CALIFORNIA STATE UNIVERSITY SAN MARCOS

THESIS SIGNATURE PAGE

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

MASTER OF ARTS

IN

PSYCHOLOGY

THESIS TITLE: Ego Depletion and the Confidence-Accuracy Relationship

AUTHOR: Nicole V. Mills

DATE OF SUCCESSFUL DEFENSE: May 3, 2016

THE THESIS HAS BEEN ACCEPTED BY THE THESIS COMMITTEE IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS IN PSYCHOLOGY.

Dustin Calvillo, Ph.D.  
THESIS COMMITTEE CHAIR  
SIGNATURE  
DATE

Carrick Williams, Ph.D.  
THESIS COMMITTEE MEMBER  
SIGNATURE  
DATE

Marie Thomas, Ph.D.  
THESIS COMMITTEE MEMBER  
SIGNATURE  
DATE
Ego Depletion and the Confidence-Accuracy Relationship

Nicole V. Mills

California State University San Marcos
Abstract

Law enforcement often uses a witness’s confidence in identification as an indicator of accuracy. This is due to the seemingly positive relationship found between self-reported confidence and accuracy in which those who are more confident in their identifications tend to be more accurate. However, the strength and reliability of this relationship has been called into question. Researchers have found a relatively high correlation between confidence and accuracy, but only for those who make a correct identification. Therefore, it is essential to examine the factors that influence reports of confidence when examining the relationship between confidence and accuracy. The goal of this study was to expand on the research into these factors by examining the effect of self-regulation on confidence in eyewitness identification. Ego depletion is the taxing of self-regulation resources, resulting in subsequent attempts at self-regulation to fail. It was hypothesized that ego depletion would be associated with the inability to monitor the ephoric experiences necessary to make a confidence judgment. Participants took part in a facial recognition task used to imitate an eyewitness identification. Results showed that there was no difference between participants who were ego depleted and those who were not on the accuracy of identifications or the calibrations of old and new faces. Individuals showed a moderately high correlation between confidence and accuracy when presented with old faces, but that correlation decreased significantly when participants were presented with new faces. The findings suggest that self-regulation does not play a role in the relationship between confidence and accuracy. However, as the identification of new faces as old faces is analogous to identifying an innocent suspect in an eyewitness identification, these findings do suggest that confidence may not be an appropriate postdictor of accuracy.
Ego Depletion and the Confidence-Accuracy Relationship

Eyewitness identification is important for the apprehension and conviction of criminals. Misidentification, however, is one of the main sources of wrongful convictions in the United States. Seventy-five percent of wrongfully convicted individuals who were later exonerated due to DNA evidence were in prison due solely to misidentifications (Innocence Project, 2014). This makes understanding what determines misidentifications important to the criminal justice system. Psychologists assist in this endeavor by examining the factors that might influence a witness’s identification of a criminal. For example, a witness’s confidence in his or her decision is often taken at the end of an identification as a measure of the accuracy, termed postdictor, in that identification (Technical Working Group for Eyewitness Evidence, 1999). This is due to the positive relationship found between confidence and accuracy, in which those who are more confident in their identifications tend to be more accurate. However, the strength of this relationship has been called into question by several researchers. For example, Bornstein and Zickafoose (1999) found that individuals have a tendency to be overconfident in their decisions in general, as well as those concerning eyewitness identifications. Therefore, it is essential to examine other factors that might disrupt or influence this positive relationship.

It has been suggested by recent research that self-reports of confidence might be influenced by one’s ability to self-regulate (DeBono & Muraven, 2013). Self-regulation is necessary to engage in goal-directed behavior and in overriding automatic responses (Fischer, Kastenmüller, & Asal, 2012; Hagger, Wood, Stiff, & Chatzisarantis, 2010; Reinhard, Scharmach, & Stahlberg, 2013). In an eyewitness scenario, making a correct identification and reporting confidence in that decision might require self-regulation by regulating distraught emotions or needing to inhibit an automatic response to be confidence in a decision. Recent
research suggests that self-regulation is a limited resource and that intense bouts of self-regulation make subsequent acts of self-regulation much more difficult (Baumeister, Bratslavsky, Muraven, & Tice, 1998). This phenomenon is termed ego depletion. Researchers have found that those who are ego depleted are more vulnerable to false memory, more susceptible to suggestions, and have decreased reasoning performance (Otgaar, Alberts, & Cuppens, 2012A; 2012B; Schmeichel, Vohs, & Baumeister, 2003). The purpose of this study was to examine the effect of self-regulation on the relationship between confidence and accuracy in a situation in which consequences of misreported confidence have a great impact: eyewitness identification.

**Eyewitness Identification and the Confidence-Accuracy Relationship**

**Eyewitness identification.** Eyewitness identification has an integral role in the United States judicial system. Witness identification is important for developing leads and identifying criminals. As a result, there is a high reliance on eyewitness identification for convictions when physical evidence is missing (Technical Working Group for Eyewitness Evidence, 1999). However, misidentification by eyewitnesses is one of the greatest causes of wrongful convictions. With 75% of cases overturned by DNA evidence being due to misidentifications, there is a strong need for research to focus on the factors that influence eyewitness identifications (Innocence Project, 2014).

