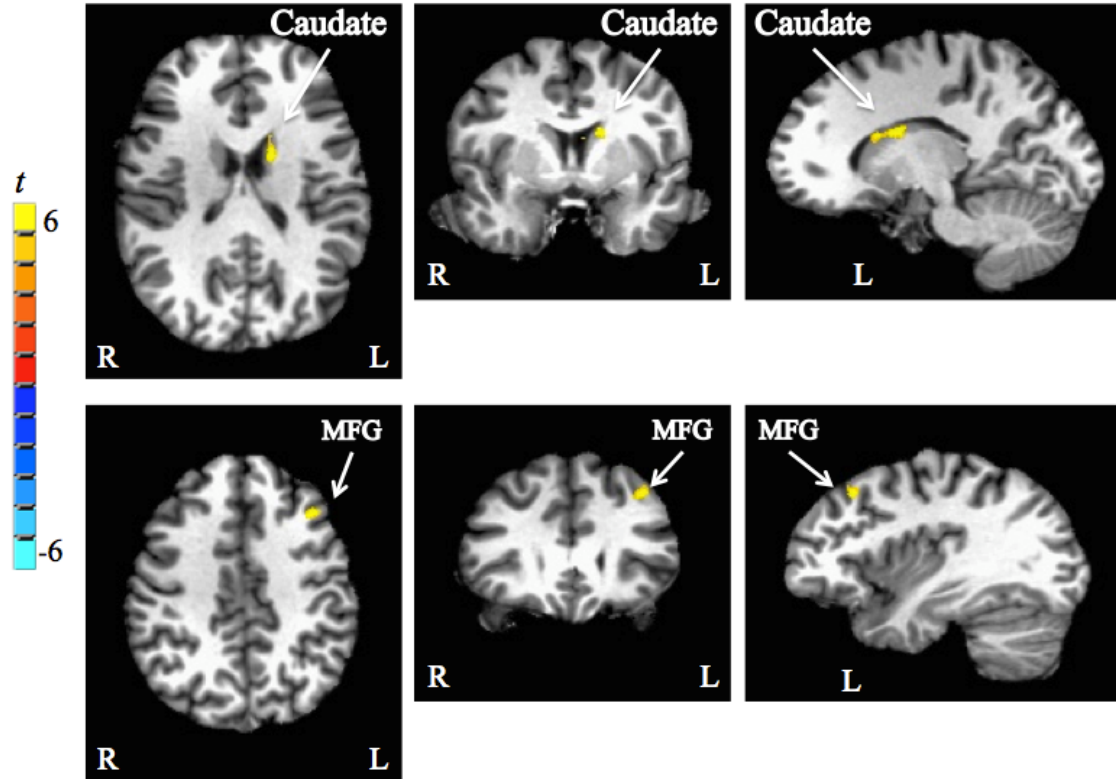


Figure 13. Consistency by Time point interaction. Regions identified in a whole-brain analysis contrasting (Terr_t + FTer_m) and (Terr_m + FTer_t). The arrows point to the regions more active for (Terr_t + FTer_m). They are the left MFG and the left Caudate.

Table 3. Group analysis results for the interaction between Consistency and Task Phase.

Cluster	Active volume	Max <i>t</i> values	Max <i>t</i> value locations				
	mm3		RL, AP, IS	Terr_t	Terr_m	FTer_t	FTer_m
L Caudate	343	7.237	14, -2, 21	.13(.07)	-.12(.06)	.04(.05)	.35(.13)
L MFG	188	6.146	36, -26, 40	.12(.06)	-.07(.11)	.01(.05)	.76(.31)

* * *

To summarize the results, consistently interpreted errors evoke dACC activity during the task performance (memory retrieval and tracking), but inconsistently interpreted errors evoke dACC activity during negative feedback. I assume that consistently interpreted errors are detected errors, and inconsistently interpreted errors are undetected errors. Under this assumption, the dACC seems to respond to errors when the system detects an error and also to negative feedback when such feedback is unexpected, i.e. when an error is not detected. This pattern of activation is consistent with the Reinforcement Learning theory of ACC function (Holroyd & Coles, 2002), and it is problematic to the Conflict monitoring theory (Carter et al., 1998). The present study confirms the Reinforcement Learning theory's assumption that the ACC is activated when an outcome is worse than expected: Both events, a detection of an error and negative feedback given after undetected errors, are worse-than-expected outcomes. The Conflict monitoring theory has difficulty explaining ACC activation after negative feedback as there is no conflict at that point. I will elaborate on these ideas in the General Discussion section.

3 GENERAL DISCUSSION

Several accounts of dACC function currently exist – the mismatch theory (e.g., Gehring et al., 1993), the conflict monitoring theory (e.g., Carter et al., 1998; Botvinick et al., 2001), the reinforcement learning theory (e.g., Holroyd & Coles, 2002), etc. The results of the present imaging study challenge the idea that the dACC performs a mismatch detection or conflict monitoring function, and they are more consistent with the idea that the dACC responds to worse-than-expected outcomes. The finding that the same region in the left dACC responds both to the errors that people are aware of and to the unexpected negative feedback is consistent with the reinforcement learning account and it is problematic for the conflict monitoring and mismatch detection theories.

In the following, I first discuss the finding that the left dACC responds to worse-than-expected outcomes. I explain why this is inconsistent with the conflict monitoring account. Then, I revisit Holroyd and colleagues' ERP study described in the introduction and explain how the imaging results of the present study and their ERP results converge. Next, I introduce an error-likelihood account of ACC function (Brown & Braver, 2005) and reinterpret the findings taken to support this account in the light of the present results. Then, I discuss the finding that the dACC responds differentially as function of error type and speculate about what the underlying mechanisms could be. Finally, I discuss the implications of the finding that the dACC responds to errors in tasks that do not involve response conflict.

3.1 The dACC in response to worse-than-expected outcomes

In the imaging experiment, I used a complex task, in which participants first performed two sub-tasks (a memory retrieval task and a multiple object tracking task) then processed feedback on the performance. This design allowed me to conduct a within subject comparison of dACC activation during task performance and during feedback processing. Another important feature of the task is participants' judgments after negative feedback. These judgments allowed me to distinguish between detected and undetected errors. On a single trial, participants performed two tasks – a memory retrieval task and a multiple object tracking task. The feedback participants received after performing these tasks was ambiguous, since it told them if they were accurate, but negative feedback did not indicate on which of two tasks they made an error. After negative feedback, participants were asked to make a judgment regarding the type of error. Most of the time they made correct judgments – when they failed to retrieve a correct item from memory they indicated that they made a memory error, and when they lost a track of a dot they indicated that they made a tracking error. However, sometimes they made incorrect judgments – when they made a memory error they indicated that it was a tracking error, and when they made a tracking error they indicated that it was a memory error. My assumption here is that participants are more likely to make a correct/consistent judgment when they actually detect an error, and they are more likely to make an incorrect/inconsistent judgment when they miss an error. Thus, the judgments allowed me to separate all errors into detected and undetected errors.

Exploring the activity in the left dACC region identified in “all errors versus correct” contrast I found an interaction between error detection and phase of the trial. The region was more active to detected errors than undetected errors during task performance

but during feedback the pattern was reversed such that the region was more active to negative feedback after undetected errors than to negative feedback after detected errors. This pattern of results is problematic for the conflict monitoring theory of ACC function, as there is no conflict during feedback processing. The results are, instead, consistent with the reinforcement learning theory of ACC function, according to which the ACC responds to worse-than-expected outcomes.

3.1.1 Evidence concerning the Conflict monitoring theory

The conflict monitoring theory can accommodate the fact that the ACC responds to detected errors during task performance. During memory retrieval task, there could be a conflict between multiple items competing for the retrieval similar to what has been observed in a verb generation task (Barch, Braver, Sabb, & Noll, 2000).

In a verb generation task, participants are presented with a noun and their task is to generate a verb associated with the noun. Some of the nouns have multiple verbs equally strongly associated with them. For example, the noun “ball” is approximately equally associated with multiple verbs, such as “hit”, “throw”, “bounce”, and “play”. In contrast, other nouns have a few or one strongly associated verb. For example, the noun “bell” is strongly associated with one verb -- “ring ” (Eighty three percent of subjects produced that verb in response to the noun in a norming study). Barch and colleagues (2000) found that nouns like “ball” lead to greater ACC activation compared to the nouns like “bell”, presumably because “ball” primes a number of verbs that compete for the retrieval. The finding that the dACC is more active in response to errors during memory retrieval can be explained in a similar vein. On memory error trials, there is no a single dominant item and multiple items create a competition that the ACC detects. On

undetected error trials, one of the items could erroneously gain greater activation compared to other items, which could lead to participants being more confident in accuracy of retrieval.