Research in eyewitness identification can be classified into two categories: research investigating estimator variables and research investigating system variables (Wells, 1978). Estimator variables are circumstantial or individual differences between witnesses that might have an effect on recognition memory such as race, gender, or the amount of time a witness sees the culprit. Results of studies that focus on estimator variables help to identify reliable witnesses, but cannot be controlled by the criminal justice system because investigators cannot
change who they have as witnesses. System variables are characterized as procedural changes that ensure that the witness is able to give the best possible recollection of a person or event and reduce the possibilities of inaccuracies. These variables are under the control of the judicial system and include topics such as proper procedure, lineup formation, lineup instructions, and communication with witnesses (Wells, 1978). For the proposed project, the witness’s self-reported confidence, used as a postdictor of accuracy, an estimator variable, and the effects of self-regulation capacity are examined. Self-regulation capacity has the potential to be a system and estimator variable. Focused attention and emotion regulation when recalling important event details, while excluding details that are not relevant to an investigating officer, could result in ego depletion. In this case, the capacity of self-regulation would be considered a system variable. However, the stress of the crime or fatigue as a result of the event may also result in ego depletion. Under these circumstances, self-regulation capacity would be considered an estimator variable. Although law enforcement cannot control for estimator variables, they can control for system variables. Should ego depletion influence the confidence-accuracy relationship, law enforcement officials can adapt their polices to prevent the ego depletion effect on their part.

Confidence-accuracy relationship. Regardless of the type of lineup shown or report taken, a measure of confidence is supposed to be taken by law enforcement at the end of an identification, rejection of lineup, or retelling of a crime (Technical Working Group for Eyewitness Evidence, 1999). This is, in part, due to the positive relationship that has been found between confidence and accuracy, where those who are accurate tend to have higher confidence in their identifications (Brewer, & Wells, 2006; Palmer, Brewer, Weber, & Nagesh, 2013; Sauer, Brewer, Zweck, & Weber, 2010; Sporer, 1993; Sporer, Penrod, Read, & Cutler, 1995). Thus a
witness’s confidence in his or her decision is often taken as a measure of the accuracy in that identification.

A meta-analysis conducted by Sporer, Penrod, Read, and Cutler (1995) found a positive correlation between confidence and accuracy ($r = .41$) for participants who made an identification from a lineup. This positive correlation suggests that an eyewitness’s self-reported confidence can adequately postdict the accuracy of his or her identification. Similar findings have been discovered by several researchers and support the idea that the relationship between confidence and accuracy is relatively high and positive (Busey, Tunnicliff, Loftus, & Loftus, 2000; Kassin, 1985; Kassin, Rigby, & Castillo, 1991; Smith, Lindsay, & Pryke, 2000). However, individuals have a tendency to be overconfident in their decisions and knowledge, not just in eyewitness scenarios, but in general knowledge cases as well. In a study that looked at the confidence-accuracy relationship for both reports of general knowledge and in an eyewitness scenario, participants were 16% overconfident in their reports of general knowledge and 19% overconfident in their eyewitness reports (Bornstein & Zickafoose, 1999). This tendency towards overconfidence is called the overconfidence bias, and there are several theories about why it occurs. For example, some researchers have found evidence that overconfidence is the result of an attempt at reducing cognitive dissonance where the desire to be accurate conflicts with a potentially inaccurate judgment (Blanton, Pelham, DeHart, & Carvallo, 2001). More recently, Koriat (2012) hypothesized that confidence judgments are measures of reliability rather than validity and that overconfidence is a result of reliability measures being consistently higher than validity measures. The severity of these findings lie in the implications an overconfident misidentification might have in court.
Jurors often cannot distinguish between accurate and inaccurate witnesses, and research has supported that the juror’s judgments are based on the perceived confidence of that witness (Kassin et al., 1991). This suggests that, when lacking appropriate forensic evidence, if a witness is extremely confident in his or her identification, then a conviction is likely, regardless of the accuracy of that identification. The negative repercussions are that innocent individuals are in danger of being convicted of crimes they did not commit.

Research has also identified several factors that influence the confidence-accuracy relationship including those that are a result of the crime itself, those that are produced from the witness, and those that surround the lineup. Factors resulting from the event of the crime can include exposure duration, the type of incident, or if the culprit was wearing a disguise. Exposure duration is the amount of time spent seeing the culprit. Palmer et al. (2013) found evidence that shorter exposures to the culprit resulted in overconfidence in identifications whereas Memon, Hope, and Bull (2003) found evidence for more appropriate confidence ratings, with higher confidence judgments given to accurate identifications, for shorter exposures in which higher confidences were reported when a correct identification was made and lower confidences were reported when incorrect identifications were made. In regards to the type of crime witnessed, Clifford and Hollin (1981) found that accuracy was better for those who saw a nonviolent crime compared to those who saw a violent crime. They also found that confidence ratings were higher for correct identifications of nonviolent crimes, whereas no difference was found in the confidence ratings of violent crimes regardless of the accuracy of the identification. While investigating the impact of varying levels of disguises on facial recognition, researchers found that when the level of disguise increased, accuracy in identifications decreased (Mansour, Beaudry, Bertrand, Kalmet, Melsom, & Lindsay, 2012). They also found mixed results for
reported confidence in these identifications. When witnesses made a correct identification, higher levels of disguise resulted in lower confidence ratings. However, when witnesses rejected the lineups by stating that the culprit is not present, the effectiveness of the disguise was dependent on the type of lineup shown, with simultaneous lineups and the disguise of a hat resulting in higher confidences in their decisions.