The conflict monitoring theory could also explain dACC activation during errors in the multiple object tracking task. People are more likely to lose a dot when it intersects with another dot. These intersections create competition between dots for attention. The intersections are probably more frequent on error trials compared with correct trials. On undetected error trials, the system perhaps did not engage a conflict resolution mechanism as much as it did on detected error trials. Therefore, the dACC responds more to detected errors compared to undetected errors.

Although the conflict monitoring theory can accommodate dACC activation to errors and especially to detected errors during performance on both tasks (i.e. memory retrieval and multiple object tracking tasks), it has difficulty explaining dACC activation in response to negative feedback. There are no competing responses or items during feedback processing. Previous research has suggested ACC involvement in negative feedback processing (e.g., Miltner et al., 1997; Holroyd & Coles, 2002; Gehring et al., 2002); however these studies employed ERPs, which are known to be imprecise in terms of spatial resolution. Some conflict monitoring theory proponents (e.g. Carter & van Veen, 2007) doubt that the ACC is the origin of feedback-related ERN, because the ACC involvement in feedback processing was not confirmed by imaging studies (van Veen et al., 2004; Nieuwenhuis et al., 2005).

In the fMRI studies (van Veen et al., 2004; Nieuwenhuis et al., 2005), participants performed a time estimation task similar to the one used in the original ERP study by

Miltner and colleagues (1997), which was described in the introduction. Recall that in this task, participants were asked to press a button one second after the onset of a trial, and feedback indicated whether the response was within an acceptable range.

Nieuwenhuis and colleagues did not find a single region in the medial frontal cortex responding more to negative feedback compared to positive feedback, and they concluded that the ACC is not the neural generator of the feedback-related negativity. In the present study, I did find that a region in the left dACC responds to negative feedback. The discrepancies between the results of our study and the previous fMRI studies of feedback-related responses in the ACC (e.g., Nieuwenhuis et al., 2005) likely involve distinguishing between expected and unexpected negative feedback in the current study. I found an ACC response only for the unexpected negative feedback after undetected errors. When I collapsed all negative feedback, including detected errors, and compared it to all positive feedback, I did not find a significant difference, $p > 0.05$, similar to previous work. Therefore, expectations regarding the accuracy of the performance rather than merely the negative valence of the feedback seem to be a key to the ACC engagement. In addition to collapsing between expected and unexpected negative feedback, Nieuwenhuis and colleagues employed a task that did not encourage the system to form any kind of performance expectations. In their task, the range of acceptable responses changed within a session to keep the participants' accuracy at 50% level. As participants were calibrating their 1 sec estimations throughout the session, the time window of correct response narrowed. So what was considered to be a correct estimation at the beginning of a session was penalized as an incorrect response at the end of the

session. This procedure could potentially discourage the system from forming any strong expectations regarding the accuracy of the performance.

3.1.2 Evidence concerning the Reinforcement learning theory

The theory that cannot only accommodate ACC activation during feedback processing but actually predicts the observed interaction between error detection and task phase is the reinforcement learning theory (Holroyd & Coles, 2002). In this theory, the ACC gets activated when the system detects that an outcome of an action is worse-than-predicted. When the error detection fails because the system does not have a criterion for what constitutes a correct action or due to fatigue or noise in the system, it waits until some external feedback provides an evaluation of performance. However, when error detection occurs because the system learned the criteria for correct actions, and the system functions properly then it does not need to wait for external feedback to issue an evaluative signal.

The results of the present imaging study map well onto the results of the ERP study by Holroyd and Coles (2002), which was discussed in the introduction. Recall that, in the ERP study, participants were asked to press a left or right button in response to a stimulus. The stimulus-response mapping was not given to participants but had to be learned throughout the experiment. One stimulus was mapped onto the left button, another stimulus was mapped onto the right button. That was the 100% mapping condition. Two stimuli were not mapped onto a particular button, and participants received negative feedback on half of the trials when the stimuli appeared – 50% mapping. The fifth stimulus was always penalized, and the sixth stimulus was always

rewarded. At the beginning of a session, all conditions elicited ERN at the point of feedback. However, as learning occurred, the ERN moved from feedback to the point of response but only for stimuli in the 100% mapping condition. Holroyd and Coles (2002) concluded that when the system learns a representation of a correct response it does not need to rely on external feedback to evaluate the performance. However, when such representations are not available as in 50% mapping condition, the system waits until the feedback.

Similarly, in the present study, I found that when the system is able to detect an error, it does not wait until the feedback to issue an error signal or a signal indicating that the performance outcome was worse-than-expected. However, when such detection fails (in this case not because it was impossible as in Holroyd and Coles' study, but due to some other factors such as fatigue, noise, etc.), the system does not issue the error signal until it processes external feedback. This is exactly the pattern of the results that would have been predicted by the reinforcement learning theory.

3.1.3 Evidence concerning the Error-likelihood theory

Recently, Brown and Braver (2005) proposed yet another theory of ACC function. In this view, the ACC represents a prediction of the likelihood of an error. I describe this study here, because it seems to provide converging evidence for the idea that expectations of a particular outcome influence ACC activation. First, I present the hypothesis of the theory regarding ACC function and the main empirical results taken to support it. Then I discuss a recent finding from this research group, which explored individual differences in forming expectations of performance outcomes (Brown & Braver, 2007). Finally, I try to reconcile and integrate the results of the present research,

the results of Brown and Braver studies (2005, 2007), and the results of Holroyd and Coles ERP study (2002) that was described earlier.

Brown and Braver (2005) observed that the ACC responds more on a high-error likelihood trial compared to a low-error likelihood trial even when there is no error or conflict on that trial. This result was taken to show that the ACC learns and stores error-likelihood information, rather than monitoring conflict. The authors employed a modified go/no go task, called a *change signal task*. In this task, participants were presented with a color cue at the beginning of a trial – blue corresponded to a high-error condition and white corresponded to a low –error condition. Then an arrow appeared pointing to the left or to the right. On 66% of the trials, the arrow indicated the required button-press – left or right index finger. These were *go* trials. However, on the 33% of the trials, a second arrow appeared signaling that the response had to be left-right reversed – if the first arrow pointed to the left, the second arrow would point to the right and the response should be made with the right index finger. These were *change* trials. The change trials were equally frequent in blue and white conditions. What made the blue condition harder or more error-prone was a longer interval between the first and the second arrow. The longer interval made it more difficult to cancel the prepared response since the second arrow appeared when the point of no return in response generation had already been reached.

Brown and Braver found that, as expected, the ACC responds more on error trials compared to correct trials – the error effect. They also found that the ACC responded more to correct-change trials compared to correct-go trials – the conflict effect. Importantly, they found that the ACC also responded more to high-error-likelihood

correct trials compared to low-error-likelihood correct trials. The effect was observed for both change and go trials. The finding that the ACC responds to high-error-likelihood trials even in the absence of any conflict (i.e., *go* trials) led to the conclusion that the ACC represents the information about error-likelihood. The authors also found that the pattern of ACC activation was reversed on error trials: the ACC responded greater to low-error-likelihood trials compared to high-error-likelihood trials. Brown and Braver claimed that the finding was consistent with their model, but did not provide detailed explanation apart from saying that “when error likelihood is lower, actual errors more highly suppress phasic dopamine cell activity.”

I believe that the results of Brown and Braver (2005) study show that the ACC responds to violations of expectations rather than that the ACC responds to high-error likelihood context. I believe that the cue at the beginning of a trial creates expectations regarding performance on that trial: in a high-error likelihood condition, people expect to make an error, and in low-error likelihood condition, people expect to perform correctly. In Brown and Braver’s study (2005), the ACC responded more to correct trials in the high-error likelihood condition than to correct trials in the low-error likelihood condition; and it responded more to error trials in the low-error likelihood condition than to errors in the high-error likelihood condition. These results are consistent with the idea that the ACC is activated when expectations regarding the outcome of the performance are violated regardless of the valence of the violation (i.e. better- or worse-than-expected).

In more recent studies, Brown and Braver (2007, 2008) investigated individual differences in learning error-likelihood information. They discovered an inverse relationship between ACC activation to error-likelihood (blue versus white cue

conditions) and ACC activation to conflict (change versus go trials). Risk-taking individuals in their study lacked an error-likelihood effect, but at the same time, they showed a greater conflict effect compared to controls. The authors concluded that risk-taking individuals were acquiring the error likelihood information at a lower rate than the control group. The results of the study are consistent with the idea that expectations of a particular performance outcome influence ACC activation. The risk-taking individuals did not learn to associate a cue with a particular error-likelihood condition; therefore, they did not show any error-likelihood effect. At the same time, if the risk-taking individuals expect positive outcomes on most of the trials then errors and conflicts would always be unexpected, which would lead to greater ACC activation on these trials.

3.1.4 Integrating the finding from ERP and neuroimaging studies

The results of all three studies, Holroyd and Coles (2002), Brown and Braver (2005), and the current study, are consistent with the idea that the ACC responds to unexpected performance outcomes. Holroyd and Coles (2002) and the present study show that the ACC responds to worse-than-expected outcomes. The Brown and Braver study (2005) shows that the ACC responds not only to worse-than-expected, but also to better-than-expected outcomes (although the authors did not interpret their results in this manner).

Figure 14 summarizes the finding from these three studies. The information about a negative outcome or the possibility of a negative outcome can become available at three time points: during a cue/stimulus presentation (future), during task performance or response (present), and during feedback (past). A red cross marks the point when the information about a negative outcome or an outcome that was different-than-expected is

presented. Describing this figure, I refer to the red cross as a worse-than-expected outcome or negative outcome for simplicity, but I believe that the same is true for better-than-expected outcomes. The upper line in the Figure 14 represents a scenario in which participants learn about negative outcome at the point of feedback. The middle line represents a scenario in which participants learn about negative outcome during task performance or as they issue a response. The lower line represents a scenario in which participants learn that the performance is likely to be worse-than-expected during a cue or a stimulus presentation. The red curves above each time point indicate the ACC response to an event in each of the scenarios: high or low.

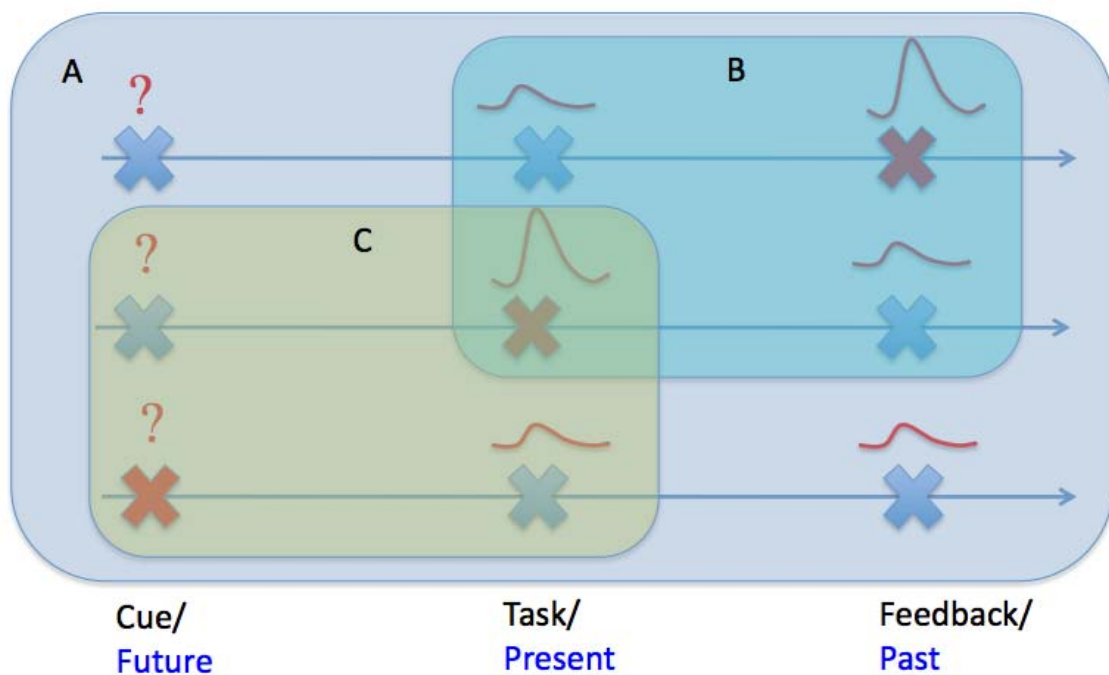


Figure 14. Summary of the results from three studies: Holroyd & Coles, 2002 (A), present study (B), and Brown & Braver, 2005 (C).

Holroyd and Coles (2002) investigated all three scenarios, see Figure 14 the grey box A. The upper line represents their 50% mapping condition, in which participant had to wait until feedback to learn about the outcome of their actions. The middle line

represents a 100% mapping condition after learning has occurred – ERN was observed at the response phase. The lower line represents their “always incorrect” condition, in which stimulus itself was informative in terms of the outcome of the performance.

The yellow box C in Figure 14 shows the scenarios investigated in the Brown and Braver (2005) study. The middle line represents errors after receiving a low-error likelihood cue, and correct trials after receiving a high-error likelihood cue; in both cases, outcomes of performance (positive and negative) are inconsistent with developed expectations. The lower line represents errors after receiving a high-error likelihood cue and correct trials after receiving a low-error likelihood cue; in both cases, the outcomes are consistent with developed expectations.

The blue box B in Figure 14 represents the scenarios considered in the current study. The upper line represents undetected errors (or inconsistently interpreted errors); the system learns about the negative outcome of the performance only at feedback. The middle line represents detected errors (or consistently interpreted errors); the system registers that the performance is worse-than-expected during task performance.

The results of the three studies seem to be consistent with the idea that the ACC responds to **unexpected performance outcomes**. The system can learn about the outcome of the performance at any point in the trial: before the task, during the task, and after the task. Importantly, the results depicted in Figure 14 are obtained with two different methodologies – ERP and fMRI, which allows us to conclude that the signal obtained during task performance and feedback processing has a common source – the ACC. The question of whether the ACC responds during cue processing remains to be addressed.

3.2 Content-specific signal in the ACC

The imaging results showed that an error related signal within the ACC is content specific. Previous studies have demonstrated that the ACC represents information about the occurrence of an error (e.g., Carter et al., 1998). The present results extend our knowledge regarding the ACC involvement in error monitoring by showing that the ACC represents information about the type of error that participants make, but not the type of error they think they made (i.e., type of interpretation).

Comparing the trials on which participants failed to retrieve correct information with the trials on which participants lost track of dots, I found an area within the left dACC that is differentially activated to the memory retrieval and tracking failures. This area was more active when participants lost track of a dot than when they failed to retrieve a correct item from memory. This differential response was observed during task performance but not during feedback processing. Comparing the trials on which participants interpreted errors as tracking and the trials on which participants interpreted errors as memory failures, I did not find an area within the left ACC that responds differentially to the memory and tracking interpretations.

The differences in the activation observed for memory and tracking errors could also be due to the difference in visual stimulation during memory and tracking phases. In the analysis reported above, in section *ROI analysis: dACC during Task phase*, I compared activation during memory and tracking errors without subtracting the activation during correct trials. Thus, the observed differences in activation could potentially be caused by differences in task demands. To test this possibility I compared

the activation during correct memory and correct tracking phases, see Appendix B. The analysis revealed a marginally significant effect of Task phase – the tracking phase seemed to result in greater ACC activation compared to the memory retrieval phase even when no mistakes were made. These results suggest that differences in the task demands (e.g., visual stimulation) could at least partially be responsible for differences in activation between memory and tracking errors. To test whether there are also differences in error-related activation above and beyond the differences on correct memory and correct tracking phases, I would have to subtract correct trials from error trials, $((\text{Terr}_t + \text{Terr}_m) - 2 \times \text{Tcor}) - (\text{Merr}_m + \text{Merr}_t) - 2 \times \text{Mcor}$. Unfortunately, that would have been a case of double dipping (e.g., Kriegeskorte, Simmons, Bellgowan, & Baker, 2010), because I identified the region in the left dACC by comparing all errors versus correct trials. One way of overcoming this problem would be to use tasks that are better matched in terms of their demands on visual system.

Assuming that the left dACC does differentiate between types of error, I see various explanations to this finding. One possibility is that the ACC contains neurons tuned to a particular type of error, similar to the feature detectors in visual cortex that respond to bars of light with a particular orientation. The neurons in the ACC could hypothetically respond to different types of error. These neurons could have different anatomy and physiology, or could differ in terms of their connections with the rest of the brain. A single cell recording study would be needed to address the question regarding the existence of various subsets of neurons tuned to different error types.

Another possibility is that people have differential sensitivity to the performance of different subsystems in the mind. For example, they could be more likely to detect a

tracking error and miss a memory error. A lower detection rate of memory errors could hypothetically lead to lower activation of the ACC region when we are looking at the responses averaged across all error trials. Behavioral studies have the potential to address this question. For example, participants could be presented with a task similar to one used in the present study with one difference – they would not be given feedback. In the present study, when participants received negative feedback, they could potentially employ a strategy, such as “if I did not make a tracking error, but I get negative feedback then it must be a memory error.” In this case, detection of a memory error during a retrieval attempt is not necessary for correctly identifying a trial as a memory error trial. Without feedback participants could no longer rely on strategies or heuristics to infer the accuracy of their performance, but they would have to rely on an error signal generated during task performance. This paradigm would allow the comparison of an error detection rate for different types of errors.