Certain qualities or actions of the witness can also influence the relationship between confidence and accuracy. For example, retrospective self-awareness is when witnesses are shown videos of their own performance or reflect on their own performance before making a confidence judgment. Witnesses who engaged in retrospective awareness showed a higher relationship between confidence and accuracy than those who did not (Kassin, 1985; Kassin et al., 1991). In addition, motivation and fatigue can contribute to the relationship. Shaw and Zerr (2003) found that when witnesses have a high motivation to make an identification, the relationship between confidence and accuracy decreases as a result of inflated confidence. It was also found that witnesses who experience sleep loss or fatigue were more likely to be overconfident in their identifications and more vulnerable to suggestion (Blagrove & Akehurst, 2000).

Lastly, factors surrounding the lineup, such as retention interval, foil similarity, and lineup type, can also influence the confidence-accuracy relationship. Retention interval is the time between the witness’s exposure to the crime and when he or she is asked to make an identification. Research has shown that when the wait between these two events is longer, witnesses have a tendency of being overconfident in their identification (Palmer et al., 2013; Sauer et al., 2010). Foil similarity is the amount of resemblance between the suspected culprit in the lineup and the fillers (called foils) put into the lineup. Charman, Wells, and Joy (2011) found
that if there are foils in a lineup that are too dissimilar to the culprit, a witness’s confidence can become inflated. Eyewitness lineups can also vary, influencing the confidence-accuracy relationship. Lineups typically happen in one of two ways. Simultaneous lineups, the most common, show all of the individuals in the lineup at one time, whereas sequential lineups present each photograph in the lineup one at a time. The confidence-accuracy relationship is greater when lineups are presented sequentially (Sporer, 1993).

These are just a few of the many factors found influencing the relationship between confidence and accuracy. These factors, in addition to how jurors are influenced by confident witnesses, provide support that the relationship between confidence and accuracy is not always reliable. In certain cases, law enforcement can control for this variability, such as using a sequential lineup rather than a simultaneous lineup. However, in cases when that variability cannot be controlled by law enforcement, such as how violent a crime is, being able to draw from previous research can inform officials about how reliable a witness is based on the circumstances of the case. Likewise, knowing if self-regulation plays a part in this relationship can assist law enforcement officials in the evaluation of the reliability of a witness.

**Confidence judgments.** Confidence is a metacognitive judgment that individuals make based on their perceived cognitive abilities to perform the task they are presented with (Costermans, Lories, & Ansay, 1992; Efklides, 2008). According to the cue hypothesis, these judgments are made retrospectively and not during the time of the identification by using internal cues. When those cues are not strong enough, the metacognitive judgments are influenced by external cues (Bradfield, Wells, & Olson, 2002). Originally, this hypothesis was created to explain the postidentification feedback effect where witnesses take cues from the lineup.
administrator to assess their confidence in an identification, leading to inflated confidence judgments (Wright & Skagerberg, 2007).

Recently, this hypothesis was adapted by Charman, Carlucci, Vallano, and Gregory (2010) to include the search and evaluation of the external cues. The selective cue integration framework (SCIF) is a three stage cognitive process in which witnesses assess their own confidence. The first stage is the assessment stage in which the witness evaluates internal cues. This means that witnesses will assess, within themselves, how confident they are in their decision. If the internal cue is weak, then the witness progresses to the search stage. During this stage, the witnesses search external cues, often from the lineup administrator, to confirm their identification. If the external cue does not confirm their identification, that cue is not integrated into the confidence assessment. If the external cue does confirm the identification, then the cognitive process continues to the evaluation stage. In the evaluation stage, the witness will evaluate the credibility of the external cue. If, at this stage, the external cue is deemed credible, then it is integrated into the confidence assessment. The cue hypothesis and adapted selective cue integration framework also provide a theoretical framework for assessing witness’s confidence outside of the postidentification feedback effect, as it allows researchers to examine factors that influence confidence at different stages of its assessment (Charman et al., 2010). I propose that making accurate confidence judgments requires the use of self-regulation resources, and, in turn, self-regulation is necessary to the SCIF process. See Figure 1 for a visual representation of this model.

As previously stated, this process begins in the assessment stage with the witness evaluating internal cues by judging ephoric experiences. Ephoric experience is the subjective similarity one perceives between a presented stimulus and one’s memory, and research has
suggested that this similarity is what determines the strength of an internal cue (Bradfield et al., 2002; Charman & Wells, 2012). In the case of eyewitness identification, the presented stimulus is the lineup of individuals that is compared to the witness’s memory for the crime. Strong ephoric experiences, such that the witness perceives a strong similarity between his or her memory for the culprit and the presented suspect, are related with accurate identifications and high reported confidences. In turn, weak ephoric experiences are often followed by inaccurate identifications and low reported confidences (Sauer, Weber, & Brewer, 2012). I proposed that this internal evaluation requires the use of self-regulation to monitor the ephoric experiences, and those individuals without the ability to self-regulate, when asked to make a confidence judgment, would be unable to appropriately estimate their level of confidence. I also proposed that this would result in the overconfidence often found in eyewitness identifications and would, in turn, produce a lower calibration between confidence and accuracy.

Calibration calculations will reveal if correct judgments are given high confidence ratings and if incorrect judgments are given low ratings by comparing concordant judgments with discordant judgments. Concordant judgments occur when higher confidence is given to accurate identifications and lower confidence is given to inaccurate identifications. Discordant judgments occur when lower confidence rating are reported when an accurate identification is made and higher confidence ratings are reported when inaccurate identifications are made.