Finally, the greater ACC response to tracking errors compared with memory errors could be explained by the differences in expectations regarding the accuracy of the performance on the two tasks. The memory task could be too difficult for participants to form strong expectations of accurate performance, whereas the tracking task could be perceived as more manageable. So when participants fail to retrieve a correct item their expectations are not violated as much as when they lose a track of a dot. This hypothesis could be tested by parametrically varying the strength of expectations of positive outcomes. If a difficult version of a tracking task leads to lower ACC activation in response to an error compared to an easy version of a tracking task then it would suggest that ACC activation varies with the strength of expectations.

3.3 ACC activation in tasks without response conflict

Most of the research on error detection and conflict monitoring employs tasks that involve response conflict such as the Eriksen task, the Stroop task, the go/no go task, etc. In these tasks, errors or conflicts arise from stimuli activating several responses simultaneously (i.e. response conflict). Research addressing the question of whether the ACC responds to conflict at all levels of processing or only to conflict at the response phase has produced mixed results. Some studies suggest that the ACC is activated only during response conflict (e.g., Milham, Banich, Webb, Barad, Cohen, Wszalek, & Kramer; 2001), whereas other studies suggest that the ACC responds to conflict at semantic level as well (e.g., Barch et al., 2000). The findings of the present study support the latter idea, that the ACC can be activated by errors that do not involve response conflict. I find that errors in memory retrieval and in the multiple object tracking tasks activate a region in the left dACC. There is no response conflict during tracking or memory retrieval and yet ACC activation is greater when performance fails on these tasks. In the context of the finding that the ACC responds to worse-than-expected outcomes, the lack of ACC activation during non-response conflict trials (i.e. Milham et al., 2001) could suggest that the system did not register the event as worse-than-expected, or the non-response conflict violated the system's expectations to a lesser degree than response conflict.

4 FUTURE DIRECTIONS

4.1 Does the ACC respond to better-than-expected outcomes?

One of the most interesting findings of the present research is an interaction between error detection and the phase of the trial. I found that the ACC responds to errors when they are detected and to negative feedback when it is unexpected. I concluded that the ACC responds to worse-than-expected outcomes. In a follow up study, I plan to investigate whether the ACC responds to outcomes that are worse-than-expected or to unexpected outcomes regardless of the valence of that outcome— negative or positive.

In this experiment, I will present participants with only a multiple-object tracking task. After the tracking phase of the trial, I would present feedback on performance negative or positive. After feedback, participants will be asked to make a judgment of whether the feedback was congruent with their evaluations of the performance. There will be four types of trials: (1) expected negative feedback – when participants correctly detect a tracking failure; (2) unexpected negative feedback – when participants fail to detect a tracking error; (3) expected positive feedback – when participants do not detect an error and there is no error; and (4) unexpected positive feedback – when participants believe they made an error, but they receive a positive feedback.

First, I expect to replicate the results of the current study and find that unexpected negative feedback evokes greater ACC activity compared with expected negative feedback. More importantly, I am interested in examining the contrast between expected positive feedback and unexpected positive feedback. The finding that the ACC is more active during the processing of unexpected positive feedback than during the processing of expected positive feedback would suggest that the ACC responds to violations of

expectations of performance outcome regardless of the valence of the expectations – worse- or better-than-expected.

4.2 Violations of expectations in the context of anxiety.

Detection of an error is accompanied with an emotional reaction. People vary in terms of the intensity and duration of the emotion they experience in relations to an error commission: anxiety prone people are known to overreact to an error and experience a higher level of frustration compared to their non-anxious counterparts. Neuroimaging studies show a heightened level of activation in the ACC and greater ERN in people scoring high on anxiety questionnaires (e.g. Fitzgerald, Welsh, Gehring, Abelson, Himle, Liberzon, & Taylor, 2005; Hajcak, McDonald, & Simons, 2003). In future studies, I plan to investigate whether anxiety-prone people overreact to negative outcomes of their actions or to unexpected outcomes regardless of the valence of the performance (better- or worse-than-predicted).

In this study, participants first will fill out the STAI-trait and state inventory, Beck anxiety scale, and BIS/BAS questionnaires. The participants will be assigned to high or low anxiety group based on the results of these questionnaires. Then each subject will participate in one fMRI session. During this session, they will perform the task described in the previous section, “Does the ACC respond to better-than-expected outcomes?” If people from the high anxiety group show greater ACC activity only to unexpected negative feedback then this finding would confirm the idea that cognitive processing in the context of anxiety is biased toward negative outcomes. However, if people from high anxiety group show greater ACC activity in both the unexpected negative and unexpected positive feedback conditions, then it would suggest that the

system is not biased toward negative outcomes, but instead is more generally sensitive to violations of expected outcomes. Also, it would be interesting to explore if there are subgroups within the high anxiety group: those who show heightened activation to unexpected negative feedback only, and those who show heightened activation to both unexpected feedback – negative and positive. Thus, people who would commonly be diagnosed with anxiety disorder could in principle suffer from failures in different parts of the system, and may require differential treatments.

4.3 Negative emotions and rate of learning.

I would also like to explore the idea that emotions influence the rate with which the system learns error likelihood information. Detection of an error is usually accompanied with a negative emotional reaction. The functional significance of this reaction is poorly understood. Recent findings (e.g., Fitzgerald, et al., 2005) show that heightened emotional reactions to errors are associated with greater activities in the brain areas implicated in error-detection, pain, and reward processing – ACC, insula, basal ganglia. However, no clear behavioral consequences of this increased activity and associated emotions have been established yet (e.g., Hajcak et al., 2003). In future work, I plan to explore the effect of emotions on learning subsequent to errors, in particular, learning of error-likelihood information.

From studies on decision making, we know that negative affect can be beneficial in learning outcome probabilities (Mueller, Nguyen, Ray, & Borkovec, 2010). In a recent study, Mueller and colleagues (2010) showed that people scoring high on anxiety questionnaires perform better than those scoring low on anxiety questionnaires on the Iowa gambling task. In this task, people are presented with four decks of cards. Drawing

from a deck of cards can result in a particular combination of gains and losses. The decks differ in the net gains/losses. With time people learn to draw from advantageous decks and avoid drawing from disadvantageous decks. People scoring high on anxiety were shown to learn the contingencies faster than people scoring low on anxiety (Mueller et al., 2010). This enhancement may be attributed to stronger emotional signals after errors in anxious individuals.

In a recent study, Brown and Braver (2008) investigated individual differences in learning error-likelihood information. They showed that the amplitude of ACC activity depended on the rate of learning error likelihood information. The authors found that risk-taking individuals did not show the error-likelihood effect (blue versus white cue) in ACC activation when compared to controls, but they did show a greater conflict effect (change versus go trials). Brown and Braver hypothesized that risk-taking individuals were acquiring the error likelihood information at a lower rate than the control group. This assumption was confirmed by a simulation, in which a computational model behaved similarly to risk-taking individuals when the rate of learning was lowered (Brown & Braver, 2008).

Therefore, in future work, I plan to address the question of whether high anxiety is associated with faster learning rates and greater error-likelihood effects in the ACC. In addition, I would also like to test whether an inverse relationship between error-likelihood and conflict effects would be observed in high anxiety individuals, who are traditionally characterized as risk-avoiding.

4.4 Lingering effect of emotions

Emotions are not just generated in response to a cognitive process (e.g., error-detection), but some components of emotions may last for some time after the cause is gone. The question I plan to address in the future is whether negative emotions generated on a current trial can spillover onto a subsequent trial and influence its evaluation.

To address this question, I plan to modify the change signal task (Brown and Braver, 2005). In this modified version, there will be three conditions – one high-error likelihood condition (blue cue) and two equally low-error likelihood conditions (yellow and white cues). In the original task employed by Brown and Braver (2005), conditions were equiprobable and randomly intermixed. In the modified version of the task, one low-error likelihood condition (e.g., yellow cue) will follow high-error likelihood condition more often than another low-error likelihood condition (e.g., white cue). Brain activity will be recorded as participants perform the task and learn how error-prone each condition is. At the end of the session, participants will be asked to rate the conditions in terms of their difficulty/number of errors they produced.

As in the original study by Brown and Braver (2005), I expect to find greater ACC activity in high-error likelihood condition compared to low-error likelihood conditions (blue > yellow or white). Importantly, out of two equally low error likelihood conditions, the one that follows high-error likelihood condition more often is expected to be associated with greater ACC activity than the one that rarely follows the high-error likelihood condition (yellow > white). These results would suggest that error-generated emotions linger over time and influence the recording of error-likelihood information associated with the subsequent performance. Participants' rating at the end of a session

will suggest if subjects are aware of the error-likelihood information they learn (as reflected in their brain activity). The lingering effect would inform models of error-detection. If found it would suggest that detection of an error/conflict per se is not necessary for learning of error-likelihood information. The affective reaction generated in response to an error is sufficient for such learning to occur.

Negative affect can potentially not only spillover to a subsequent action, but it could also travel up a hierarchy of actions. The lingering effect of error-generated emotions could be a mechanism by which the system reinforces an action at a more abstract level. For example, spilling sugar on the floor, while trying to add it to a cup of coffee, can potentially negatively reinforce not only “adding sugar” (an elementary action) but the whole “coffee-making” procedure (an action at a higher level of abstraction). Follow-up experiments addressing these questions would have important implication for theories of action control in the normal population (e.g. Cooper & Shallice, 2000; Botvinick & Plaut, 2004) and in patients suffering from action disorganization syndrome and apraxia (e.g., Humphreys & Forde, 1999; Schwartz, Reed, Montgomery, Palmer, & Mayer, 1991).

4.5 Lingering effect of emotions in a context of anxiety

In this study, I would like to explore the individual differences in the lingering effect of error-generated emotions. In particular, I would like to address the question of whether people scoring high on negative affect (e.g. anxiety) would produce a greater lingering effect or a more lasting effect compared to people scoring low on anxiety.

First, participants will fill out the STAI-trait and state inventory, Beck anxiety scale, and BIS/BAS questionnaires. They will be assigned to high or low anxiety group

based on the results of these questionnaires. Then each subject will participate in one fMRI session. During this session, they will perform the modified version of change detection task described in the previous session “*Lingering effect of emotions*”. Anxiety prone people tend to overreact to errors and experience more intense and prolong negative emotions. Findings that the lingering effect can be modulated by the level of anxiety (train or induced) would confirm the assumption that error-generated emotions are the primary source of the learning error-likelihood information.

5 CONCLUSIONS

The results of the present work challenge the idea that the dACC performs conflict monitoring, and they are more consistent with the idea that the dACC responds to unexpected performance outcomes. In an fMRI study, I found that the same region in the left dACC responds both to detected errors and to unexpected negative feedback.

Although the conflict monitoring theory can accommodate the finding that errors in a memory retrieval task and in a tracking task evoke ACC activity, it is unclear how the theory would explain the finding that the ACC responds to unexpected negative feedback since there is no conflict during feedback processing. The results of the present study are consistent with the reinforcement learning account, since both detection of errors and processing of unexpected negative feedback are worse-than-expected outcomes. Future studies are needed to address the question of whether the ACC responds to worse-than-expected outcomes, or whether it responds to unexpected performance outcomes regardless of valence – positive or negative.

The imaging results also showed that an error related signal within the left dACC is content specific. Tracking errors led to greater ACC activity compared to retrieval errors. Several explanations are possible and future work is needed to distinguish between them. One possibility is that the ACC contain neurons that are tuned to a particular type of error. Alternatively, participants might have differential expectations regarding the accuracy of their performance on two tasks. The errors on the task that has greater expectations associated with it result in greater ACC activation.

APPENDIX A

Analysis of regions other than the left dACC

Although the contrast “all errors versus all correct trials” revealed only one region, the left dACC, to be active more for error trials, other important contrasts showed multiple brain regions to be involved in processing of errors and unexpected feedback. Here I address the question whether these areas perform a function similar to the left dACC. In particular, whether they exhibit the same pattern of activation as the left dACC does – responding to consistently interpreted errors during the Task phase and to inconsistently interpreted errors during the Feedback phase.

Consistently interpreted errors versus correct trials

Contrasting consistently interpreted errors and correct trials ($Terr_t + Merr_m$) – ($Tcor + Mcor$), I found a greater activation for consistently interpreted errors ($Terr_t + Merr_m$) in the following regions: the left dACC (two adjacent regions: a small inferior region and a large superior region), the left medial frontal gyrus (MFG), and the left superior frontal gyrus (SFG), see Figure 9. Inconsistently interpreted errors do not show activity in these regions when compared to correct trials in the whole brain analysis.

In the ROI analysis, I explored activation in these regions during feedback processing. I was interested to see if the regions would show a pattern of activation similar to what I found for the left dACC in “all errors versus all correct trials” contrast, that is if they will be more active on inconsistently identified errors compared to consistently identified errors during feedback processing. I expected to see a large superior region in the left dACC showing this pattern, as it encompasses the left dACC identified in “all errors vs. correct trials” contrast. I did not have predictions for other

regions as they might perform a function different from the left dACC and show consistency effect only during error detection but not during negative feedback processing.

ROI analysis: a large region in the left dACC.

Figure A1 shows percentage signal change in the large superior dACC region as a function of consistency of interpretation and error type. A repeated-measures ANOVA with Consistency (consistent versus inconsistent) and Error type (tracking versus memory) factors revealed a marginally significance main effect of consistency, $F(1, 16) = 3.317, p = 0.087$. The main effect of Error type and the interaction were not significant, $p > 0.7$. To explore the marginally significant effect of Consistency further, pair-wise t -tests were conducted on tracking and memory errors separately. The effect of Consistency was not significant for memory errors, $p = 0.398$. The effect of Consistency was small but significant for tracking errors in a one-tailed t -test, $t(16) = 1.996, p = 0.032$. After excluding two outliers (2 SDs from the mean in FTer_m condition), I find the effect of Consistency to be significant even in a two-tailed t -test, $t(14) = 2.592, p = 0.021$. Thus, the results show that the activation pattern in the large region of the left dACC in “consistently interpreted errors vs. correct trials” is similar to the pattern of activation observed in the left dACC in “all errors vs. correct trials”.

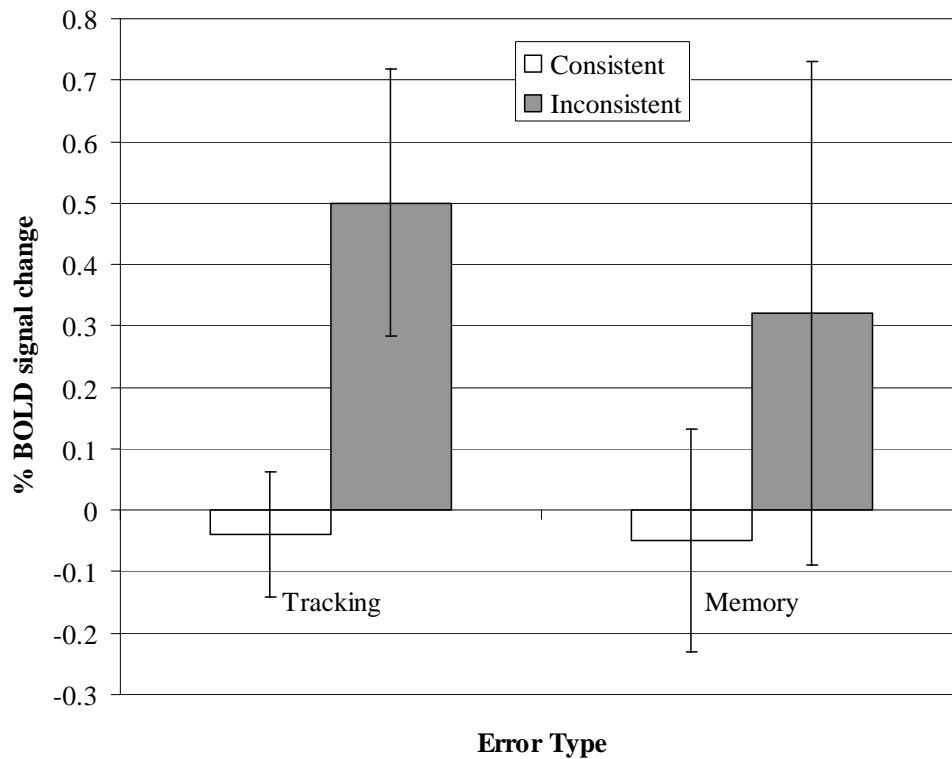


Figure A1. ROI analysis: a large region of the left dACC. During negative feedback processing, the region shows Consistency effect for tracking errors, but not for memory errors. The error bars show SEM.

ROI analysis: a small region in the left dACC, the left SFG, and the left MFG.

In the ROI analysis, I explored activation in a small region of the dACC, the left SFG, and the left MFG during feedback processing, see Table A1. Repeated-measures ANOVAs with Consistency (consistent versus inconsistent) and Error Type (tracking vs. memory) as within subject factors showed neither of the main effects nor the interaction of the factors was significant, $p > 0.05$. In addition, pair-wise t-tests on tracking and memory errors did not show consistency effects either, $p > 0.05$. Thus, the small region in the left dACC, the left SFG, and the left MFG did not exhibit the pattern of activation observed in the left dACC in the “all errors vs. correct trials” contrast. Inconsistently

interpreted errors did not lead to greater activation compared to consistently interpreted errors.