**The Strength Model of Self-Regulation**

Self-regulation is one’s ability to override an automatic response that is not wanted in a situation. Functions of self-regulation include goal-directed behavior, inhibiting automatic or unwanted responses or actions, error monitoring, emotion control and self-monitoring (Kalory, 2012). For example, self-regulation is used when trying to control emotions in a tense situation.
or refraining from automatically saying one’s first thought. According to the strength model of self-regulation, self-regulation is seen as a limited resource and when that resource is taxed, a decrease in the success of subsequent self-regulation attempts is found (Baumeister et al., 1998). The taxing of these resources is called ego depletion. The strength model of self-regulation views self-regulation as a muscle. When a muscle is overused, successive attempts to employ it are more difficult and less successful (Baumeister et al., 1998; Hagger et al., 2010). In a similar vein, when individuals are ego depleted due to recent acts of self-regulation, their following attempt to employ self-regulation is more difficult and less successful. This inability to self-regulate is a concern because self-regulation is important in goal-directed behaviors and controlling automatic responses. According to some, ego depletion is associated with a lack of control over dominant or automatic responses (Fischer et al., 2012; Hagger et al., 2010; Reinhard et al., 2013).

Ego depletion is manipulated using a dual task paradigm (Baumeister, Bratslavsky, Muraven, & Tice, 1998). This paradigm consists of two consecutive tasks used to simulate the taxing of the self-regulation resource. Those who are assigned to the ego depletion condition receive two tasks requiring self-regulation and those who are assigned to the control condition only receive the self-regulating exercise in the second of the two tasks. For example, Reinhard, Scharmach, and Stahlberg (2013) examined the effect that depleted self-regulation played on the ability to detect deception. For their manipulation, they had participants first transcribe a provided article with those in the ego depletion condition being asked to omit common letters. Participants in the ego depletion condition had to employ self-regulation by inhibiting the automatic response to type normally. For the second task, all participants were given a deception
task that required self-regulation to monitor for accurate responses. As a result, those in the ego depletion condition showed a decrease in their deception detection ability.

Ego depletion has been linked to increases in risk-taking behaviors, vulnerability to false memory, and susceptibility to suggestions (Fisher et al., 2012; Otgaar, Alberts, & Cuppens, 2012A, 2012B). It has also been associated with decreases in logic, reasoning, cognitive extrapolation, feelings of guilt, prosocial behavior, and event-based prospective memory (Li, Nie, Zeng, Huntoon, & Smith, 2013; Schmeichel, Vohs, & Baumeister, 2003; Xu, Bègue, & Bushman, 2012). Hagger et al. (2010) reviewed the effect of depleted self-regulation resources in a meta-analysis and found that those who were ego depleted also showed decreases in effort, positive affect, self-efficacy, and blood glucose. Most of these factors are related to functions of self-regulation such as goal directed behavior, inhibiting automatic or unwanted responses or actions, error monitoring, emotion control. I propose that self-monitoring is the function of self-regulation that is associated with monitoring ecphoric experiences as it is associated with monitoring internal and external events in the pursuit of a goal, in this case reporting confidence.

One study has looked at the effects of self-regulation resources on self-reporting confidence (DeBono & Muraven, 2013). These researchers found that participants who experienced ego depletion also showed decreased levels of reported confidence, controlling for overconfidence. This suggests that self-regulation could influence the confidence-accuracy relationship. Specifically, Debono and Muraven (2013) examined the role self-regulation played when predicting performance on a future task and that individual’s reported confidence in the prediction. However, this differs from the proposed project, as confidence judgments in eyewitness identifications are postdictors of accuracy and not predictors of accuracy, meaning
that confidence judgments are made after the identifications and not before identifications. In addition, these researchers focused on confidence in the context of an individual’s ability to physically accomplish a task, playing a video game, while the present study focuses on the confidence in a previous memory judgment, identifying a previously seen face, which is a judgment based on an individual’s ability to cognitively perform a task.

Furthermore, the study’s intention was to reduce the tendency to be overconfident, in doing so, suggesting that the automatic response brought about by the depletion of self-regulation is to be underconfident in one’s decisions (Debono & Muraven, 2013). While this may be the case when predicting behavior, I theorize that depleted self-regulation is one of the factors that brings about overconfidence when postdicting accuracy.

**Present Study**

The current study aimed to determine the effect of self-regulation on the relationship between confidence and accuracy in an identification paradigm. This project was the first of its kind, to my knowledge, to look at decreased self-regulation resources in the eyewitness identification paradigm. In addition, there appears to be no research looking at the effects of decreased self-control abilities on the confidence-accuracy relationship in an eyewitness identification scenario. Specifically, this study aimed to determine if those who are cognitively depleted would not have the self-regulation capacity needed to monitor the ecphoric experiences presented in the assessment stage of the SCIF process. This study used a facial recognition paradigm to simulate an eyewitness identification scenario and ego depletion was manipulated by using a letter e task. After being presented with a set of faces to learn, participants were assigned to a high or low ego depletion condition, and then they saw a larger set of faces and decided if each face was old or new and reported their confidence in their decisions. These
responses enabled the assessment of accuracy and confidence for each identification or rejection. In addition, it has been suggested that a difficult task, without self-regulation, might elicit the same results as a product of mental fatigue. As a result, a difficult task, analyzing anagrams, was used as a third condition to insure that this is not the case, as it did not ask the participants to inhibit any response, but rather to read the words presented and determine if they employed the same letters.

It was theorized that making accurate confidence judgments required the use of self-regulation resources. Specifically, that the internal evaluation one makes in the assessment stage of the SCIF process required the use of self-regulation to monitor the ephoric experiences. In turn, when individuals without the ability to self-regulate are asked to make a confidence judgments they would be unable to appropriately estimate their level of confidence and produce a lower calibration between confidence and accuracy. I further proposed that the function of self-regulation that is associated with monitoring ephoric experiences was self-monitoring, as it is associated with monitoring internal and external events in the pursuit of a goal.

I hypothesized that:

1) The depletion of self-regulation resources would not influence the accuracy of an identification, such that those in the low ego depletion, high ego depletion, and difficult task conditions would have similar sensitivity.

2) The depletion of self-regulation resources would result in lower calibration between accuracy and confidence when presented with old faces than low ego depletion and difficult task conditions.
3) Similarly, the depletion of self-regulation resources would result in lower calibration between accuracy and confidence when presented with new faces than low ego depletion and difficult task conditions.

**Method**

**Participants**

Participants for this study were recruited using the Human Participant Pool at California State University San Marcos and received class credit for their time. The number of participants was determined after running a pilot study to establish the effect sizes of self-regulation on the calibration of accuracy and confidence. Results of the pilot study established medium effect sizes for the effect of self-regulation on the calibration when recognizing old faces ($d = .68$) and on recognizing new faces ($d = .66$). Using G*Power, it was concluded that to achieve a power of .80, this experiment would need 34 participants in each condition to assess the recognition of old faces and 36 participants in each condition to assess recognition of new faces (Faul, Erdfelder, Lang, & Buchner, 2007). Three hundred and six participants were originally recruited. This was to ensure that approximately 50 participants were available for analysis in each condition, in an effort to control for a participant’s lack of false alarms or perfect hit rates. One hundred and twenty-one participants were excluded from analysis due to participant’s error, including missing responses or inappropriate responses, such as reporting a number other than “1” or “0” when making an identification or reporting a number over 10 when making a confidence judgment. Twenty-nine participants were excluded as the result of research assistant or computer error. Lastly, 15 participants were excluded due to perfect hit rates or a lack of false alarms. This was done because calculating gamma correlations for the calibration of old and new faces requires that a participant produce both hits and false alarms. As a result, 141 participants (114 women)
aged 18 to 44 ($M = 20.99, SD = 3.96$) were used for the following analyses, with 49 participants in the low ego depletion condition, 45 participants in the high ego depletion condition, and 47 participants in the difficult task condition. Participants varied on ethnicity with the plurality of participants being Caucasian (39.72%), followed by Hispanic (27.66%), Asian or Pacific Islander (21.28%), other (6.38%), and African American (4.96%) participants.

**Design and Planned Analyses**

The study was an experimental design in two parts. The independent variable manipulated was self-regulation with the three between-subjects conditions being low ego depletion, high ego depletion, and difficult task. The dependent variables were accuracy of facial recognition and confidence. To test Hypothesis 1, accuracy of facial identifications was examined between the levels of ego depletion and anagram task using an Analysis of Variance (ANOVA) to compare sensitivity measures of each participant. To test Hypotheses 2 and 3, another ANOVA test was used to compare the gamma correlations of each participant across the two levels of ego depletion and the anagram task.

**Materials**

**Demographic questionnaire.** Before starting the study, participants were asked to complete a brief demographic questionnaire. This included questions about age, gender, and ethnicity.

**Face stimuli.** Face stimuli were taken from the TarrLab face database\(^1\). This database contains photographs of both males and females across several ethnicities. Faces were chosen from this database because the TarrLab provided multiple images of the same individuals and controlled for resolution and lighting. For the purposes of this study, 30 male Caucasian face
images were chosen for their similarity and lack of distinctive features. The Appendix includes some examples of faces.

**Manipulations and manipulation check.** Ego depletion was manipulated using an automated version of the letter e task presented using E-Prime 2 (Schneider, Eschman, & Zuccolotto, 2002). In this task, all participants saw a series of words. Participants in the low ego depletion condition were asked to press the “1” key if they see the letter e in the word presented whereas those in the high ego depletion condition were told to only press the “1” key if they see a lonely e. An e is lonely when there are no vowels (a, e, i, o, u, and y) within the two letters before the e or two letters after it. All participants were told not to press any key if the corresponding letters are not there. The difficult task condition consisted of an anagram task, in which participants were asked to determine if the two ‘words’ presented on the screen were anagrams of one another by pressing the ‘1’ key if they were and pressing nothing if they were not.