Table A1. ROI analysis: Percentage BOLD signal change observed during feedback processing for consistently and inconsistently interpreted tracking and memory errors in a small region in the dACC (a) and in the left SFG (b). Repeated measures ANOVAs showed neither the main effects (Consistency or Error Type) nor the interaction between factors.

	Tracking Errors		Memory Errors	
	Consistent Mean (SD)	Inconsistent Mean (SD)	Consistent Mean (SD)	Inconsistent Mean (SD)
small L dACC	-0.19 (0.38)	0.07 (1.11)	-0.10 (0.74)	-0.43 (0.94)
L SFG	-0.02 (0.56)	-0.18 (0.72)	-0.18 (0.83)	-0.47 (1.40)
L MFG	0.04 (0.51)	0.22 (0.90)	-0.15 (0.58)	0.23 (1.91)

To summarize, contrasting consistently interpreted errors with correct trials, I identified a number of regions more active for consistently interpreted errors. They are two regions in the left dACC (big and small), a region in the left SFG, and a region in the left MFG. ROI analysis in these areas showed that only the big region in the left dACC exhibited the pattern of activation similar to the pattern found in the left dACC from the “all errors versus all correct trials” contrast. The other regions – a small inferior region in the left dACC, the left SFG, and the left MFG – although active for consistently interpreted errors during task performance, do not show a consistency effect during negative feedback processing. The results suggest that these regions perform a function

different from the left dACC. Although they are involved in error processing, they do not seem to respond to worse-than-expected outcomes as the left dACC does.

Consistently interpreted tracking errors versus correct tracking trials

Contrasting consistently interpreted tracking errors and correct tracking trials ($Terr_t - Tcor$), I found a greater activation for consistently interpreted tracking errors ($Terr_t$) in the following regions: the dACC bilaterally, the left medial frontal gyrus (MFG), the left superior frontal gyrus (SFG), and the inferior parietal gyri bilaterally, see Figure A2. Inconsistently interpreted tracking errors ($Terr_m$) do not show activity in these regions when compared with correct tracking trials.

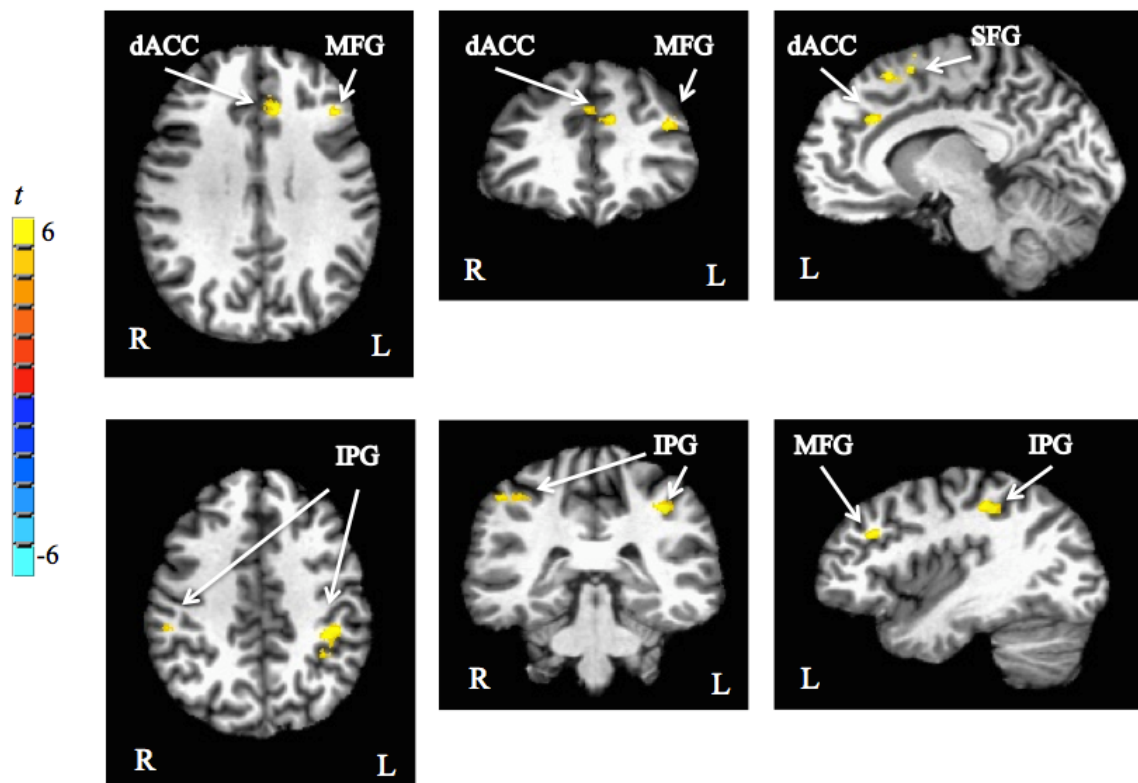


Figure A2. Consistently interpreted tracking errors. Regions identified in a whole brain analysis contrasting consistently interpreted tracking errors and correct trials ($Terr_t - Tcor$). The arrows point to the regions more active for consistently interpreted tracking errors. They are the dACC bilaterally, the left medial frontal gyrus (MFG), the left superior frontal gyrus (SFG), and the inferior parietal gyri (IPG) bilaterally.

ROI analysis.

In the ROI analysis, I explored activation in these regions during feedback processing. In particular, I was interested to see if these regions would show the same pattern of activation as the left dACC identified in all errors versus correct trials contrast – greater activation for inconsistently interpreted errors during feedback processing. Contrasting the activation on consistently interpreted trials with activation on inconsistently interpreted trials (FTer_t – FTer_m), I did not find a consistency effect for neither of the regions, $p > 0.05$, see Figure A3 and Table A2. This result indicates that the regions shown in Figure A2 perform a function different from left dACC identified in all errors versus all correct contrast. The regions also do not show greater activation for errors compared with correct trials during feedback (i.e., negative feedback versus positive feedback) $p > 0.05$. This pattern of results suggests that these areas participate in error detection, but not in evaluation of the performance outcome. They can potentially carry the information about occurrence of an error to other areas such as dACC or basal ganglia, which later issue an evaluative signal suggesting that performance is worse than expected.

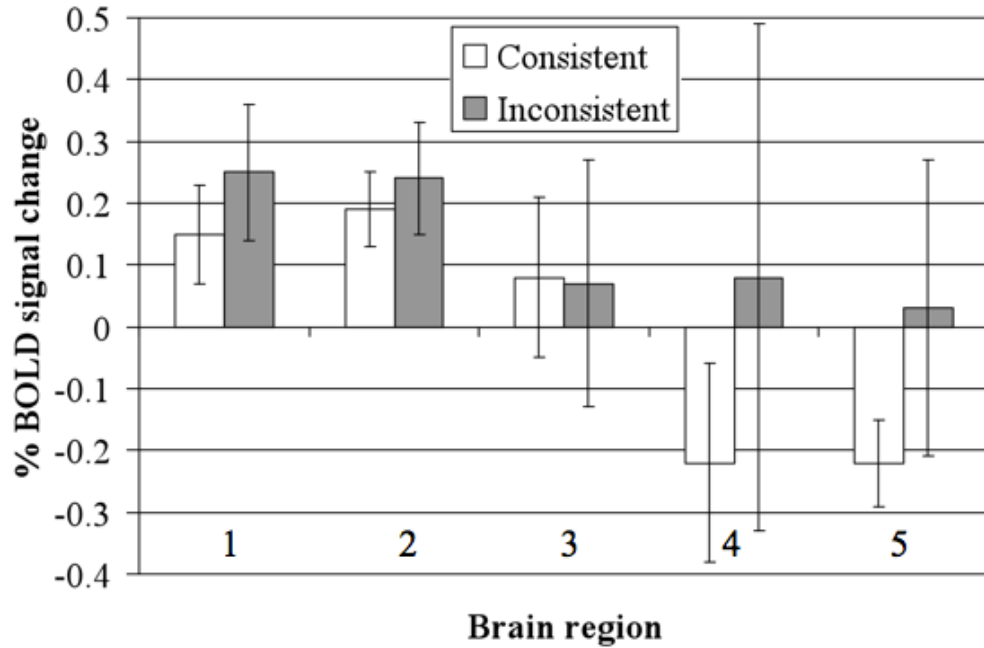


Figure A3. ROI analysis: consistently interpreted tracking errors versus inconsistently interpreted tracking errors during feedback processing in brain regions identified with the contrast (Terr_t – Tcor). 1 = RIP, 2 = LIP, 3 = LMF, 4 = RACC, 5 = LACC. Neither of the comparisons reaches significant level. Thus the bilateral IP, the left MFG, and the bilateral ACC do not show a consistency effect during feedback processing. (The region in the left ACC described here is more anterior and inferior compared to the regions in the left ACC identified in “all errors vs. correct trials” contrast and in “consistent interpretations vs. correct trials” contrast.)