**Procedure**

Participants received instructions stating that they would see a set of 15 faces and to pay close attention to these faces. Participants were then shown a set of 15 Caucasian male faces for 5 seconds each. Each face was angled either 30 degrees to the right or 30 degrees to the left. After this task, participants were randomly assigned to one of the two ego depletion conditions or the difficult task condition. Each task consisted of 150 trials lasting 7 minutes and 30 seconds. As a manipulation check, every participant was asked about the difficulty of the task and the level of exhaustion on a 5 point Likert-scale after the task was completed. Previous research has indicated that ego depletion is associated with higher levels of exhaustion and greater perceived difficulty (Hagger et al., 2010).
After the manipulation check participants took part in a recognition task. They were presented with a set of 30 Caucasian male faces, the 15 original faces and 15 new faces, and asked to identify the faces as being either old or new. These faces were counterbalanced: For half of the participants, the previously new faces were presented as the original faces. All faces in the recognition task were presented to the participant facing forward. The different angling of the faces was done in an effort to ensure that the participants were recognizing the individuals in the lineup, rather than recognizing a previously seen picture. Participants indicated seeing an ‘old’ face by pressing the ‘1’ key and a ‘new’ face by pressing the ‘0’ key. After each old and new judgment, the participants were asked how confident they were in their decision on an 11-point Likert-type scale, from 0 (not at all confident) to 10 (completely confident). When this task was completed the participants were debriefed, thanked for their time, and assigned credit.

Results

Assumptions

As ANOVAs were run to examine all hypotheses, the assumptions for this test were examined before any analysis was run. The three statistical assumptions of ANOVAS are observations within samples must be independent, homogeneity of variances, and the variables are normally distributed for each condition. All observations within the sample were independent as each participant was used for only one condition.

Homogeneity of variance was examined using Levene’s test. The results for level of exhaustion, $F(2, 138) = 2.12, p = .125$, sensitivity, $F(2, 138) = 2.09, p = .128$, calibration of old faces, $F(2, 138) = .86, p = .427$, and calibration of new faces, $F(2, 138) = .67, p = .514$, were not significant suggesting that the sample had a homogeneity of variances. The results for task
difficulty, $F(2, 138) = 3.82, p = .024$, were significant suggesting that the sample did not have homogeneity of variances.

A Kolmogorov-Smirnov test was used to examine if variables are normally distributed. The test revealed that none of the variables examined were normally distributed. After an attempt to correct this with a square change transformation failed, it was decided to proceed with the analyses with caution using the original untransformed data.

**Manipulation check**

As a manipulation check, two separate ANOVAs were run to examine the differences on task difficulty and level of exhaustion between the two levels of ego depletion and the anagram task. It was anticipated that for task difficulty those in the high ego depletion condition and anagram task condition would report higher levels of task difficulty than those in the low ego depletion condition. Additionally, it was predicted that for level of exhaustion those in the high ego depletion condition would report higher levels of exhaustion than those in the low ego depletion condition and anagram task condition. The ANOVA revealed significant differences in reports of task difficulty between the three conditions, $F(2, 138) = 93.47, p < .001, \eta^2_p = .58$. Post hoc Tukeys test showed that those in the low ego depletion condition reported significantly lower task difficulty scores than both the high ego depletion condition and the anagram task condition. However, the second ANOVA showed no significant differences in the level of exhaustion between the three conditions, $F(2, 138) = 1.10, p = .336, \eta^2_p = .02$. Implications of the failed manipulation check will be reviewed in the Discussion. Table 1 contains the means and standard deviations for the manipulation check questions.
Sensitivity

To assess if those in the high ego depletion condition were just as accurate in making identifications as those in the low ego depletion condition (Hypothesis 1), a one-way ANOVA was conducted using the self-regulation manipulations of low and high ego depletion and anagram task as the between subjects independent variable and sensitivity ($d'$) as the dependent variable. Sensitivity was calculated for each participant using signal detection theory to compare hit rates with false alarm rates. Hit rates were calculated by counting the number of times each participant correctly identified an ‘old’ face as being ‘old’ and dividing that number by the amount of potential hits (15). Similarly, false alarm rates were calculated by counting the number of time each participant identified a ‘new’ face as being ‘old’ and dividing that number by the amount of potential false alarms (15). As predicted, the ANOVA revealed no significant differences in sensitivity between the three conditions, $F(2, 138) = 1.10, p = .336, \eta_p^2 = .02$.

Table 1 contains the means and standard deviations for sensitivity.

Calibration of Old and New Faces

To assess the calibration between confidence and accuracy for the recognition of old faces (Hypothesis 2) and new faces (Hypothesis 3), two separate ANOVAs were conducted using the ego depletion condition and anagram task as the between subjects independent variable and gamma correlations of old and new faces as the dependent variables. Goodman–Kruskal gamma correlations were calculated for each participant to measure the relationship between the accuracy of the participant’s identification of an old face to the corresponding confidence judgment. This process was repeated for the calibration of new faces. Positive scores indicate the occurrence of more concordant judgments, whereas negative scores indicate an occurrence of more disconcordant judgments. The ANOVA revealed no significant difference between the
EGO DEPLETION AND CONFIDENCE-ACCURACY

three conditions for the calibration of old faces, \( F(2, 138) = 1.55, p = .215, \eta^2_p = .02 \), or for the calibration of new faces, \( F(2, 138) = 1.45, p = .237, \eta^2_p = .02 \). Table 1 contains the means and standard deviations of the gamma correlations.

**Exploratory Analyses**

While conducting the analyses for Hypotheses 2 and 3, I noticed that the overall gamma correlations for old faces appeared to be greater than with new faces. This suggests that participants’ confidence judgments are better calibrated for old faces than new faces. I tested whether these differences were statistically significant using Steiger’s (1980) recommendations for comparing independent correlations. The overall calibration between confidence and accuracy of old faces was significantly greater than the overall calibration between confidence and accuracy of new faces, \( z = 2.01, p = .045 \). This comparison was also examined separately for each condition with no significant differences found: low ego depletion, \( z = 1.50, p = .133 \); high ego depletion, \( z = 0.00, p = 1.000 \); anagram task, \( z = -1.74, p = .083 \).