Table A2. Group analysis for the contrast: “Consistently interpreted tracking errors > correct tracking” during the Task phase.

Cluster	Active volume	Max <i>t</i> values	Max <i>t</i> value locations		
	mm3		RL, AP, IS	Terr_t	Tcor
R IPG	788	7.82	-30, 44, 47	1.01(0.20)	0.79(0.21)
L IPG	487	7.625	38, 30, 41	0.77(0.13)	0.57(0.12)
SFG	465	7.392	7, -14, 50	0.45(0.18)	0.18(0.14)
L & R dACC	345	6.605	9, -29, 29	0.37(0.14)	0.08(0.10)

L MFG	248	6.446	38, -30, 25	0.33(0.06)	0.05(0.05)
-------	-----	-------	-------------	------------	------------

Inconsistently interpreted errors versus consistently interpreted errors during

Feedback

Contrasting inconsistently interpreted errors with consistently interpreted errors ($F_{Ter_m} + F_{Mer_t} - (F_{Ter_t} + F_{Mer_m})$), I found a single region in the left IPG to be active more for inconsistently interpreted errors, see Figure A4 and Table A3. ROI analysis did not show a consistency effect in this region during task performance. A repeated-measures ANOVA with Consistency (consistent vs. inconsistent) and Error type (tracking vs. memory) showed neither the main effects nor the interaction between factors, $p > 0.7$. Pair-wise t -tests did not show consistency effects for tracking errors (for consistently interpreted errors, $M = -0.15$, $SD = 0.34$; for inconsistently interpreted errors, $M = -0.13$, $SD = 0.83$). Pair-wise t -test did not show consistency effect for memory errors also (for consistently interpreted errors, $M = -0.15$, $SD = 0.34$, for inconsistently interpreted errors, $M = -0.11$, $SD = 0.39$). Thus, this region seems to respond to unexpected negative feedback, but not to detection of an error during task performance. This pattern of results suggests that this region performs a function different from the left dACC identified in “all error versus all correct trials” contrast.

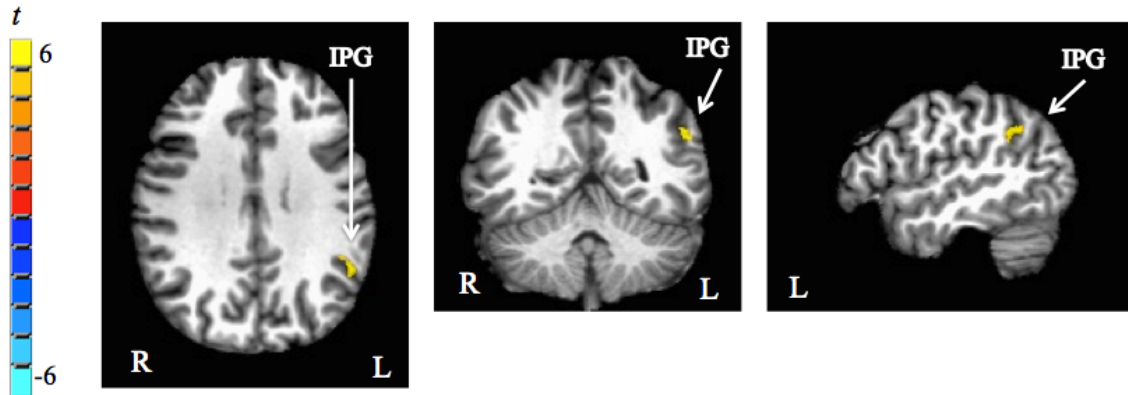


Figure A4. Feedback: Inconsistently interpreted errors. Contrasting inconsistently interpreted errors and consistently interpreted errors during Feedback ($F_{Ter_m} + F_{Mer_t}$) – ($F_{Ter_t} + F_{Mer_m}$), I find a single area in the left IPG responding greater to inconsistently interpreted errors.

Table A3. Group analysis results for the contrast: “Inconsistently interpreted errors > Consistently interpreted errors” during the Feedback phase.

Cluster	Active volume	Max t values	Max t value locations				
	mm3		RL, AP, IS	FMer_t	FMer_m	FTer_t	FTer_m
L IPG	186	5.914	52, 47, 25	0.65(.09)	0.05(.14)	0.07(.07)	0.45(.20)

* * *

Exploring the contrasts “consistently interpreted errors versus all correct trials” and “consistently interpreted tracking errors versus correct tracking trials”, I found several brain areas, other than the ACC, being active for consistently interpreted errors during the Task phase. They were the left SFG, left MFG, right IPG and left IPG. The ROI analysis of these areas revealed that these areas do not show the same pattern of activation as the left dACC does. Although, during the Task phase, these areas were active selectively for consistently interpreted errors, during the Feedback phase, the effect

of the consistency of interpretation was not significant. At the same time, a brain area, identified in the contrast “inconsistently interpreted errors versus consistently interpreted errors during feedback processing”, the left IPG, exhibited an opposite pattern. It showed a consistency effect during the Feedback phase but not during the Task phase.

The presence of the areas that respond to detected errors but not to unexpected negative feedback, and the areas that respond to unexpected negative feedback but not to detected errors suggests that error detection and processing of unexpected negative feedback are different phenomena. However, these phenomena might have common processes, for example, evaluation of performance outcome. Left dACC might be involved in this process, since this area responds to both error detection and unexpected negative feedback.

The present study does not allow making conclusions regarding the functions of the other areas involved in processing of errors and negative feedback. Although from previous studies we know that the left MFG, for example, is involved in working memory, especially it is important for storage of spatial information (e.g. Leung & Goldman-Rakic, 2002). The fact that we find it to be active for detected errors and not during feedback processing is consistent with the idea that this region supports memory processes. Possibly, after detecting an error the system needs to retain the information about the occurrence of the error, so later it can make a correct judgment about the type of the error. In the present study, we find that the left and right IPG are active during detected errors compared to correct trials during the Task phase. In the previous literature, these areas were also found to be active during correct tracking (e.g. Culham, Brandt, Cavanagh, Kanwisher, Dale, Roger, & Tootell, 1998). Hypothetically, the areas,

performing tracking function, work harder when tracking becomes difficult and errors are likely to occur. Thus the additional recruitment of these areas could serve as a signal to high-level areas that tracking is likely to go wrong. These conclusions are merely suggestive since they are based on reverse inferences. Further research is required to determine the role of these areas in processing errors and negative feedback.

APPENDIX B

Correct Memory phase versus Correct Tracking phase

The present study shows that the left dACC responds greater to tracking errors compared to memory errors. One of the possible explanations is that memory and tracking phases impose differential demands on visual system. In this case, we should observe a differential ACC activation to tracking and memory phases even in the absence of an error. A pair-wise *t*-test, comparing Tcor with Mcor (see Figure A5), showed a marginally significant effect of the Trial phase, $t(16) = 1.935$, $p = 0.071$. Thus tracking seems to engage the left dACC region greater than memory retrieval even in the absence of an error.

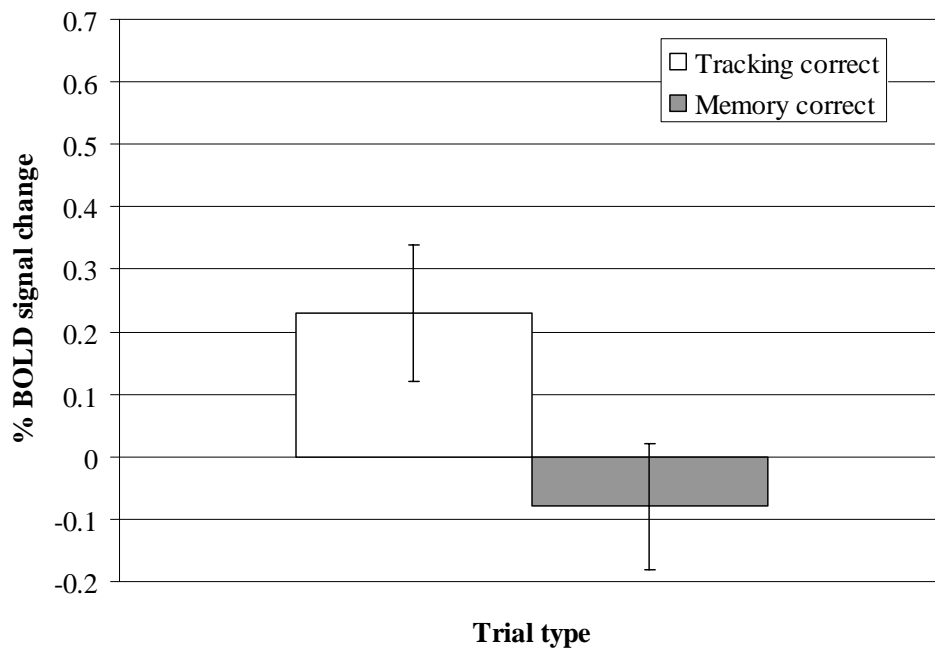


Figure A5. ROI analysis: the left dACC. The region shows an effect of the Trial Phase: correct tracking phase is associated with a greater dACC activation compared to correct memory retrieval phase.

REFERENCES

- Barch, D. M., Braver, T. S., Sabb, F. W., & Noll, D. C. (2000). Anterior cingulate and the monitoring of response conflict: Evidence from an fMRI study of overt verb generation. *Journal of cognitive Neuroscience*, *12*, 298-309.
- Botvinick, M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624-652.
- Botvinick, M., Cohen, J. D. & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, *8*, 539-546.
- Botvinick, M., Nystrom, L. E., Fissell K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, *402*, 179-181.
- Botvinick, M., & Plaut, D. C. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review*, *111*, 395-429.
- Brown, J. W. (2009). Conflict effects without conflict in anterior cingulate cortex: Multiple response effects and context specific representation. *NeuroImage*, *47*, 334-341.
- Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science*, *307*, 1118-1121.
- Brown, J. W., & Braver, T. S. (2007). Risk prediction and aversion by anterior cingulate cortex. *Cognitive, Affective, & Behavioral Neuroscience*, *7*, 266-277.
- Brown, J. W., & Braver, T. S. (2008). A computational model of risk, conflict, and individual difference effects in the anterior cingulate cortex. *Brain Research*, *1202*, 99-108.
- Buckner, R.L., Koutstaal, W., Schacter, D. L., Wagner, A. D., & Rosen, B. R. (1998). Functional-anatomical study of episodic retrieval using fMRI. I. Retrieval effort versus retrieval success. *Neuroimage*, *7*, 151-162.
- Cabeza, R., Dolcos, F., Graham, R., & Nyberg, L. (2002). Similarities and differences in the neural correlates of episodic memory retrieval and working memory. *Neuroimage*, *16*, 317-330.
- Carter, C. S, Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*, 747-749.

- Carter, C. S., van Veen, V. (2007). Anterior cingulate cortex and conflict detection: An update of theory and data. *Cognitive, Affective & Behavioral Neuroscience*, 7, 367-379.
- Cooper, R., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology*, 17, 297-338.
- Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L. & Petersen, S. E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science*, 248, 1556-1559.
- Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1997) Transient and sustained activity in a distributed neural system for human working memory. *Nature*, 386, 608-611.
- Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance images. *Computers and Biomedical Research*, 29, 162-173.
- Culham, Brandt, Cavanagh, Kanwisher, Dale, Roger, & Tootell, (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology*, 80, 2657-2670.
- Culham, Cavanagh, & Kanwisher, (2001). Attention response functions: characterizing brain areas with fmri activation during parametric variations of attentional load. *Neuron*, 32, 737-745.
- Donaldson, D. I., Wheeler, M. E., & Petersen, S. E. (2009). Remember the source: Dissociating frontal and parietal contributions to episodic memory. *Journal of Cognitive Neuroscience*, 22, 377-391.
- Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception & Psychophysics* 25, 249-263.
- Falkenstein, M., Hohnsbein, J., Hoorman, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology* 78, 447-455.
- Fitzgerald, K. D., Welsh, R. C., Gehring, W. J., Abelson, J. L., Himle, J. A., Liberzon, I., & Taylor, S. F. (2005). Error-related hyperactivity of the anterior cingulate cortex in obsessive-compulsive disorder. *Biological Psychiatry*, 57, 287-294.
- Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, 4, 385-390.
- Hajcak, G., McDonald, N., & Simons, R. F. (2003). Anxiety and error-related brain activity. *Biological Psychology*, 64, 77-90.

- Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109, 679-709.
- Howe, P. D., Horowitz, T. S., Morocz, I. A., Wolfe, J., & Livingstone, M. S. (2009). Using fMRI to distinguish components of the multiple object tracking task. *Journal of Vision*, 9, 1-11.
- Huk, A.C., Dougherty, R.F., & Heeger, D.J. (2002). Retinotopy and functional subdivision of human areas MT and MST. *Journal of Neuroscience*, 22, 7195-7205.
- Humphreys, G. W., & Forde, E. M. E. (1999). Disordered action schema and action disorganization syndrome. *Cognitive Neuropsychology*, 15, 771-811.
- Jonides, J., Lewis, R.L., Nee, D.E., Lustig, C.A., Berman, M.G., & Moore K.S. (2008). The mind and brain of short term memory. *Annual Review of Psychology*, 59, 193-224.
- Jovicich, J., Peters, R. J., Koch, C. Braun, J., Chang, L., & Ernst, T. (2001). Brain areas specific for attentional load in a motion-tracking task. *Journal of Cognitive Neuroscience*, 13, 1048-58.
- Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, 303, 1023-1026.
- Kim, D. H., Adalsteinsson, E., Glover, G. H., Spielman, D. M. (2002). Regularized higher-order in vivo shimming. *Magnetic Resonance in Medicine*, 48, 715-722.
- Konishi, S., Nakajima, K., Uchida, I., Kameyama, M., Nakahara, K., Sekihara, K., & Miyashita, Y. (1998). Transient activation of inferior prefrontal cortex during cognitive set shifting. *Nature Neuroscience*, 1, 80-84.
- Milham, M. P., Banich, M. T., Webb, A., Barad, V., Cohen, N. J., Wszalek, T., & Kramer, A. F. (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Cognitive Brain Research*, 12, 467-473.
- Miltner, W. H. R., Braun, C. H., Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error detection. *Journal of Cognitive Neuroscience*, 9, 788-798.
- Mueller, E. M., Nguyen, J., Ray, W. J., & Borkovec, T. D. (2010). Future-oriented decision-making in generalized anxiety disorder is evident across different versions of the Iowa Gambling task. *Journal of Behavior Therapy and Experimental psychiatry*, 41, 165-171.

- Nee, D.E., & Jonides, J. (2008). Neural correlates of access to short-term memory. *PNAS*, *15*, 14228-14233.
- Norman, D. A. (1981). Categorization of action slips. *Psychological Review*, *88*, 1–14.
- Nieuwenhuis, S., Slagter, H. A., von Geusau, N. J. A., Heslenfeld, D. J., Holroyd, C. B. (2005). Knowing good from bad: Differential activation of human cortical areas by positive and negative outcomes. *European Journal of Neuroscience*, *21*, 3161-3168.
- Perrish, T. B., Gitelman, D. R., LaBar, K. S., Mesulam, M. M. (2000). Impact of signal-to-noise on functional MRI. *Magnetic Resonance in Medicine*, *44*, 925-932.
- Pylyshyn Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179-197.
- Runganath, C., Johnson, M. K., & D'Esposito, M. (2003). Prefrontal activity associated with working memory and episodic long-term memory. *Neuropsychologia*, *41*, 378-389.
- Reason, J. (1990). *Human error*. Cambridge, England: Cambridge University Press.
- Rizzolatti, G., Riggio L., & Sheliga B.M. (1994). Space and selective attention. In Umiltà C, Moscovitch M.(eds) *Attention and performance*, XV Bradford Book Cambridge MA, 231-265.
- Schwartz, M. F., Reed, E. S., Montgomery, M. W., Palmer, C., & Mayer, N. H. (1991). The quantitative description of action disorganization after brain damage: A case study. *Cognitive Neuropsychology*, *8*, 381–414.
- Talairach, J., & Tournoux, P. (1988). Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional proportional system: an approach to cerebral imaging. New York, New York: Georg Thieme Verlag.
- Talmi, D., Grady, C. G., Goshen-Gottstein, Y., & Moscovitch, M. (2005). Neuroimaging the serial position curve: A test of single-store versus dual-store models. *Psychological Science*, *16*, 716-723.
- Van Veen, V., Holroyd, C. B., cohen, J. D., Stenger, V. A., & Carter, C. S. (2004). Errors without conflict: Implications for performance monitoring theories of anterior cingulate cortex. *Brain & Cognition*, *56*, 267-276.
- Ward, B. (2000). Deconvolution analysis of fMRI time series data. Biophysics Research Institute, Medical College of Wisconsin, Milwaukee, WI.
- Zeki, S. (1993). *A Vision of the Brain*. Oxford, England: Wiley-Blackwell.