**Discussion**

This study examined the effects of self-regulation on the relationship between confidence and accuracy. It was hypothesized that (1) the depletion of self-regulation resources would not influence the accuracy of an identification, such that those in all conditions would have similar sensitivity, (2) the depletion of self-regulation resources would result in low calibration between accuracy and confidence when presented with old faces and, similarly, (3) the depletion of self-regulation resources would result in low calibration between accuracy and confidence when presented with new faces.

The results from the study only showed support for the first hypothesis, that the depletion of self-regulation resources would not influence the accuracy of an identification, resulting in
similar sensitivity scores for those in the low ego depletion, high ego depletion, and difficult task conditions. However, as these results support a null hypothesis, it could be the case that this study simply did not have enough power to find support against it. Hypotheses 2 and 3 were not supported as similar calibrations were found for all three conditions in the recognition of old and new faces.

However, the overall average calibration for old faces was significantly higher than the overall average calibration for new faces. Positive gamma correlations indicated a greater amount of concordant judgments, with higher confidence scores given to accurate assessments and lower confidence scores given to inaccurate assessments. Negative correlations indicated the opposite pattern, higher confidence ratings given to inaccurate judgments and lower confidence ratings given to accurate judgments. This decrease in relationship causes problems when trying to use confidence as a postdictor of accuracy. This means that when presented with a new face, participants reported higher confidence judgments more indiscriminately, to both accurate and inaccurate identifications, than when they were presented with an old face, where participants gave higher confidence judgments to accurate identifications. The real world parallel to this situation would be when a witness identifies someone who is innocent as the culprit of a crime and reports a high confidence in that identification. As there is a positive relationship found when accurate identifications are made, law enforcement officials assume that the identification is correct as a high confidence was reported. The consequences of this decision are severe. In the best-case scenario it could result in an innocent person now having a criminal record, in the worst-case scenario, it can result in an innocent person spending the rest of their life in prison. Given that this is an inaccurate identification, using confidence as a postdictor of accuracy is problematic. These reports support similar findings that although correct
identifications have significantly greater reported confidences, incorrect identifications are also receiving higher confidence levels than anticipated (Sporer et al., 1995). The results of the exploratory analysis, along with the previous findings suggest that confidence is not an appropriate postdictor of accuracy in eyewitness identification, particularly when the face has not been previously seen. This comparison was also examined for each condition separately and there was no significant differences found between the calibration of old faces and the calibration of new faces for any of the three conditions. However, the lack of significance in these analyses could be the result of an insufficient amount of participants to find an effect.

It should be noted that although the manipulation check for task difficulty confirmed that two ego depletion conditions were perceived by the participants to be at different levels of difficulty, the manipulation check for level of exhaustion did not. One explanation for this could be that the length of the task was too long, depleting the self-regulation resources in all conditions. However, this same paradigm for ego depletion, including the duration of the task, was also used by a recent registered replication report on the subject (Hagger et al., 2016). Those researchers did find statistical differences in measures of fatigue and effort. It could be possible that the phrasing used in this study of “level of exhaustion”, as opposed to ‘effort’ or ‘fatigue’, influenced the participants reports of effort and could potentially explain these differences.

There are several explanations for the results produced in this study. First, the results suggest that self-regulation does not play a role in reporting confidences. One might argue that because the SCIF, proposed by Charman et al. (2010), is the first framework of its kind to propose a theory for how individuals assess internal cues, that it does not adequately explain how confidence judgments are assessed internally. For example, Koriat (2012) proposes the self-
consistency model of confidence, where confidence judgments are a result of an individual's attempt to be consistent. In this context, that would mean that overconfidence in identifications is a result of witnesses attempting to be consistent in their overall ability to recognize people, rather than their ability to recognize a specific individual. However, it is just as likely that self-regulation is simply not needed for the monitoring of ephorhic experiences. It is possible that this monitoring works outside of self-regulation, or that it uses so little self-regulation that the depletion of these resources does not pose a threat to making confidence judgments.

Another explanation for the lack of results could be due to the absence of an ego depletion effect altogether. Although Hagger et al.’s (2010) meta-analysis revealed a small to medium effect size, depending on the manipulation used for ego depletion, a recent registered replication report suggests otherwise (Hagger et al., 2016). Using a standardized procedure and multiple labs, these researchers found the effect to be essentially nonexistent and propose that the results found in recent literature, Hagger et al’s (2010) meta-analysis included, is likely due to publication bias. In other words, self-regulation is not a limited resource like the strength model suggests, but, because publications tend to publish significant results, one perceives an effect that is not actually there. The question is, then, what did the researchers who found support for ego depletion actually find. As stated previously ego depletion has been linked to a variety of other factors including mental fatigue, decreases in effort, positive affect, self-efficacy, and blood glucose. An argument could be made that instead of self-regulation, studies that have found support for ego depletion, actually just found support for the effects of one or more of those factors. Even in this study, those who completed the anagram task, a difficult task that was not meant to utilize self-regulation, responded similarly to those in the high ego depletion task on all measures. This similar response could mean that difficult tasks are getting at the same
mechanism as ego depleting tasks but it might also indicate that the anagram task was not 
different enough from the letter e task and could use self-regulation by inhibiting the response to 
say that the two words are anagrams of one another when they are not. This is a limitation of the 
current study.

A second limitation of this study is the high omission rate of participants due to missing 
or incorrect data reported on the part of the participant. Participants were excluded from analysis 
if they did not report whether they believe a face was old or new, or if they did not provide a 
confidence judgment. There was no pattern for the missing data on a particular face or condition 
so the high rate of missing data might indicate that the identification and judgment task overall 
was too difficult. However, these results could also be due to the way that the E-Prime program 
was constructed in that the participants were not limited to the number key and did not allow for 
participants to go back and change errors. Another limitation of this study is the use of only 
young adult male Caucasian faces for identification. These faces were used as a way to ensure 
consistency across identifications. However, several studies have supported the idea that 
individuals are more accurate in identifying individuals of their own race, gender, and age 
(Messiner & Brigham, 2001; Palmer, Brewer, & Horry, 2013; Wright & Stroud, 2002). Given 
that this study did not limit its participants to male Caucasian participants, it is possible that the 
accuracy in identifications might have been influenced by the use of these faces. An additional 
limitation of this study is the artificiality of the face recognition paradigm. Although this study 
used a typical facial recognition paradigm, this is not typical of an eyewitness identification 
scenario. In typical eyewitness studies, the participants would witness a crime, such as a culprit 
robbing a house, and then make the identification of the culprit. This model most accurately 
reflects how a witness would experience an eyewitness identification. Although the current
study allowed for the analysis of multiple identifications, it does not accurately reflect a real world eyewitness identification. The final limitation of this study is the lack of normality within the examined variables. The nature of the data, which included negative values and zeros, made the use of traditional transformations inappropriate. Continuing with the analyses while failing to meet an assumption for that analysis could mean that there was an effect present but due to the negatively skewed nature of the data, that went unfound.

**Conclusion and Future Directions**

Results of this study have expanded on the current literature on self-regulation and the use of confidence judgments as a reference for accuracy. Because confidence in an identification is seen as a measure of accuracy it is important to understand what factors can influence that judgment.

Although this study did not find evidence for the influence of self-regulation on confidence judgments, it did provide some interesting data on the relationship between confidence and accuracy regardless of self-regulation. It was found that the overall calibration between confidence and accuracy in the recognition of old faces was significantly higher than the overall calibration for the recognition of new faces. The lower calibration found for the recognition of new faces suggests that there was a lower rate of concordant judgments made (i.e., that fewer participants gave high confidence ratings to accurate identifications and low confidence rating to inaccurate identifications). This supports the idea that confidence is not an appropriate measure of accuracy, regardless of the calibration for the identification of old faces being higher than that of the new faces, as the recognition of new faces as old faces is synonymous to a witness identifying an innocent person as the culprit of a crime. At the very
least, this could provide additional incentive to continue research examining the relationship between confidence and accuracy.

As the manipulation for ego depletion in this study might not have adequately paralleled how a witness might experience depleted self-regulation, it would be interesting to further investigate self-regulation’s effect on confidence judgments using a manipulation such as emotion regulation. Future research might also examine the role of self-regulation on the relationship between confidence and accuracy outside of the ego depletion paradigm. It would be beneficial to see if the results found in this study differed when using a more typical eyewitness identification paradigm or in a study that varied the ethnicity, age, or gender of the target.
References


doi:10.1016/j.jesp.2012.03.007

1Stimulus images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University, http://www.tarrlab.org/. Funding provided by NSF award 0339122.
A witness attempts to assess confidence

Assess internal cues

Internal cues are strong

Report confidence

Internal cues are weak

Cue present that disconfirms one’s identification

Search for external cues

Cue not integrated into confidence assessment

Cue present that confirms one’s identification

Information available that undermines credibility of cue

Evaluate credibility of the cue

No information available that undermines credibility of cue

Cue not integrated into confidence assessment and confidence is

Cue integrated into confidence assessment and confidence is

Figure 1. The selective cue integration framework (Charman et al., 2010)
Table 1

*Means and Standard Deviations for Experimental Variables and Manipulation Check*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Low Ego Depletion ((n = 49))</th>
<th>High Ego Depletion ((n = 45))</th>
<th>Anagram Task ((n = 47))</th>
<th>Overall ((N = 141))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(M)</td>
<td>(SD)</td>
<td>(M)</td>
<td>(SD)</td>
</tr>
<tr>
<td>Task Difficulty</td>
<td>1.33</td>
<td>.55</td>
<td>3.20</td>
<td>.89</td>
</tr>
<tr>
<td>Level of Exhaustion</td>
<td>2.53</td>
<td>1.21</td>
<td>2.80</td>
<td>1.01</td>
</tr>
<tr>
<td>Hit Rate</td>
<td>.69</td>
<td>.16</td>
<td>.62</td>
<td>.16</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>1.31</td>
<td>.53</td>
<td>1.17</td>
<td>.72</td>
</tr>
<tr>
<td>Calibration of New Faces</td>
<td>.12</td>
<td>.61</td>
<td>.30</td>
<td>.55</td>
</tr>
<tr>
<td>Calibration of Old Faces</td>
<td>.41</td>
<td>.49</td>
<td>.30</td>
<td>.46</td>
</tr>
</tbody>
</table>
Appendix

Faces presented at learning phase:

Faces presented at recognition phase: