2013-01-01

An Intra-Individual Event-Related Potential-Based Concealed Attitude Test

David R. Herring
University of Texas at El Paso, david.herring41@gmail.com

AN INTRA-INDIVIDUAL EVENT-RELATED POTENTIAL (ERP)-BASED CONCEALED ATTITUDE TEST

David R. Herring

Department of Psychology

APPROVED:

Stephen L. Crites Jr., Ph.D., Chair

John J. B. Allen, Ph.D.

Lawrence D. Cohn, Ph. D.

Christian A. Meissner, Ph. D.

Osvaldo F. Morera, Ph. D.

Benjamin C. Flores, Ph.D.
Dean of the Graduate School

Dedication

To Fred, for inspiring me to work hard.

AN INTRA-INDIVIDUAL EVENT-RELATED POTENTIAL (ERP)-BASED CONCEALED ATTITUDE TEST

by

David R. Herring, M.A.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

THE UNIVERSITY OF TEXAS AT EL PASO

December 2013

Acknowledgements

Thank you, Steve, for wonderful mentorship and being a friend. I will miss having discussions of Star Trek on a regular basis. Thanks John, Larry, Chris, and Ozzie for your suggestions and making me better. Many thanks to my labmates over the years—Lupe, Jen, Kat, Linsa, and Ciara—for such wonderful support and input. Last, thank you Kasey, Dean, Gage, Bentley, and Storie for your love, support, and making me smile.

Abstract

The evaluative oddball is an implicit measure for detecting (concealed) attitudes. In evaluative oddball studies, low occurrence stimuli such as negative pictures are presented among high occurrence context stimuli such as positive pictures. Late positive potentials (LPPs) of the event-related potential (ERP) are larger to evaluatively incongruent stimuli such as negatives compared to congruent stimuli such as positives with the context (e.g., positives). In prior evaluative oddball paradigms, this evaluative congruity effect of the LPP was reduced when participants concealed compared to truthfully reported attitudes. Because prior evaluative oddballs have been focused on the *group level* analysis, it has been overlooked how this reduced evaluative congruity effect for concealing might affect the *individual level* classification of attitudes. The evaluative oddball paradigm was extended by using a Bayesian classification approach to determine whether the evaluative oddball task influenced classification of concealment trials. Fifty-two participants performed an evaluative oddball task in which they either counted or made key presses to evaluatively incongruent pictures. Altogether, evaluatively incongruent pictures were correctly classified at 76.9%. There was no difference between counting (19/25 or 76%) and key pressing (19/27 or 70%) classification rates of participants' attitudes during concealment trials. During truth telling, there was a non-significant effect for a larger percentage of participants' attitudes classified for key pressing (24/27 or 89%) than counting (18/25 or 72%). These data suggest that this Concealed Attitude Test (CAT) is promising for detecting (concealed) attitudes at the individual level using an LPP-based Bayesian approach.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Detecting evaluations (i.e., good/bad assessments[1] of objects) is noteworthy in several situations. A company may want to know the extent of an employee's bias toward certain groups (e.g., gay men and lesbians) because this bias potential *may be* associated with promoting employees unfairly. A security agency (e.g., NSA) may wish to know which employees are patriots based on positive attitudes toward their country, which *could be* associated with being less of a security risk (e.g., leaking information). In both situations, people may try to conceal their true attitudes if they are inconsistent with the agency's agenda or social norms. Several efforts over the past decades have been made to devise measures that can sidestep misreporting evaluations by obtaining implicit measures that rely on psychophysiological or (e.g., EEG) or behavioral measures (e.g., reaction times; Cameron, Brown-Iannuzzi, & Payne, 2012; Greenwald, Andrew, Uhlmann, & Banaji, 2009; Herring et al., in press). However, much of the research has focused on basic theoretical issues. To date, little is known about the accuracy of implicit tests of evaluation. The objective of the current investigation was to develop an implicit technique for detecting concealed attitudes and to use a Bayesian classification scheme to determine its accuracy at the individual level.

There are two prominent approaches to psychophysiological detection. One approach to psychophysiological detection is the Control Question Test (CQT; Iacono, 2007; Lykken, 1998; Raskin & Honts, 2002), which is commonly referred to as the polygraph. The CQT has traditionally involved measuring autonomic nervous system (ANS; i.e., cardiovascular, respiratory, and electrodermal responses) responses to control and relevant questions. Control questions such as "Have you ever taken something that didn't belong to you?" are minor transgressions that most people have committed. On

[1] *Evaluation* refers to the operations of assessing the goodness of an object (Cacioppo & Berntson, 1994a; Herring et al., in press). Our use of *affect* is consonant with emotion literature as a state along a positive/negative continuum that also ranges along a sleepy/excited continuum (i.e., Bradley & Lang, 2007; Russell & Feldman Barrett, 1999).

the other hand, relevant questions, such as "Did you take Mr. Waddams' Swingline stapler?" are specific to the investigation. Examinees are led to believe before the CQT that they should respond “no” to both question types. For innocent people, CQT theory predicts larger fear responses to control questions compared to relevant questions because relevant questions are meaningless to innocent people and should thus produce minimal physiological responding. The converse is expected for guilty people—larger fear responses for relevant relative to control questions should occur because the relevant questions are more meaningful. Despite law agencies in the U.S. relying on the technique, few psychophysiologists support the CQT (Ben-Shakhar, 2002; Iacono & Lykken, 1997; National Research Counsel, 2003).

For one thing, the CQT has questionable validity. The hypothesis that fear or anxiety are associated with a unique autonomic nervous system signature is controversial and has not been confirmed (Russell, 2003). Multiple meta-analyses have failed to demonstrate clearly dissociable physiological patterns of fear from other emotions (Cacioppo, Berntson, Klein, & Poehlmann, 1998; Lench, Flores, & Bench, 2011; Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012). Further, important properties of the CQT such as estimates of reliability and accuracy remain elusive (National Research Counsel, 2003). Assessing reliability is made challenging by the testing procedures of the polygraph (e.g., pre-polygraph interviews, chart interpretation, etc.) introducing variability into the outcome of the polygraph. Assessing accuracy is made challenging by the varying thresholds polygraphers set, which determines the trade-off between the percentage of truly deceptive examinees detected (i.e., the sensitivity) and truly honest examinees detected (i.e., specificity).

The second method of psychophysiological detection is the Concealed Information Test (CIT; Ben-Shakhar & Elaad, 2002; Ben-Shakhar, 2012; Verschuere, Ben-Shakhar, & Meijer, 2011). The principal difference between the CQT and CIT is that the former is a test of deception whereas the latter is a test of information typically someone guilty would know (e.g., a murder weapon). Whereas the

CQT is not theoretically sound, the CIT is based on the well-established psychophysiological principle of orienting—heightened physiology (e.g., increased skin conductance) to novel or significant stimuli (Sokolov, 1963; Stern, Ray, & Quigley, 2001). In a typical CIT, participants' psychophysiological responses are recorded while questions pertaining to either probe or context (irrelevant) items are asked. Probes are information relevant to the case typically someone guilty would know (e.g., a murder weapon) whereas context items are insignificant to all examinees. For instance, if someone was stabbed and killed with a butcher knife a series of questions could be asked: "Was Mr. Smith (a) beat with a bat (context item)…(b) beat with a hammer (context item)…(e) stabbed with a butcher knife (probe item)?" Alternatively, individual pictures of weapons could be presented followed occasionally by a probe—"bat"… "gun"… "butcher knife" (probe).

Researchers initially adapted the CIT for autonomic nervous system-based recording. In a classic study, Lykken (1959) had some participants perform mock-crimes (i.e., "guilty" condition), one of which involved stealing an item from a desk and hiding it in a locker. Other participants did not perform the mock-crimes (i.e., "innocent" condition). During the CIT, participants were asked where the item was hidden ("Was it (a) in the men's room…(e) in the locker?") while skin conductance was measured. Because only "guilty" participants performed the theft, these participants' skin conductance responses were larger to question "e" because this item was meaningful but items "a" through "d" were not. Innocent participants, however, did not orient to question "e" and had similar skin conductance responses to all choices. Lykken's classification scheme of differentiating guilty from innocent participants yielded 94% accuracy (see Matsuda, Nittono, & Allen, 2012 for review on classification methods). Several meta-analyses support the validity and reliability of the CIT with autonomic nervous system-based approaches. For example, larger physiological responses to probe relative to context stimuli (i.e., CIT effect) range for skin conductance from 1.5-1.7 (*d*) translating to *r*s from .55-.58 (Ben-Shakhar & Elaad, 2003; Meijer & Ben-Shakhar, 2012). Whereas, the best estimate of the CQT is an

area under the curve (A) ranging from .81 to .91 (note these are not overall accuracy estimates; National Research Counsel, 2003; see also Gougler et al., 2011; Kircher, Horowitz, & Raskin, 1988), which can be difficult to interpret. Nevertheless, several studies have demonstrated that autonomic nervous system-based CITs are susceptible to mental and physical strategies people use to distort the test (i.e., countermeasures; Elaad & Ben-Shakhar, 1991). Researchers then turned to CITs using event-related potentials (ERPs) with the hope this might be less susceptible to countermeasures (Allen, Iacono, & Danielson, 1992; Farwell & Donchin, 1991; Rosenfeld, Angell, Johnson, & Qian, 1991; cf. Mertens & Allen, 2008; Rosenfeld, Soskins, Bosh, & Ryan, 2004).

ERPs are scalp-recorded electrical potentials tied to events embedded in the electroencephalogram (EEG) that reflect different psychological operations (Fabiani, Gratton, & Federmeier, 2007; Kappenman & Luck, 2012; Luck, 2005). One could, for instance, present probes and context items separately and track the ERPs that unfold after presentation of these stimuli on the order of milliseconds. When conceptualizing these psychological operations (e.g., attention, categorization, response inhibition) it is more accurate to refer to ERPs as latent *components*—cognitive operations associated with specific neuroanatomical modules (Luck, 2005). In contrast, ERP *waveforms* are the superficial peaks and troughs with negative or positive polarity. The virtue of ERPs is that they have excellent temporal resolution compared to skin conductance and functional magnetic resonance imaging (fMRI). Another advantage of ERPs is that they offer incremental validity to other psychophysiological measures in CITs. For instance, Matsuda and colleagues (2011) found that five out of the seven participants who were missed with an autonomic-based CIT were detected with the addition of an ERP based assessment. Further, meta-analysis reveals the ERP-based CIT effect is significantly larger (2.55) than autonomic nervous system-based CITs, which could enhance individual level classification rates for the ERP-based CIT (Meijer & Ben-Shakhar, 2012).

Most ERP-based CITs rely on the P3 component of the ERP (Donchin & Coles, 1988; Donchin, 1981; Johnson, 1986; Nieuwenhuis, Aston-Jones, & Cohen, 2005; Picton, 1992; Polich, 2012; Pritchard, 1981). The P3 is typically investigated using an oddball paradigm in which low probability oddball stimuli occasionally appear among high probability context stimuli. For instance, a simple oddball paradigm might consist of presenting "Os" 20% of the time and "Xs" 80% of the time (e.g., X…X…X…X…O). A robust finding is that the P3 amplitude is larger to oddballs relative to context stimuli (i.e., oddball effect; Nieuwenhuis et al., 2005). In the ERP-based CIT, probes are presented infrequently and context items are presented often. Further, remembered stimuli are associated with larger P3 amplitudes than forgotten stimuli (Fabiani, Karis, & Donchin, 1986). That is, if participants perform an oddball and are unexpectedly asked to recall as many stimuli from the oddball paradigm as possible, P3s to recalled stimuli are larger than P3s to forgotten stimuli. In the ERP-based CIT, probes are usually learned by performing a mock crime and the context items are novel items. Thus, because probes are rare and learned they should produce larger P3s than frequent, novel context stimuli.

Rosenfeld and colleagues (1987) adapted the CIT for using ERPs (see also; Johnson & Rosenfeld, 1992; Rosenfeld et al., 1991; Rosenfeld, 2011). "Guilty" participants were presented with a number of items (e.g., pearls, ring, etc.), chose one of the items to write about, and then pretended during an ERP-based CIT that the chosen item was something they stole and silently concealed that item. "Innocent" participants went through the same procedure but never saw the chosen item during the CIT. This P3-based CIT was an oddball paradigm in which the chosen item (i.e., probe) was the oddball and novel items not presented earlier were the context. For guilty participants the P3 amplitude was significantly larger for probes relative to the context; no significant difference was found between probes and the context for innocent participants. This initial study was important for demonstrating in principle that the P3 could be used for detecting concealed information at the *group level*. Subsequent early investigations extended this study by developing *individual* classification techniques yielding

impressive accuracy ranging from 88% (Rosenfeld et al., 1991) to 94% (Allen et al., 1992; see also Farwell & Donchin, 1991). A shortcoming of the P3-based CIT is that it has limited application because it is most useful for detecting information about past but not future behavior. Other forms of knowledge such as attitude might be more useful to this end.

Attitudes are evaluative summaries (good/bad) of objects (Cacioppo & Berntson, 1994b; Eagly & Chaiken, 1993; Fazio, 2007). Notably, attitude is a potentially useful construct for predicting behavior because attitudes predict behavioral intentions that in turn predict behavior (Ajzen, 1991, 2011). For instance, someone with a strong negative attitude toward the U.S. likely has a negative behavioral intention toward the U.S. (e.g., wanting to be uninvolved in U.S. politics) that could lead to actual negative behavior (e.g., not voting for government officials). Traditionally, attitudes have been assessed using self-report; these measures involve objects (e.g., the U.S.) being assessed typically with a bipolar scale (e.g., -3 to +3) in which higher values indicate more favorable attitudes (Krosnick, Judd, & Wittenbrink, 2005). Self-report measurement of attitudes, however, is ineffective when people misreport extreme attitudes (e.g., negative attitudes toward the U.S.) because of recognition that these attitudes go against social norms.

Implicit measure of assessing attitudes, such as the evaluative oddball paradigm, can be used to detect extreme attitudes people wish to misreport (Cacioppo, Crites, Berntson, & Coles, 1993; Cacioppo, Crites, Gardner, & Berntson, 1994; Ito, Larsen, Smith, & Cacioppo, 1998). In an evaluative oddball paradigm, attitude object words or pictures are presented individually similar to the ERP-based CIT. Oddballs are attitude objects presented infrequently (e.g., 20%) that are evaluatively *incongruent* (e.g., negatives) with a context of attitude objects (e.g., positives) presented frequently (e.g., 80%). The attitude stimuli are objects commonly associated with valence (e.g., snake = negative), or are idiosyncratic stimuli associated with valence (e.g., Jon Stewart = positive). Either way, these are objects that participants should have knowledge of with respect to valence. For instance, Crites and colleagues

used an evaluative oddball paradigm with normed trait adjectives (Crites, Cacioppo, Gardner, & Berntson, 1995). Participants viewed trait adjectives in either a positive or negative context—e.g., “honest”, “sincere”, “calm”, “orderly”, “understanding”, “loyal”—followed by either evaluatively congruent (“thoughtful”) or incongruent (“liar”) trait adjectives. Because they were rare and stored in memory, evaluatively incongruent stimuli elicited larger P3-like late positive potentials (LPPs) relative to evaluatively congruent adjectives (i.e., evaluative incongruity effect; Cacioppo et al., 1993; Cacioppo et al., 1994).

A few evaluative oddball studies have involved participants misreporting some attitudes while truthfully reporting other attitudes (Cacioppo, Crites, & Gardner, 1996; Crites et al., 1995; Crites, Mojica, Corral, & Taylor, 2010; Güereca et al., 2010). For example, even though participants may have known a trait (e.g., “honest”) was associated with a positive connotation, they reported this stimulus as negative. In the Crites et al. (1995) study, whether participants were truth telling or concealing did not moderate the evaluative congruity effect. In follow-up studies, more complex attitudes toward people (vs. traits) were assessed using pictures (e.g., Tony Romo, Jon Stewart, Bill O’Reilly; Crites et al., 2010; Güereca et al., 2010). The evaluative congruity effect (i.e., larger LPPs to evaluatively incongruent than congruent trials) was smaller when participants concealed relative to truth telling. For instance, in Güereca et al. (2010), the evaluative congruity effect (i.e., *d*) was 1.55 and .57 for truth telling and concealment trials, respectively. Similar diminished effects on the LPP have been found in the CIT literature when participants use countermeasures (Bergström, Anderson, Buda, Simons, & Richardson-Klavehn, 2013). The diminished evaluative congruity effect to concealment presumably occurred because of the cognitive load of misreporting attitudes during these trials (Crites et al., 2010; Johnson, Henkell, Simon, & Zhu, 2008). That is, misreporting attitudes divides attention between primarily evaluating stimuli and misreporting as a secondary task, and it is well established that secondary tasks diminish the LPP (Nieuwenhuis et al., 2005). These findings are also consonant with the hypothesis that

focused attention on the evaluative dimension is integral for quickly and strongly activating attitudes (i.e., attentional sentizing; Herring et al., in press; Kiefer, Adams, & Zovko, 2012; Spruyt, Houwer, & Hermans, 2009).

The reduction of an oddball effect at the *group level* is noteworthy because it may lead to diminished accuracy at the *individual level*. For example, Mertens and Allen (2008) conducted a CIT in which guilty participants responded typically (i.e., to target stimuli requiring a response) or used countermeasures to make non-probe stimuli salient (e.g., imagining being slapped or tightening the sphincter muscle to context stimuli). If context stimuli were more salient or became more task-relevant then oddball stimuli, P3s should be enhanced for context stimuli. Thus, discriminating between contexts and probes would be more difficult. At the group level, typical guilty participants displayed a significant CIT effect whereas none of the countermeasure groups displayed a CIT effect. These different group effects translated to more accurate individual classification using a Bayesian approach for the typical relative to countermeasure guilty participants. No individual assessments have been conducted on evaluative oddball paradigms. Nevertheless, provided that diminished CIT effects translate to diminished individual level effects, it is tenable that diminished evaluative congruity effects by misreporting translate to diminished individual classifications. With this in mind, in the current study modifications were made that are described below to enhance the group level effects in an evaluative oddball paradigm in an effort to optimize individual level classification.

If concealing attitudes leads to diminished individual classification compared to truth telling, then it is important to develop an evaluative oddball paradigm using different techniques that enhance the evaluative congruity effect. Given misreporting attitudes diminishes the evaluative congruity effect, one alteration to an evaluative oddball paradigm that would free attention is to have participants conceal by withholding responses instead of misreporting (e.g., Ambach, Bursch, Stark, & Vaitl, 2010; Matsuda, Nittono, & Ogawa, 2012). For instance, participants could conceal by merely withholding responses to

attitude objects. An issue with this approach is that truth telling should still yield a larger evaluative congruity effect because stimuli requiring responses elicit larger LPPs (Duncan-Johnson & Donchin, 1977). This is noteworthy since truth telling and concealment are the key comparison groups in evaluative oddball paradigms. Nevertheless, replacing misreporting with withholding responses may significantly enhance the evaluative congruity effect compared to previous evaluative oddballs (i.e., d = .57; Güereca et al., 2010) by removing the cognitive load associated with misreporting. In the present study, we developed an evaluative oddball paradigm by replacing misreporting instructions from previous studies with conceal instructions that directed participants to make no response. In this way, we made it more likely that the evaluative congruity effect for concealing would be enhanced at the group level and thus classification would be enhanced at the individual level.

Another modification to an evaluative oddball paradigm that could enhance the evaluative congruity effect during concealment trials is to switch from key pressing to counting stimuli. In non-evaluative oddball tasks participants typically count oddball trials because the oddball effect is larger than counting context stimuli. For instance, participants could see predominately “Xs” and occasionally “Os” and would be required to mentally count the “Os” in each block and report the number of “Os” and the end of each block. Several non-evaluative oddball studies have demonstrated that the oddball effect is enhanced for counting relative to key pressing toward oddball stimuli because counting requires a constant mental update drawing more attention toward the task (Barrett, Neshige, & Shibasaki, 1987; Polich, 1987; Salisbury, Rutherford, Shenton, & McCarley, 2001). On the other hand, key pressing, which has been predominately used in evaluative oddball paradigms, captures attention to a lesser extent since there is no requirement that keeps participants mentally updating such as counting. That is, with key pressing participants simply make a quick key press and have “down time” in between stimuli in which the task is momentarily off the mind. In the present study, participants either counted or key

pressed toward incongruent trials to determine whether counting led to an enhanced evaluative congruity effect and thus potentially greater individual classification.

There is one methodological consideration when comparing counting and key pressing conditions in oddball paradigms. In non-evaluative oddball studies, counting and key pressing tasks were similarly directed at oddballs. Because these tasks were not different—e.g., one task directed at both oddball *and* context stimuli and another only at oddballs (Acosta & Nasman, 1992)—comparison between these response tasks was fair. In previous evaluative oddball paradigms, participants used key pressing to indicate or misreport their attitudes to both context and probe stimuli (i.e., "good" key vs. "bad" key). Thus, to compare key pressing and counting in the present evaluative oddball paradigm required the removal of key presses to context stimuli so a fair comparison could be made between counting and key pressing conditions.

1.1 The Concealed Attitude Test

The purpose of the present study was to develop a Concealed Attitude Test (CAT). Participants performed an evaluative oddball paradigm in which the context was pictures from the International Affective Picture System (IAPS)—normed ratings from the IAPS were used so the same context stimuli could be used across participants. The context valence alternated from block to block; for instance, if the first block was of positive IAPS then the second block would be negative. Other pictures from the IAPS and a celebrity set (e.g., athletes, politicians, etc.) that participants rated beforehand were also presented during the evaluative oddball paradigm. These pictures were either evaluatively congruent or incongruent with the context. Participants were truthful toward half of these pictures (*controls*) and withheld responses toward the other half of pictures (*probes*). To prevent participants from learning to associate responses (i.e., counting or key pressing) to pictures they were always truthful toward, "catch" stimuli were drawn from whichever context set was not in use. This procedure made it less likely

participants would lookout for their idiosyncratically rated stimuli and thus kept them on the task of evaluating stimuli as either positive or negative. For instance, if the context was positive pictures, then the "catch" trials were pictures from the negative IAPS context stimulus set. Participants were randomly assigned to either count or make a key press toward evaluatively incongruent pictures.

The design of the current CAT enabled a couple of important extensions of prior evaluative oddball paradigms. First, because several non-evaluative oddball studies indicate that counting enhances the oddball effect relative to key pressing (Barrett et al., 1987; Polich, 1987; Salisbury et al., 2001), we randomly assigned participants to either perform counting or key pressing to evaluatively incongruent pictures. This extends prior evaluative oddball paradigms that have relied on just key pressing as the response task. Counting and key pressing occurred only to evaluatively incongruent pictures during truth telling and catch trials because as mentioned we replaced misreporting of probes with concealing via making no response. Counting and key pressing toward evaluatively incongruent truthful trials made comparisons fair between the two tasks since counting is typically performed on just oddballs. The primary question examined in the present work was whether enhanced attention by counting during truth telling would transfer to probe trials. Specifically, does this transfer of attention from truth telling to concealment during counting evoke a larger evaluative congruity effect compared to key pressing?

Second, this CAT was the first evaluative oddball paradigm to make individual classification using the Bayesian approach from the CIT literature (Allen et al., 1992; Allen & Iacono, 1997; Allen, 2002). Specifically, using Bayes' rule we were able to assess the probability that a given participant's attitude was incongruent provided a combination of LPP indicators. We thus extended previous evaluative oddballs by comparing individual classification between probes and controls, as well as comparing individual classification between response tasks (counting vs. key pressing) for probe trials.

1.2 Group-level Hypotheses

Hypothesis 1. Across truth telling and concealing, evaluatively incongruent compared to congruent pictures would evoke larger LPPs (i.e., *evaluative congruity effect*; Cacioppo et al., 1993, 1996, 1994; Crites et al., 1995, 2010; Ito et al., 1998).

Hypothesis 2. Because stimuli that require responses elicit larger incongruity effects on LPPs than stimuli of the same probability not requiring responses, the evaluative congruity effect should be larger during truth telling compared to probe trials (Duncan-Johnson & Donchin, 1977).

Hypothesis 3. We anticipated that the evaluative congruity effect would be larger for counting than key pressing based on the non-evaluative oddball literature (Barrett et al., 1987; Polich, 1987; Salisbury et al., 2001). We were particularly interested in this effect of task on evaluative congruity during concealment. One might expect that the enhancement of evaluative congruity via of counting would only occur during truthful trials because these trials are task relevant. On the contrary, we reasoned that the sustained attention from counting evaluatively incongruent pictures during truth telling would spill over to concealment trials, thus enhancing the evaluative congruity effect for counting relative to key pressing during concealment trials (Barrett et al., 1987; Polich, 1987; Salisbury et al., 2001). We reasoned that counting from the truth telling trials would not merely "turn off" rendering the concealment trials unaffected—participants would still be attuned to the evaluative dimension during concealment because they were constantly evaluating valence by making mental updates of their counts.

1.3 Individual-level Hypotheses

We extended previous evaluative oddball paradigms by determining whether individual classification rates were hampered for probes relative to truth telling across response tasks.

Additionally, we examined whether response task (key pressing vs. counting) would influence classification for probe trials. For these analyses accuracy was determined using a Bayesian approach (Allen et al., 1992).

Chapter 2: Method

2.1 Participants

Fifty-two participants (25 females; *M* = 22.42 [*SD* = 6.47] yrs. old) were included in the analysis. Participants were recruited from an online psychology participant pool and via flyers posted throughout a southwestern campus. Participants received either course credit or $30 for online pool and flyer recruitment, respectively. All participants had normal or corrected-to-normal vision and eight participants were left-handed (Oldfield, 1971). Data from seven participants were discarded because of poor behavioral performance (*n* = 1; i.e., accuracy 2.5 *SD*s below the grand mean), experimenter error (*n* = 2; e.g., an incorrect response device was used), or high artifact rejection rates (*n* = 4). An additional six participants were recruited but did not complete the study because they took 45 minutes or greater to perform the individual picture selection phase (see below).

2.2 Stimuli

Context/catch stimuli. Context stimuli were 84 pictures (42 positive, 42 negative; 999 X 1350 pixels) from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008). The IAPS is a set of 1,196 scenes varying in semantic categories (e.g., people, animals, and environments) and affective content (e.g., pleasant, unpleasant, and neutral images). The IAPS was normed on valence (1 = *unhappy*, 9 = *happy*) and arousal (1 = *unaroused*, 9 = *aroused*) dimensions with the nonverbal self-assessment manikin (SAM; Bradley & Lang, 1994). From the 84 IAPS, separate IAPS sets were constructed for male and female participants given gender differences rating the IAPS—half (42) of the pictures were constructed using normed female ratings while the other half were constructed using normed male ratings (Lang & Bradley, 2007). These same context stimuli were also used as “catch” stimuli during the experimental task.

To confirm construction of the IAPS sets, separate two-way ANOVAs (gender X picture type: positive vs. negative) were run on valence and arousal ratings. Positive pictures (*M* = 7.50, *SD* = .42)

were significantly more pleasant than negative pictures (*M* = 2.86, *SD* = .52), *F* (1, 81) = 1,839.22, *p* < .001, and these valence ratings did not differ between female (*M* = 5.19, *SD* = 2.31) and male (*M* = 5.17, *SD* = 2.47) stimulus sets, *F* (1, 81) = .01, *p* = .90. Further, the positive (*M* = 5.65, *SD* = .68) and negative (*M* = 5.58, *SD* = .71) pictures did not differ on arousal ratings, *F* (1, 81) = .21, *p* = .65, and these arousal ratings did not differ between male (*M* = 5.68, *SD* = .73) and female (*M* = 5.54, *SD* = .66) stimulus sets, *F* (1, 81) = .82, *p* = .36.

Idiosyncratic stimuli. Eight pictures (4 positive, 4 negative; 999 X 1350 pixels) were selected using participants' idiosyncratic assessments. Four of the pictures were selected from a set of celebrities (hereafter "celebrity set; Corral, 2009; Crites et al., 2010; Güereca et al., 2010; Taylor, 2010). The celebrity set is a stimulus pool of 649 colored pictures of famous people (e.g., athletes, politicians, musicians) gathered from various search engines. Each picture was a portrait selected with the criteria that 1) there were no distracting features (e.g., atypical hairstyles and backgrounds) and 2) there were no strong facial expressions (e.g., prototypical emotional expressions such as anger or disgust). The other four pictures were from the IAPS. The difference between these IAPS and the context IAPS (described above) is that participants rated these IAPS, whereas for the context stimuli normed ratings were used. To avoid confusion, hereafter we refer to the context IAPS as "normed-IAPS" and the individually rated IAPS as "idio-IAPS".

2.3 Procedure

Overview. After informed consent, participants completed a brief questionnaire regarding demographics and handedness (Oldfield, 1971). Next, participants completed the idiosyncratic picture selection phase and upon completion were prepared for electroencephalography (EEG; 15 kΩ or less). While participants were prepped for EEG, an E-Prime script for the experimental session was modified to include participants' ratings from the idiosyncratic picture selection phase. Participants were then

taken to an isolated chamber to complete the experimental session. Participants viewed high occurrence (i.e., 83.3% probable) context stimuli (e.g., positives) and occasionally (16.6%) viewed idiosyncratic pictures either evaluatively incongruent (e.g., negatives) or congruent (e.g., positives) with context pictures across four discrete segments of trials (i.e., "blocks"). These low occurrence (non-context stimuli) were comprised of idio-IAPS, celebrities, and "catch" stimuli. "Catch" stimuli, which were context stimuli from the valence not presented in a current block (e.g., negatives if the context was positive) were also presented so participants would not pair their responses with idiosyncratic pictures. In each block, participants' task was to count or press a key (between-subjects) to two evaluatively incongruent pictures (e.g., one idio-IAPS and one celebrity; truthful *controls*); for the other two evaluatively incongruent pictures (e.g., one idio-IAPS and one celebrity; *probes*) participants concealed their attitudes by not responding (i.e., neither counting nor key pressing). Once the CAT was completed participants revealed their attitudes toward the probes, completed a brief post-experimental survey rating task difficulty ("How difficult was this experiment?"; 0 = *extremely easy*, 8 = *extremely difficult*), and were debriefed (see Figure 1).

Idiosyncratic picture selection phase. The idiosyncratic picture selection phase consisted of two parts in the following order: 1) the control stimulus set ratings and 2) the probe stimulus set ratings. For both assessments, participants viewed IAPS and celebrity pictures presented with Psychtoolbox 3.0.10 (Kleiner et al., 2007) on a 22" monitor (60 Hz refresh rate). Instructions for rating the IAPS and the celebrity pictures followed the IAPS recommendations[2]. Pictures were presented similarly for the

[2] The celebrity set has been rated (Corral, 2009; Crites, Mojica, Corral, & Taylor, 2010; Güereca et al., 2010) using a general evaluative 7-point bipolar scale (-3 to +3) from the attitude literature (Crites, Fabrigar, & Petty, 1994; Krosnick, Judd, & Wittenbrink, 2005; Osgood, Suci, & Tannenbaum, 1957). Nonetheless, there is evidence that participants can rate non-affective celebrities with the SAM (Gobbini, Leibenluft, Santiago, & Haxby, 2004). Further, the semantic differential, that our general bipolar scale is based on, is strongly correlated with the SAM (*r*s > .90; Bradley & Lang, 1994; Lang & Bradley, 2007).

control and probe assessments—on each trial one picture was randomly drawn from the IAPS or celebrity set (6000 ms; Bradley & Lang, 1994; see Figure 1 for details)[3]. Participants either made an

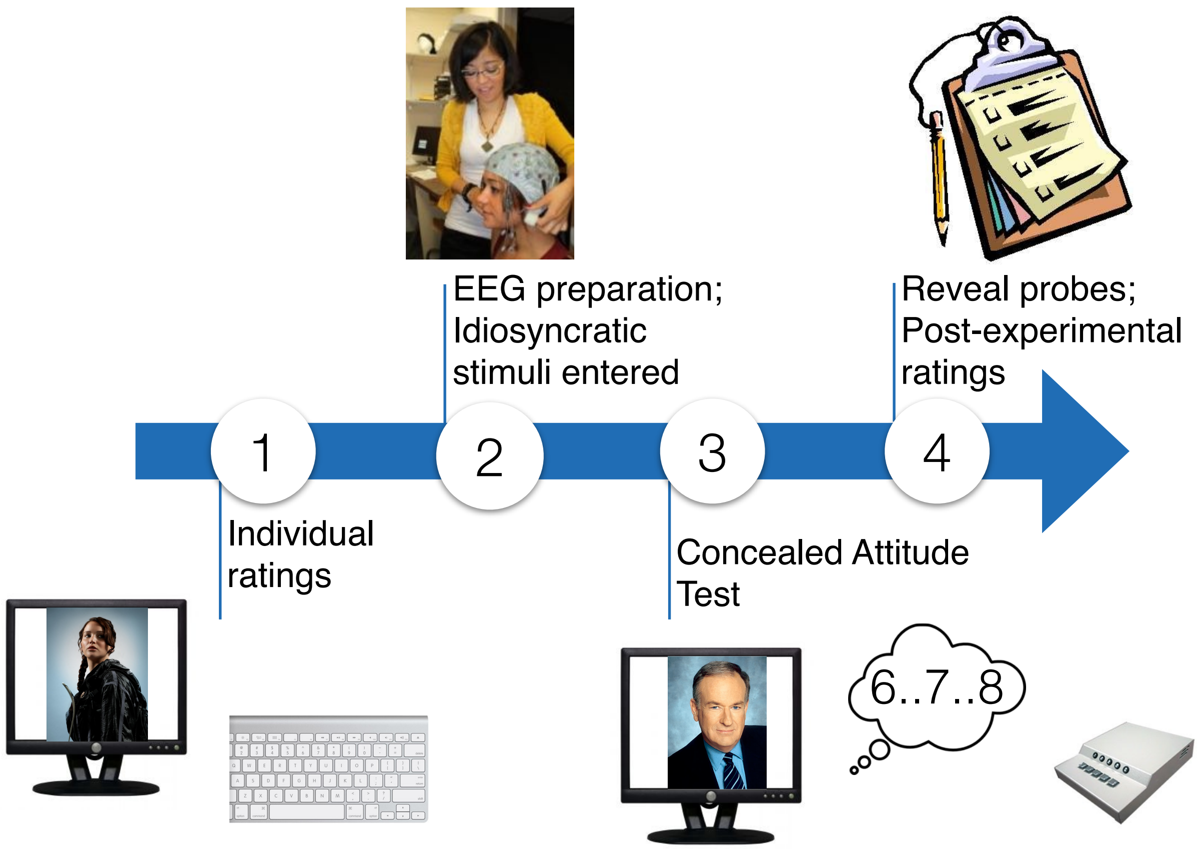


Figure 1: Timeline of experimental procedures.

1) The idiosyncratic selection phases for control and probe stimuli. 2) EEG preparation and time period in which the idiosyncratic stimuli obtained from time point 1 were entered for the Concealed Attitude Test. 3) The evaluative oddball LPP-based Concealed Attitude Test time period. Participants were randomly assigned to either mentally count or key press toward controls. 4) Post-experimental procedures in which attitudes toward probes were revealed and ratings of task difficulty were made.

[3] Neutral stimuli from the IAPS (i.e., $M \sim 5$ on valence) were not presented to participants, nor were IAPS pictures that could not be resized similar to the celebrity set ($n = 536$).

explicit rating to the Self-Assessment Manikin (SAM; control assessment) or made a "match response" (probe assessment; see below). In either case, if participants did not know one of the celebrities during either assessment phase they pressed the "enter" key ("N/?") on a numeric keypad. None of the context IAPS stimuli were presented during the idiosyncratic picture selection phase to prevent the same stimulus from being both a context and an idiosyncratic stimulus. If the participant took 45 minutes or greater to complete either of the assessments, participants did not proceed to the experimental phase (n = 6). This procedure was to prevent the experimental session from exceeding the already long three hour average. Also, it was reasoned that if participants took well of 30 minutes during the control session then the probe session would be lengthy.

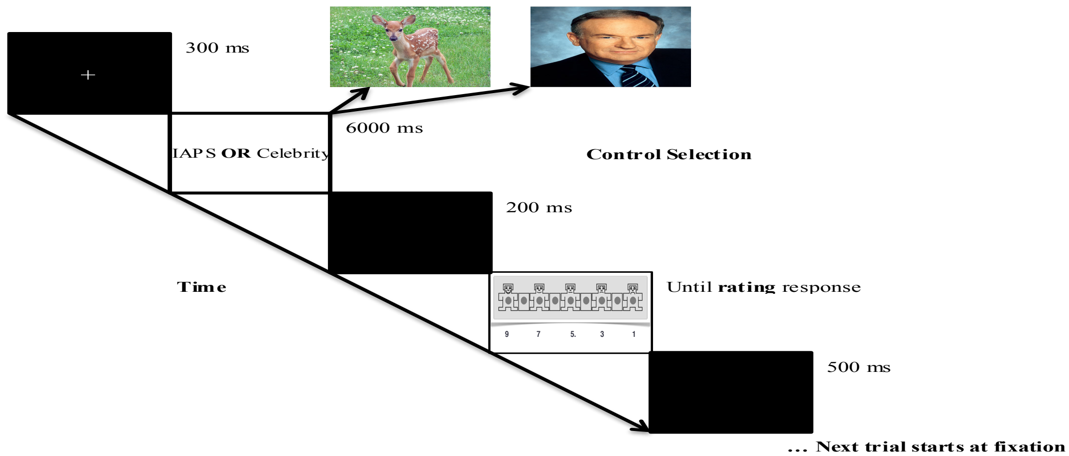
The principal function of the control stimulus selection phase was to collect extreme valence ratings to four pictures from each participant. Following the randomly presented IAPS or celebrity picture, participants made an *explicit* numeric keypad valence rating of pictures using the 9-point SAM. Specifically, these pictures were selected once participants made one positive IAPS and one positive celebrity rating 7 or greater on valence and one negative IAPS and one negative celebrity rating less than 3 on valence. After the control stimulus selection phase, ratings of the four controls that met these criteria were verified[4].

The principal function of the probe assessment was to collect "match responses" to two *new* celebrities and two *new* IAPS for each participant. To guide these "match responses" participants saw the recently selected four controls (two idio-IAPS, two celebrities) in a montage at the beginning of each trial (5000 ms) after the fixation cross but before stimulus presentation. Following a randomly presented

---

4 Participants performed four additional control trials in which each control stimulus was presented separately and ratings of each control picture appeared twice after the SAM. So, the alteration to the regular controls trials (i.e., before verification; see Figure 2) was that 1) instead of a randomly presented idio-IAPS or celebrity one of the four pictures just selected was presented, 2) the SAM was presented for 4000 ms, 3) the rating that was provided for that picture appeared ("8") until participants verified ("Y") or disconfirmed ("N/?") after the 500 ms inter-trial interval, and 4) the same rating that was provided for the pictured appeared a second time ("8") until participants verified ("Y") or disconfirmed ("N/?"). One participant underwent the control script twice because one of the pictures was disconfirmed twice in the initial verification procedure.

IAPS or celebrity picture, participants made an *implicit* "match response" to pictures by pressing the "space" key to the SAM ("Y").



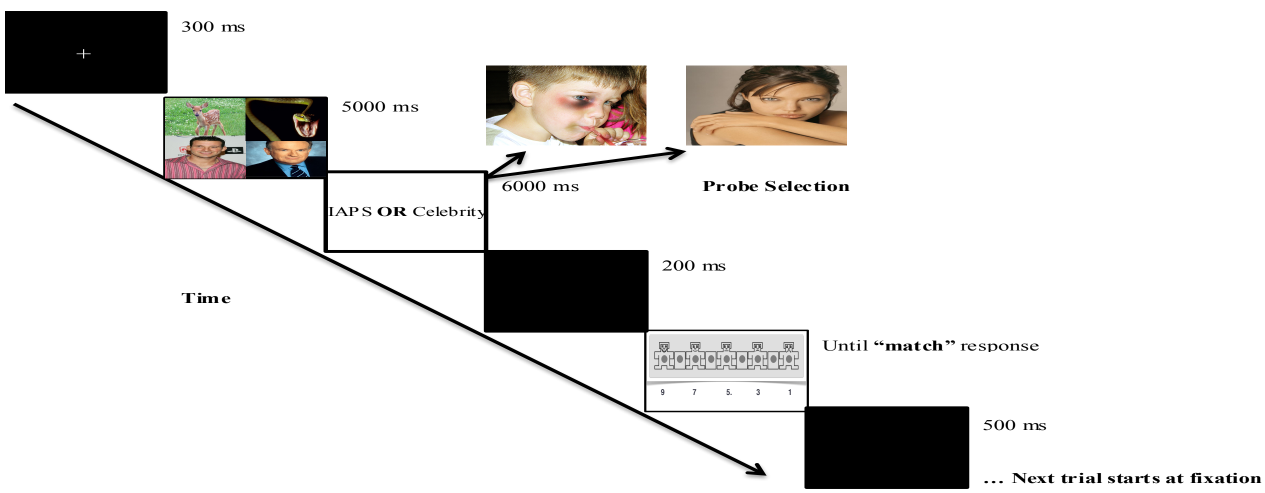


Figure 2: Individual trials of the idiosyncratic selection phase for controls (top) and probes (bottom).

The "rating" response for the controls involved an explicit 9-point rating, whereas "match" responses were key presses signaling that one of the four picture types in the montage (e.g., positive celebrity = Romo) was matched to a new stimulus of the same stimulus type (e.g., Jolie).

Participants would make a match response to the indefinitely presented SAM only when a picture was comparable in valence strength to one the four controls in the montage. Participants were instructed that once the SAM was displayed if the stimulus was similar to a previously rated positive or negative control to write the stimulus and valence rating down. Then participants pressed the space key ("Y") to advance to the next trial. Once participants made four "match responses", the script terminated and participants folded the paper, placed it in their pocket, and did not remove it until they completed the post-experimental questionnaire. Participants were told to keep these attitudes to themselves; if pressed, research assistants were vague whether participants would later indicate their attitudes toward probes. Last, participants were shown a montage of the four probes they made "match responses" to and verified that the pictures in the montage were the ones they intended to match to controls.

Experimental phase. After participants finished the idiosyncratic picture section phase, they were prepped for EEG, taken to an isolated chamber, and received instructions for concealing their attitudes toward probes (adapted from Allen et al., 1992). Participants were instructed to count or make a key press (between-subjects) to evaluatively incongruent control and "catch" pictures as quickly and accurately as possible. Alternatively, participants would not make a response if pictures were probes (congruent or incongruent), evaluatively congruent controls, or context pictures.

Trials were divided into four blocks so participants could break from the experiment and to establish a context—each block included either positive or negative IAPS context stimuli (see Herring et al., in press, for review on evaluation and breaking). For the count condition, participants entered the number of evaluatively incongruent and catch pictures using a numeric keypad at the end of each block. Participants in the key pressing condition made responses during individual trials so did not report any values at the end of each block. In each block, the global probability of an evaluatively incongruent picture was 16.6% and context/context congruent stimuli were 83.3%. A quasi randomization technique ensured at least two context stimuli were between evaluatively incongruent pictures to maintain an

acceptable target-to-target interval to elicit the evaluative congruity effect (Croft, Gonsalvez, Gabriel, & Barry, 2003; Gonsalvez et al., 1999).

The four blocks were divided into two "low" and two "high" blocks in which context trials were 105 and 147, respectively. The purpose of low and high blocks was to prevent participants from learning the number of evaluatively incongruent and catch pictures in the count condition. That is, if the trials in each block were always the same (e.g., 18 responses) participants could guess what the next count total would be instead of paying attention to the task during the block. In each block, participants could make up to 15 or 21 responses (i.e., counting or key pressing) in low and high blocks, respectively, to lessen response predictability. There were four block sequences (randomly assigned across participants); for instance, sequence A from block 1 to block 4 was low positive (LP), high negative (HN), high positive (HP), low negative (LN). The remaining three sequences were HP, LN, LP, HN (Sequence B); LN, HP, HN, LP (Sequence C); and LN, HP, HN, LP (Sequence D)[5]. The breakdown of trials by conditions is in Table 1.

Stimuli were presented with E-Prime 2.0 (Schneider, Eschman, & Zuccolotto, 2002) against a dark background on a 22" monitor (60 Hz refresh rate) with a viewing distance of 0.6 m. Each trial included a sequential presentation of a fixation cross (500 ms), the picture stimulus (1000 ms), and an ITI of 1500 ms. On each trial during stimulus presentation, either an IAPS context (e.g., negative), evaluatively incongruent (e.g., positive; IAPS or celebrity, probe or control), context congruent (e.g., negative; IAPS or celebrity, probe or control), or evaluatively incongruent "catch" (e.g., positive; IAPS drawn from other context valence set) picture was presented.

---

[5] The reason for this sequencing of the blocks was to further prevent participants from ignoring the valence of stimuli by pairing responses to the evaluatively incongruent and catch pictures. For instance, if the first two blocks were both positive contexts, participants would eventually learn that all of their responses are to negative stimuli and they may lookout for these stimuli instead of evaluating each stimulus. Because the valence context alternated each block, pictures that were evaluatively incongruent controls in the previous block were not responded to in the subsequent block and thus the opposite valence control pictures were then responded to.

Table 1: Example of a block when the context was positive and negative controls and catch trials were responded to with key presses or mental counts.

Trial types	Example	Task: Press	Task: Count	Average trials (*across four blocks*)	"Low" trials (total *across two blocks*)	"High" trials (total *across two*)
CONTROLS						
Positive *IAPS* in Positives	ppppp…P			6 (24)	5 (10)	7 (14)
Positive *celebrity* in Positives	ppppp…P			6 (24)	5 (10)	7 (14)
***Negative *IAPS* in Positives**	ppppp…N	Negative	Count	6 (24)	5 (10)	7 (14)
***Negative *celebrity* in Positives**	ppppp…N	Negative	Count	6 (24)	5 (10)	7 (14)
PROBES						
Positive *IAPS* in Positives	ppppp…P			6 (24)	5 (10)	7 (14)
Positive *celebrity* in Positives	ppppp…P			6 (24)	5 (10)	7 (14)
Negative *IAPS* in Positives	ppppp…N			6 (24)	5 (10)	7 (14)
Negative *celebrity* in Positives	ppppp…N			6 (24)	5 (10)	7 (14)
CONTEXT & CATCH						
Positive context	ppppp…			126 (504)	105 (210)	147 (294)
***Negative *catch (IAPS)* in Positives**	ppppp…N	Negative	Count	6 (24)	5 (10)	7 (14)
Column totals				180 (720)	150 (300)	210 (420)

Note. In each block incongruent oddballs are 16.6% probable (i.e., four evaluatively incongruent oddball types + "catch" trials; boldfaced) and context stimuli are 83.3% probable (i.e., context stimuli + four context congruent types; *not* boldfaced). Lower case letters (e.g., ppppp…) represent the context whereas capital letters oddballs (e.g., N) and context congruent comparisons (e.g., P). Press = task in which participants make key presses to evaluatively incongruent controls and catch oddballs (e.g., negatives). Count = task in which participants mentally count evaluatively incongruent controls and catch oddballs (e.g., negatives). Boldface conditions = evaluatively incongruent oddballs. * = Stimuli requiring a responses. Note that there are no asterisks next to probes because participants are concealing.

2.4 Signal Processing

Bioelectrical activity was recorded using Ag/AgCl electrodes in an Aegis Array cap (Electrode Arrays) with standard positioning (American Encephalographic Society, 1994). EEG activity was recorded from 62 scalp locations and referenced to the left mastoid. Electrical activity was also recorded from the right mastoid so a digitally linked reference could be used following data collection. A ground electrode was located at AFz. Activity was recorded below the right eye by an infraorbital electrode to compute vertical electrooculography (VEOG). Neuroscan SynAmps were used to amplify,

filter (bandpass: 0.05-30 Hz), and digitize (250 Hz) the bioelectrical signals that were recorded continuously.

A number of steps were taken to reduce and quantify the bioelectrical data offline using MATLAB 8.0 (R2012b) with EEGLAB 11.0.5.4b (Delorme & Makeig, 2004) and ERPLAB 3.0.2.1 (http://www.erpinfo.org/erplab) toolboxes. Bad electrodes were identified by manually reviewing each channel, and corrected using spherical spline interpolation by using the entire scalp distribution to reconstruct poor scalp sites (Perrin, Pernier, Bertnard, Giard, & Echallier, 1987; Pizzagalli, 2007). The EEG was digitally filtered 0.1-20 Hz at 12 dB/octave. EEG were then re-referenced to a digital, linked-mastoids reference (Hagemann, Naumann, & Thayer, 2001). An independent components analysis (ICA) procedure for removing ocular artifacts was applied (Jung et al., 2000; Makeig & Onton, 2012; Stone, 2002) using the ADJUST algorithm (Mognon, Jovicich, Bruzzone, & Buiatti, 2011). In short, ICA is a method for extracting independent, uncorrelated source signals from a mixture of signals (e.g., eye artifacts, muscle activity, environmental noise, and psychologically relevant cortical activity) inherent in the EEG. The ADJUST algorithm is a method for automatically detecting artifacted ICs by using spatial and temporal features (e.g., capturing abrupt changes at frontal sites for eye blinks). Epochs associated with each stimulus (300 ms prestimulus-1500 ms poststimulus) were extracted. Each epoch and electrode site was baseline corrected to the mean of the pre-stimulus period. Epochs containing extreme activity at any scalp site ($\pm$ 100 μV) were rejected using a moving window (200 ms) peak-to-peak 50 ms step function. This moving window technique is preferable over an absolute voltage threshold because it maximizes the sensitivity of artifact rejection by being less sensitive to baseline drift.

The EEG recordings over each electrode site for each participant were averaged separately within each of the experimental conditions. A minimum of 20 critical trials were required for each participant for the four critical experimental conditions (i.e., evaluative congruity X response

instruction) to ensure stable LPPs (Cohen & Polich, 1997). The LPP local peak amplitude was quantified by selecting maximum (positive) voltages between 475–775 ms (*M* = 532 ms) following stimulus onset, falling within this component's known waveform latency range (Cacioppo et al., 1993; Crites et al., 1995, 2010).

Chapter 3: Results

3.1 Group Analysis

Response accuracy and task difficulty. Overall behavioral performance (*hits plus correct rejections*) was computed[6] for both response conditions to identify outliers, and then *accuracy* was computed for the primary analysis below to compare across response conditions. One participant from the key press condition was removed from the analysis for accuracy 2.5 *SD*s below overall behavioral performance (M = 92.9%, SD = 3.9%; counting condition M = 97.6%, SD = 1.9). Accuracy for the count condition was computed as the total of evaluatively incongruent (control and catch) pictures reported across blocks divided by 72. Instances where a value greater than 72 was reported were scored as 72 because the assumption was that over reports were false alarms. For example, if a participant's total count was 73 across blocks, which is one higher than the total possible report of evaluatively incongruent controls and catches, then that participant's hits were scored as 72 and the additional report was scored a false alarm. To facilitate comparison with the counting condition, the key press condition was computed as the number of correct evaluatively incongruent stimuli responded to divided by 72. That is, across the four blocks the total possible presentations of evaluatively incongruent controls and catches—the two stimuli participants responded to—was 72.

Comparison of accuracy between the response conditions revealed that the counting condition (M = 85.3%, SD = 19.2%) was significantly more accurate than the key pressing condition (M = 61.7%, SD = 26.2%), t (50) = 4.44, $p < .001$, d = 1.25[7]. Additionally, the key pressing condition (M = 3.59, SD

---

[6] For the counting condition, correction rejections were scored perfectly unless participants over-reported the total counts across blocks. Thus, this measure exaggerates correct rejections in the counting condition.

[7] Analysis was on the arcsine transformation of the percentages.

= 1.80)[8] was rated more difficult than the counting condition (*M* = 2.40, *SD* = 2.00), *t* (50) = 2.26, *p* = .02, *d* = .63.

LPP analysis. The LPP group level analysis was divided into a priori analysis comparing local peak amplitudes among the primary experimental conditions and an exploratory analysis of the overall evaluative congruity effect (i.e., evaluatively incongruent-congruent trials) across conditions, scalp sites, and select time points. Averaged electrode arrays were computed for the a priori analyses The LPP group level analysis involved comparing local peak amplitudes among the primary experimental conditions. Averaged electrode arrays were computed for the a priori analyses to ease interpretation, better fit the ANOVA model (Dien & Santuzzi, 2005), and to foster comparison to prior research (Crites et al., 2010; Güereca et al., 2010). An averaged array of six parietal electrodes were chosen a priori (P1, Pz, P2, PO3, POz, PO4; see Herring, Taylor, White, & Crites, 2011). The LPP local peak amplitudes were analyzed with a 2 (evaluative congruity: congruent vs. incongruent) by 2 (response instruction: control vs. probe) by 2 (response task: key press vs. count) by 2 (stimulus type: IAPS vs. celebrity) mixed ANOVA with the response task between-subjects. The exploratory evaluative congruity effect of the scalp was conducted with *t*-tests from 100-1000 ms using the permutation-based strong control *t* max test for multiple comparisons with the Mass Univariate ERP Toolbox (Groppe, Urbach, & Kutas, 2011). To reduce the number comparisons, these data were down sampled from 250 Hz to 125 Hz.

A priori analysis. We first tested the three a priori predictions: 1) larger LPPs to evaluatively incongruent than congruent pictures, 2) larger LPP evaluative congruity effect (evaluatively incongruent-congruent trials) to control relative to probe trials, and 3) larger LPP evaluative congruity effect for counting than key pressing. As expected, there was an effect of congruity, F (1, 50) = 28.81, $p < .001$, $\eta_p^2 = .36$; LPPs were larger to pictures evaluatively incongruent (*M* = 8.51 μV, *SD* = 2.96 μV)

[8] The average correct response time for the press condition was 802 ms (*SD* = 258 ms).

than congruent (*M* = 7.11 μV, *SD* = 3.60 μV) with context pictures[9] (see Figure 3). Contrary to our hypothesis that evaluative congruity would be larger to controls than probes, the congruity effects for controls (*M diff* = 1.57 μV, *SD* = 2.47 μV) and probes (*M diff* = 1.24 μV, *SD* = 2.71 μV) were no different, $F\ (1, 50) = .37$, $p = .55$, $\eta_p^2 = .00$. Also, the congruity effect for the counting task (*M diff* = 1.14 μV, *SD* = 1.98 μV) was no different than the congruity effect for the key pressing task (*M diff* = 1.65 μV, *SD* = 1.77 μV), $F\ (1, 50) = .94$, $p = .33$, $\eta_p^2 = .02$. In sum, the evaluative congruity effect was replicated, but the evaluative congruity effect was not moderated by response instruction or by response task.

Unexpectedly, the evaluative congruity by response instruction (control vs. probe) by response task (count vs. key press) interaction was significant, $F\ (1, 50) = 4.65$, $p = .036$, $\eta_p^2 = .09$. For control trials the evaluative congruity effect was larger for the key press condition (*M diff* = 2.31 μV, *SD* = 2.74 μV) than the count condition (*M diff* = .77 μV, SD = 1.88 μV), $F\ (1, 50) = 5.51$, $p = .023$, $\eta_p^2 = .09$, whereas for the probe trials the evaluative congruity effect did not differ between key press (*M diff* = .99 μV, *SD* = 2.26 μV) and count conditions (*M diff* = 1.52 μV, *SD* = 3.15 μV), $F\ (1, 50) = .49$, $p = .48$, $\eta_p^2 = .02$ (see Figure 4 & Appendix A). In sum, key pressing during control trials displayed a significantly larger congruity effect than counting during controls trials.

[9] The Pitman-Morgan test of differences between correlated variances revealed that evaluatively congruent ($s^2 = 8.76$) and incongruent ($s^2 = 12.96$) variances were different, $t\ (50) = 2.67$, $p = .01$ (Kenny, 1953).

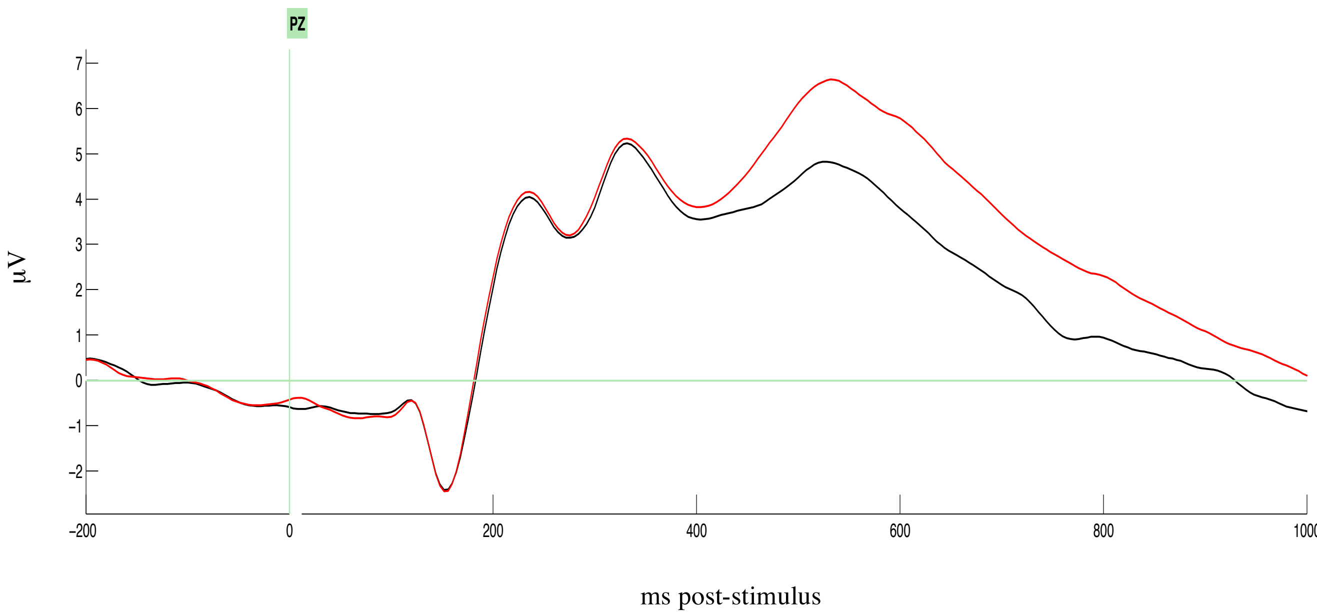


Figure 3: Individual trials of the idiosyncratic selection phase for controls (top) and probes (bottom).

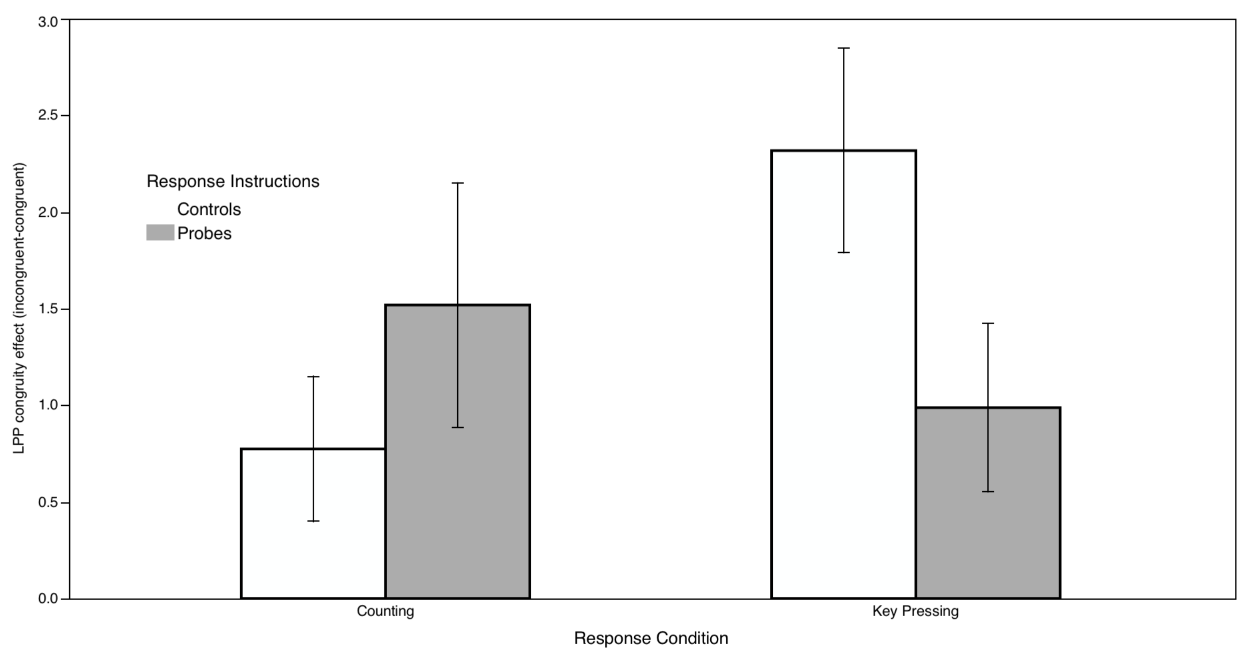


Figure 4: Evaluative congruity effect (incongruent-congruent) of the LPP for controls (white bar) and probes (gray bar).

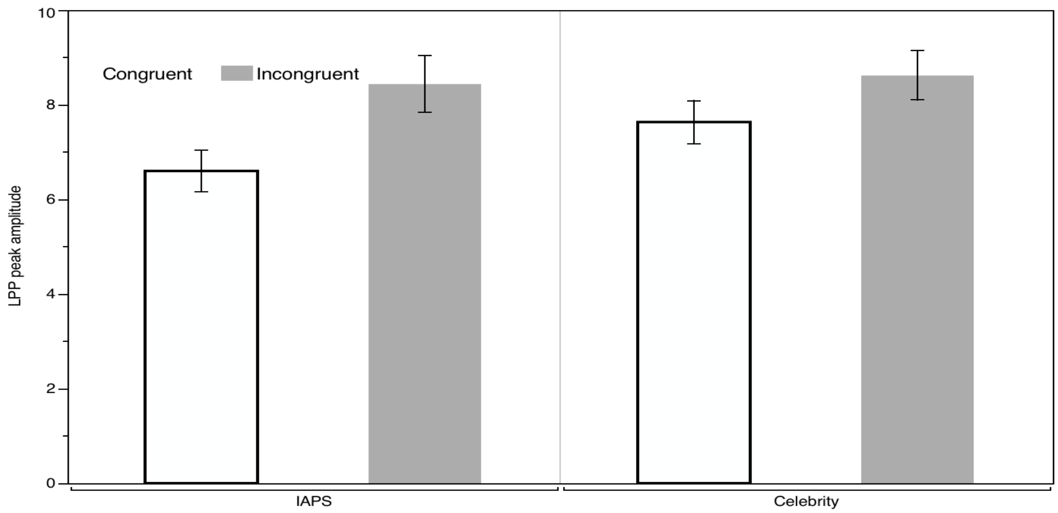


Figure 5: LPP local peak amplitude for congruent (white bar) and incongruent (gray bar) trials. Error bars are one standard error of the mean.

There were two other unanticipated findings, both of which were marginal. First, there was an evaluative congruity by stimulus type interaction, F (1, 50) = 3.89, p = .054, η_p^2 = .07. LPPs were larger to celebrity congruent (M = 7.63 μV, SD = 3.28 μV) than idio-IAPS congruent (M = 6.60 μV, SD = 3.13 μV) pictures, F (1, 51) = 9.18, p = .004, η_p^2 = .15, but LPPs were no different between celebrity incongruent (M = 8.61 μV, SD = 3.75 μV) and idio-IAPS incongruent (M = 8.61 μV, SD = 4.28 μV) pictures, F (1, 51) = .13, p = .71, η_p^2 = .00 (see Figure 5 & Appendix B). Thus, these differences between the idio-congruent and celebrity congruent conditions drove the evaluative congruity by stimulus interaction, leading to a larger evaluative congruity effect for idio-IAPS pictures compared to celebrity pictures. Second, there was a stimulus type by response condition interaction, F (1, 50) = 3.27, p = .077, η_p^2 = .06; LPPs were larger to celebrities (M = 8.89 μV, SD = 3.51 μV) than IAPS (M = 7.65 μV, SD = 3.19 μV) for key pressing, F (1, 26) = 6.28, p = .02, η_p^2 = .20, but celebrities (M = 8.61 μV,

SD = 3.75 μV) and IAPS (M = 8.61 μV, SD = 3.75 μV) were no different for the counting condition, $F(1, 24) = .02$, $p = .89$, $\eta_p^2 = .00$.

Mass univariate analysis. We permuted congruity and incongruity averages across participants using 2500 iterations to create a null distribution. This analysis indicated that a t value +/- 4.25 (i.e., the t max statistic) would be significant using a family-wise error rate of .05 (test-wise = .000083). Evaluative congruity effects were largest early frontally beginning 468 ms centrally (e.g., Fz, F3, AF4), and largest later centrally ending 972 ms (e.g., CP3, CPz, C6). Notably, the chosen electrodes comprising the parietal-occipital averaged array for the a priori analyses were typically not found significant (but see Pz 580-644 ms; see Appendix D).

3.2 Individual Analysis

A Bayesian classification approach adapted from the CIT literature was applied to each participant's idiosyncratic evaluatively incongruent control and probe averages and the context averages (see Allen, 2002; Allen et al., 1992 for review). This CAT adapted Bayesian approach enables answering, what is the *posterior probability* (see Appendix C for details) that an individual's LPP occurred to a picture evaluatively incongruent (e.g., negative) with a preceding context (e.g., positive) given a combination of LPP indicators? These probabilities were calculated using local peak amplitudes as indicators from the six electrodes used in the group analysis (i.e., P1, Pz, P2, PO3, POz, PO4; see Table 2) with the program Zbayes (Allen et al., 1992)[10].

The posterior probabilities by response condition, evaluatively incongruent averages[11], and context averages across the six indicators are presented in Table 3. As seen in Table 3, the distribution

[10] Because these indicators are strongly correlated, caution is needed in interpreting the posterior probabilities in the present study (Kruschke, 2010). Future studies would profit from including more unique indicators (e.g., mean amplitude, latency, etc.).

[11] Posterior probabilities could not be computed for evaluatively congruent controls and probes because the base rate of congruent trials was 13.3%. The base rate of congruent trials needed to be 16.6%, as the incongruent trials were, to compare

of posterior probabilities fell either below 21% or above 64.9%. Using this natural break in the distribution, we classified posterior probabilities greater than 64.9% as evaluatively incongruent and posterior probabilities less than 21% as context. (Using conventional significance did not alter classification by much. See footnote 10 below.)

Table 2: Correlations Among the Six Electrode Indicators.

Pz	P1	P2	PO3	POz	PO4
1					
.85	1				
.84	.86	1			
.80	.89	.78	1		
.82	.73	.69	.82	1	
.68	.68	.79	.77	.76	1

With this classification criterion, the six electrode indicators had a combined sensitivity of 76.9%. As seen in Table 3, controls and probes each have 52 averages because there was a total of 52 participants. Response instruction was within-subjects so the total evaluatively incongruent averages were 104. Counting each posterior probability in the control and probe columns that exceed 64.9%, controls were equal to 42 and probes were equals to 38. Thus, 80/104 (76.9%) evaluative incongruent averages (i.e., controls and probes) were correctly classified as incongruent. Treated separately, 42/52 (81%) of participants' control trials were correctly classified and 38/52 (73%) of participants' probe trials were correctly classified. Further, these six indicators had a combined specificity of 84%. As seen in Table 3, there are 10 context averages (see Appendix C for details on the rationale for context averaging) for each participant yielding 520 context averages. Counting each posterior probability in

these averages to context trials. Similarly, posterior probabilities were not computed independently for the idiosyncratic IAPS and celebrity pictures because the base rate of celebrities (13.3%) was not equal to the base rate of the IAPS (16.6%). This imbalance of base rates for the idiosyncratic stimuli was introduced by the inclusion of the "catch" stimuli (see Table 1 & Appendix C).

the context columns that were below 21% equals 437. Thus, 437/520 (84%) context averages were correctly classified as context (see Table 4 for a breakdown by indicators)[12].

Group level comparison of individual classifications. We then submitted the above individual classifications to group level comparisons with nonparametric tests. This enabled us to examine for the first time whether there are differences in classification rates between probe and control trials. Also, we were able to determine whether classification rates would be elevated for counting over key pressing during probe trials. Specifically, using hit rates for the control and probe conditions, we 1) compared probes and controls across response conditions and 2) examined the effect of response task on probe trials. First, we compared within-subject group differences between probes and controls with McNemar's test. This test ($N = 52$) revealed no significant differences ($p = .42$) in hit rates between control ($M = 81\%$, $SD = 39\%$) and probe trials ($M = 73\%$, SD = 44%).

W e also conducted a 2 (classification: correct vs. incorrect) by 2 (response task: key press vs. count) chi-square test of homogeneity for the probe trials. However, the counting ($M = 76\%$, $SD = 43\%$) and key pressing ($M = 70\%$, $SD = 46\%$) hit rates for the probe trials we no different, χ^2 (1) = .21, $p = .65$ (see Figure 6)[13]. Taken together, these data suggest that hit rates are not significantly hampered by concealing relative to truth telling (i.e., control trials), nor are they elevated for counting (vs. key pressing) during probe trials.

[12] Rather than using the natural break in the distribution of posteriors, one could use taxometric procedures for finding cut points (Ruscio, Haslam, & Ruscio, 2006). An analysis was performed using TaxProg in RStudio version 0.97.332. However, because the between-group indicator validity was low (all *ds* < 1.25), this approach was not implemented.

[13] Thought not germane to our aims, we nonetheless examined the effect of response condition on controls trials. This test too revealed non-significance between the key pressing ($M = 89\%$, $SD = 32\%$) and counting conditions ($M = 72\%$, $SD = 45\%$), χ^2 (1) = 2.38, $p = .12$.

Table 3: Posterior Probabilities by Participant, Response Condition Evaluatively Incongruent Controls and Probes, and Context (across electrode indicators).

Sub	RT	Con	Pro	Context 1	Context 2	Context 3	Context 4	Context 5	Context 6	Context 7	Context 8	Context 9	Context 10
1	Counting	0.999	**0.149**	0	0	0	**0.999**	0	0	0.738	0	0.001	0
2	Counting	**0.099**	0.999	0.01	0.149	0	0.007	0	0.015	0.015	0	0	0
3	Key Press	0.999	0.999	0	0	0	0	0	0	0	0	0	0
4	Key Press	0.999	**0**	0	0	**0.999**	0	0	0	0.001	0	0	**0.999**
5	Counting	0.999	0.999	0	0	0	0.001	0	0	0	0	0	0.006
6	Counting	0.999	0.999	0	0	0	**0.691**	0.015	0.139	0	0	0	0
7	Counting	0.999	0.999	0	0	0	0	**0.999**	**0.793**	0	**0.755**	0	0
8	Key Press	0.999	0.999	0	0	0	**0.793**	0.097	0	0	0.001	0	0
9	Counting	0.999	**0.118**	0	0	0	0	0.001	0	**0.999**	0	0	**0.65**
10	Key Press	0.999	0.999	0	0.012	0	0	0	0	0.178	0.012	0	0
11	Counting	0.985	0.985	0	0	0	**0.999**	0	**0.999**	0	0	0	0.097
12	Key Press	0.999	0.999	0	0	0	0.001	0.097	**0.999**	0.97	0	0	0
13	Key Press	0.999	0.999	0	0	0	0	0.001	0.001	0.015	0.007	0	0
14	Counting	**0.138**	**0**	0.15	0	0	0	**0.999**	**0.999**	**0.976**	0.001	0	0
15	Counting	**0.001**	0.999	**0.691**	0	0	**0.999**	**0.999**	0	0	**0.999**	0	0
16	Key Press	0.999	**0**	**0.98**	0	0	0	0	0	0.01	0	0	0

17	Counting	0.985	0.999	0	0	0	**0.999**	0	**0.999**	0	0	0	**0.985**
20	Key Press	0.999	0.999	0	0	0	0	0	0	0	0	0	**0.999**
21	Counting	0.999	0.999	**0.755**	0.001	0	0	**0.999**	0	0	0	0	0
22	Key Press	0.999	0.999	0	0	0	0	**0.999**	**0.999**	**0.98**	0	0	0
23	Key Press	0.999	**0.01**	0	0	0	0.738	0.007	**0.733**	**0.999**	0	0.136	0
24	Counting	**0**	0.98	**0.999**	0.097	0	**0.976**	0.01	0	0.001	0	0.001	0.001
25	Key Press	0.98	0.98	0.001	0	0	0	0	0.001	**0.999**	0	0	0.097
26	Counting	0.755	0.976	0	0	0	0	**0.98**	0.7	0.001	0	0	0.118
27	Key Press	0.999	0.999	0	0	0	0	0.001	0	0	**0.999**	0	0
30	Counting	0.999	**0**	**0.98**	0.001	0	0.001	**0.98**	0.182	0.207	0	0.007	0
31	Counting	**0.009**	0.999	**0.999**	0	0	0	**0.999**	**0.97**	0	0	0	0.001
32	Key Press	0.999	**0.007**	**0.999**	**0.999**	0	0.006	0	0.001	0	0	0	0
33	Key Press	**0.113**	**0.01**	0	0.001	0.209	0.001	0.015	**0.976**	0.009	0	0.096	0.007
34	Key Press	**0**	**0**	0	0	0	**0.999**	**0.999**	**0.999**	0.207	0	0	0
35	Counting	0.795	0.999	**0.755**	0	0	0.209	**0.999**	0	0	0.001	0	0
36	Key Press	0.999	0.999	0.012	0.006	0	0	0	0	0	0.001	0	0
37	Counting	0.698	0.999	0	0	0	0	0	0	0.006	0	0	0
39	Counting	0.755	0.976	0.209	0	0.006	0.007	0	0	0.149	0.009	0.015	0.001
40	Key Press	0.999	**0**	0	0	0	**0.976**	**0.82**	**0.976**	**0.976**	0	0	**0.703**
41	Counting	0.999	0.985	0	0	0	0.007	0	0	**0.999**	0	0.007	0
43	Counting	0.999	0.999	0	**0.999**	0	0	0	0	**0.999**	0	**0.999**	0

44	Counting	**0**	**0**	0.01	0	0	0.755	0.755	0.118	0.999	0	0	0.009
45	Key Press	0.97	0.999	0	0	**0.999**	0	0	0.115	0	0	0	0
46	Key Press	0.999	0.999	0	0.114	0	0	0.001	0	0	0.001	0.097	0
48	Key Press	0.999	0.999	0	0	0	**0.999**	0	0	0.009	0	0	0
49	Counting	**0**	0.65	0.015	0.01	0	0	0	**0.999**	**0.999**	**0.999**	0	0
50	Counting	0.999	0.999	0	0	0	**0.999**	0	0	0.009	0	0	0
51	Key Press	0.793	0.97	**0.736**	0.01	0.012	0.006	0.001	0.152	0	0.001	0	0.001
52	Counting	0.999	0.999	0	0	0	**0.999**	0	0	0	0	0	0
53	Counting	0.999	0.999	0.012	0	0	0	0.009	0	0	**0.999**	0	0
54	Key Press	0.999	0.999	**0.985**	0	0	0	0	0.097	0.001	0	0	0.7
55	Key Press	0.999	**0**	**0.98**	0	0.007	0	0	0.207	0	**0.65**	0.097	0
56	Key Press	0.999	0.999	0	0.097	0	0	0.01	0	0	0	0	**0.755**
57	Key Press	**0.097**	0.999	0	0	0.001	0.149	0	**0.999**	0	0.207	0	0
58	Key Press	0.999	0.999	0	0	0	0	**0.999**	**0.999**	**0.999**	0	0	0
59	Key Press	0.999	0.999	0	0	0	0	**0.65**	**0.999**	0	0	0	0

Note. Bold probabilities reflect misses and false alarms. There are 10 context averages to accurately reflect the base rate of these trials (i.e., 10/12 or 83.3%). See Appendix C for details. Original participant numbers are displayed above (e.g., 59), but only the 52 accepted participants' data are presented above and were analyzed. Sub = subject; RT = Response task; Con = Controls; Pro = Probes.

Table 4: Sensitivities and Specificities by Electrode Site Indicators.

Indicator	Sensitivity	Specificity	Z Cutpoint
P1	.77	.78	.47
Pz	.77	.80	.56
P2	.76	.80	.53
PO3	.80	.76	.42
POz	.80	.78	.44
PO4	.77	.75	.39

Note. Percentage of evaluatively incongruent averages classified correctly (i.e., sensitivity) and percentage of context averages classified correctly (i.e., specificity) by electrode site indicators. Z cutpoint is the *z*-score that maximally separated evaluatively incongruent from context averages (see Appendix C for details).

Chapter 4: Discussion

Consistent with prior research (Cacioppo et al., 1993, 1994), the LPP was larger to pictures evaluatively incongruent (e.g., negatives) than congruent (e.g., positives) with a context of pictures (e.g., positives). Notably, we extended previous evaluative oddball paradigms by developing a concealed attitude test (CAT) for detecting evaluatively incongruent (vs. context) trials at the individual level using a Bayesian approach (Allen et al., 1992). Previous evaluative oddball paradigms focused exclusively on using the LPP for examining group differences between concealing and truthfully reporting attitudes, thus overlooking the potential utility of the

LPP for making individual classification. Altogether, we found that evaluatively incongruent trials were correctly classified at 76.9% (i.e., the sensitivity) and context trials were correctly classified at 84% (i.e., the specificity). There was thus a 23.1% and 16% probability that evaluatively incongruent and context trials, respectively, were misclassified. The sensitivity and specificity of the current CAT are comparable to previous CITs using the Bayesian approach when taken together (Allen et al., 1992 = 94% & 96%; Allen & Mertens, 2009 = 65% & 96%; see also Matsuda, Nittono, & Allen, 2013; Matsuda et al., 2011 = 70% & 70%; Mertens & Allen, 2008). These data provide the first evidence of individual level classification of attitudes using the LPP in an evaluative oddball paradigm.

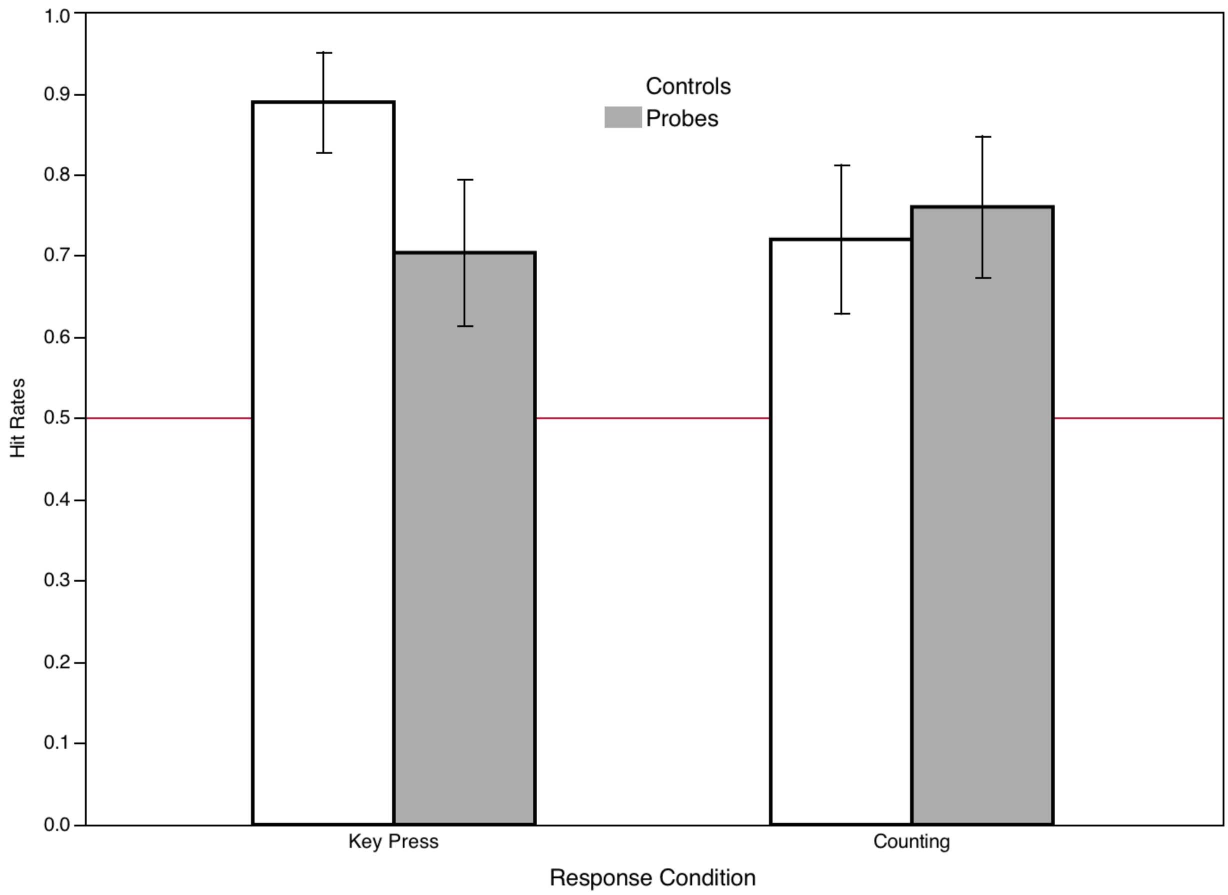


Figure 6: Percentage of evaluatively incongruent averages classified correctly (i.e., hit rates) for controls (white bar) and probes (gray bar).

4.1 Response Instruction

One of the two key variables we examined in the present study was the influence of truth telling versus concealing (i.e., response instruction) on the evaluative congruity effect. We anticipated that the evaluative congruity effect would be moderated by response instruction based on previous evaluative oddball paradigms. There was, however, no moderating effect of instruction on evaluative congruity to support our hypothesis. Prior evaluative oddball paradigms found that the evaluative congruity effect was smaller for stimuli participants concealed than truthfully reported their attitudes toward (see Table

5). The diminished evaluative congruity effect for concealment in previous studies was presumed due to cognitive load via misreporting (Crites et al., 2010; Johnson et al., 2008). That is, it was hypothesized that misreporting attitudes led to cognitive load because misreporting involves both paying attention to the stimulus ("is it positive…negative?") as well as remembering to report the opposite attitude during the task ("that's Jon Stewart who is positive so I must indicate negative"). It may be that our response instruction findings differ with prior evaluative oddball findings because of the influence of task (see discussion below on task, attention, and difficulty). We computed the standardized mean difference effect size (Cohen's *d*) for truthful and concealed responses for each task to compare with previous studies (Cumming, 2008). For the counting task, the effect sizes were comparable for concealing and truth telling—.46 and .38, respectively. For the key pressing task, truth telling was larger than concealment, in line with previous studies (.84 vs. .43, respectively). Notably, we are also speculating about the null, which is always problematic, but a post hoc (*not retrospective*) power analysis suggested that we were adequately powered (i.e., .99) to detect an effect of instruction on evaluative congruity (see Faul, Erdfelder, Lang, & Buchner, 2007).

Alternatively, there may have been no influence of instruction on evaluative congruity because participants were explicitly reminded before each block about only probe trials, whereas in previous evaluative oddballs participants were reminded about both controls and probes (Güereca et al., 2010). A comparison of the evaluative congruity effect involving retrieval instructions to controls from Güereca and colleagues to the current control condition lacking retrieval instructions to controls indicates a sizable difference—1.55 versus .64. This idea is further corroborated by a recent study demonstrating that LPPs were larger for tasks that involved retrieval instructions compared to those that do not (see Weymar, Bradley, El-Hinnawi, & Lang, 2013).

Moreover, we extended previous evaluative oddballs by examining the effect of response instruction on individual level classification rates. The Bayesian approach enabled us to estimate for

each subject for both truth telling and concealment the probability LPPs were in response to an evaluatively incongruent picture given a combination of indicators (i.e., multiple electrode sites). We then used this probability to classify the percentage of participants whose attitudes were incongruent with the context during both truth telling and concealment. This technique importantly extended previous evaluative oddball paradigms in which only group level comparisons were made between truth telling and concealment.

Table 5: Evaluative Congruity Effects during Concealment in Prior Studies.

Study	Truth Telling Effect (*d*)	Concealment Effect Size (*d*)
Crites et al., 1995 (Experiment 1)		.98/.70
Crites et al., 1995 (Experiment 2)		.57
Crites et al., 2010	.96	.69/.79
Güereca et al., 2010	1.55	.57

Not surprisingly given the lack of group level differences between truth telling and concealment, a similar number of participants were classified as having attitudes incongruent with the context for both truth telling (42/52 or 81%) and concealment (38/52 or 73%). This non-significant trend for greater classification rates for truth telling than concealment is consistent with previous CITs in which countermeasure conditions displayed lower individual classification rates than standard guilty conditions (Mertens & Allen, 2008; Rosenfeld et al., 2004). Even though there were no statistical differences between truth telling and concealment, the correct trend was apparent—when participants were actively

trying to beat the test or conceal, accuracy was lower compared to when participants were *not* trying to beat the test (i.e., truth telling). Upon first glance, these data seem useful because they suggest when participants conceal responses via merely withholding a response, that individual classification during concealment is not much lower than truth telling. That is, because previous evaluative oddball studies found smaller evaluative congruity effects for concealment than truth telling at the group level (Güereca et al., 2010, 2010), it might be expected that individual classification rates during concealment would be considerably lower. These data, however, must be considered in light of participants' task during the evaluative oddball.

4.2 Response Task

The other key variable we examined in the present study was whether counting or key pressing toward evaluatively incongruent (i.e., truthful and "catch") trials influenced the evaluative congruity effect. Based on non-evaluative oddball paradigms, we reasoned that the evaluative congruity effect would be enhanced for counting relative to key pressing (Barrett et al., 1987; Polich, 1987; Salisbury et al., 2001), but there was no effect at the group level of response task on evaluative congruity to support our hypothesis (see Mertens & Polich, 1997; Potts, 2004). There was, nonetheless, a moderating effect in which response instruction elucidated the effect of response task on evaluative congruity. Specifically, the evaluative congruity effect was larger for key pressing than counting during truth telling, but the evaluative congruity effect was no different between response tasks during concealment. In hindsight, that the evaluative congruity effect was influenced by response task during truth telling but not concealment trials should not be surprising. It is tenable for no difference in the evaluative congruity effect between response tasks for concealment because both tasks similarly involved withholding a response (i.e., pressing or counting) to a stimulus. Therefore, participants may have quickly learned in both response tasks to ignore probes despite not being told to. It was unexpected,

however, that the evaluative congruity effect was larger for key pressing than counting during truth telling—this was the opposite of what several studies from the non-evaluative oddball literature found. There are some non-evaluative oddball studies that have found larger oddball effects for key pressing than counting, thus providing a different account of response task (Brázdil, Roman, Daniel, & Rektor, 2003; Johnson, 1986).

According to Johnson (1993; 1986), different oddball effects between tasks can be explained by considering task difficulty in addition to the influence of attentional processing. Specifically, Johnson proposed that the amount of information extracted from a stimulus (e.g., positivity or negativity) is determined by the attentional resources available in addition to how difficult the task is. So how would Johnson's model reconcile previous assertions that counting leads to more attentional resources thus enhancing the oddball effect? Johnson proposed that attention moderates task difficulty, which then influences the oddball effect. One possible revision to Johnson's model is that the effects of attention and task complexity are bidirectional. That is, attention may be initially captured during truthful trials (relative to concealing) because they are task-relevant, which opens the door for a potential moderating effect of task to begin with. But the task may then influence attention depending on how demanding it is for participants. The accumulation of bidirectional effects between attention and task complexity ultimately affects evaluative congruity, and there may be a number of avenues to drive this effect. For example, when counting and key pressing are comparable in task difficulty, counting may be more attentional engaging. This might explain why some non-evaluative oddball paradigms found greater oddball effects for counting than key pressing. In contrast, when the tasks are unequal in task difficulty (e.g., key pressing is more difficult), task complexity might then enhance attention for the more difficult task (i.e., key pressing). There must be, however, a point where extreme difficulty would hamper the evaluative congruity effect (Yerkes & Dodson, 1908; cf. Kok, 2001). Subjective report and behavioral data lend some support to this idea. In non-evaluative oddball paradigms in which oddball effects were

greater for counting, there was no difference in behavioral performance suggesting that the two tasks were comparable in complexity (Polich, 1987; Salisbury et al., 2001). Whereas in the present study, where the evaluative congruity effect was larger for key pressing than counting during truthful trials, participants were more accurate for counting than key pressing suggesting that key pressing was more difficult. What is more, the subjective report data from the present study suggest greater task difficulty for key pressing—key pressing was rated significantly higher in task difficulty than counting. A caveat is that this explanation of the interplay among response instruction, response task, and evaluative congruity with respect to attention and task complexity was post-hoc.

Furthermore, we examined the effect of response task on classification rates. This allowed us to examine whether counting during concealment led to better individual classification relative to key pressing during concealment. Comparison of classification rates revealed no difference between counting (76%) and key pressing (70%) during concealment. Thus, response task was ineffective in enhancing classification during concealment. Though not the focus of the current study to compare classification rates for truthful trials between response tasks, there was a trend indicating that key pressing (89%) had greater classification rates relative to counting (72%; see footnote 9). This 17% percent difference between response conditions during control trials, though marginal (p = .12), is striking because it highlights how much poorer the counting task is at detecting attitudes even when participants are truthful.

4.3 Stimulus Type

Up to this point, we have focused on the effects of response task and response instruction on evaluative congruity because these variables were involved in testing our primary hypotheses; we also varied whether non-context stimuli were idio-IAPS versus celebrities. This was the first evaluative oddball paradigm in which we attempted to use normed-IAPS to establish a context in conjunction with

idiosyncratic non-context celebrity stimuli. This was an effort to build upon prior evaluative oddball paradigms that relied exclusively on idiosyncratic stimuli for both context and non-context stimuli (Crites et al., 2010; Güereca et al., 2010). Using idiosyncratic stimuli for both context and non-context stimuli was ecologically invalid because it would be too difficult to use a CAT in an applied setting in which the valence of all stimuli had to be determined using the LPP. In this way, if the valence of context stimuli were already known, then only the valence of the non-context stimuli would need to be ascertained using the LPP. However, if we were to use normed-IAPS as context stimuli with idiosyncratic celebrities as non-context stimuli, there would have been no way to know what influence there was on evaluative congruity by having context and non-context stimuli that differed in stimulus content (i.e., IAPS vs. celebrity pictures). Our solution for avoiding this stimulus confound was to also have idio-IAPS as non-context pictures to compare to the celebrity pictures.

Unexpectedly, there was an interaction between stimulus type (idio-IAPS vs. celebrity) and the evaluative congruity effect. Specifically, the evaluative congruity effect was larger for idio-IAPS than celebrity pictures. This enhanced evaluative congruity effect for idio-IAPS relative to celebrities was driven by larger LPPs to evaluatively congruent celebrities pictures than evaluatively congruent idio-IAPS. Evaluatively incongruent LPPs for celebrities and idio-IAPS were comparable. It is curious for the evaluative congruity effect to be affected by a task irrelevant feature (i.e., celebrity vs. idio-IAPS) because the LPP was initially thought to be highly sensitive to task-relevance (Nieuwenhuis et al., 2005). In the present case, task relevance is whether pictures were positive or negative, not whether they were IAPS or celebrities. However, recently the LPP has been found to be sensitive to a number of task-irrelevant features such as people versus non-people, attractiveness, and familiarity, which may help explain the present findings (Corral, 2009; Ito & Cacioppo, 2000; Wang, Kitayama, & Han, 2011). For example, Ito and Cacioppo (2000) had participants perform an oddball paradigm that was both evaluative and non-evaluative. That is, positive and negative IAPS pictures could involve people (e.g.,

couple embracing) or no people (e.g., chocolate bar). Thus, evaluation and people/no-people were orthogonally manipulated. This enabled the researchers to examine both evaluative and non-evaluative (i.e., larger LPPs to non-evaluatively incongruent than congruent trials) congruity effects. In one task, participants counted positive and negative stimuli. When the context was of non-people (e.g., chocolate bar) and participants saw incongruent people pictures (e.g., couple embracing), LPPs were larger than if they saw congruent non-people pictures (e.g., littered beach). This non-evaluative congruity effect was found despite evaluation of valence being task-relevant since participants were supposed to count the valence of these stimuli. Thus, there was evidence of a task-irrelevant, non-evaluative congruity effect. Because in the present study the majority (87%) of the pictures were IAPS (i.e., normed-IAPS and idio-IAPS), the celebrity stimuli were rarer than IAPS allowing them to elicit larger LPPs. There was evidence of a non-evaluative congruity effect—larger LPPs to celebrities than idio-IAPS during evaluatively congruent trials—despite being task-irrelevant. Thus, although participants were supposed to focus on evaluating pictures as positive or negative, they also noticed differences between celebrities and idio-IAPS pictures.

4.4 Future Directions and Limitations

The joint effects of task and instruction on evaluative congruity provide clues for potential avenues to extend the CAT. As noted above, during truthful but not concealment trials the evaluative congruity effect was largest for key pressing relative to counting. There was also a trend for greater individual classification rates for key pressing (89%) than counting (72%). Therefore, in future CATs if researchers could devise a method for participants to truthfully key press while concealing their attitudes, then individual classification rates might be improved during concealment. So how could one devise an approach that enables participants to conceal while making truthful key presses? One approach is to mimic the CIT. In a standard CIT, regardless if participants are guilty or innocent they

will look for stimuli called targets, make a key press to them, and usually press another key for non-targets (i.e., probes and context). In this way, participants are not misreporting or withholding responses, as has been the case in evaluative oddball paradigms, and are able to conceal knowledge of a mock crime. Similarly, a CAT could be devised so participants make three responses instead of two: positive, negative, and positive/negative. For example, participants could see normed IAPS as the context and press the positive and negative key to these pictures on the majority of trials (e.g., 80%). Critically, when participants see idio-IAPS and celebrities, they could press the positive/negative key. So, evaluation would be salient, which is critical for eliciting attitudes with implicit measures (Herring et al., in press; Spruyt et al., 2009), and the task would neither involve withholding a response nor misreporting. As a result, the individual classification might improve during concealment.

It is worth noting a key limitation of the CAT that leads to discussion of another potential future direction. The CAT detects concealed attitudes, which are related but theoretically distinct from intentions (i.e., a committed plan one's own action; Granhag, 2010). Intention is comprised of attitudes, and is the most proximal construct leading to actual behavior (Ajzen, 2011). The positive associations between attitudes, intentions, and behavior are well established (Glasman & Albarracín, 2006; Hrubes, Ajzen, & Daigle, 2001; Sheeran & Taylor, 1999). For instance, having a negative attitude toward actress Jennifer Lawrence makes it more likely one avoids planning to see her movies. Thus, it is even more unlikely one might go out and spend money to see her movies. However, there are situations where one might have a negative attitude toward Jennifer Lawrence and see her movies anyway because of the enjoyment of relaxing at the movies regardless of who is acting. There is not always correspondence among attitudes, intentions, and behavior, and this limits the CAT with respect to predicting future behavior (see Cameron et al., 2012).

One approach is to better develop implicit measures of intentions and use them in conjunction with the CAT. Recently, Ask and colleagues (2013) developed an implicit measure of true and false

intentions. Participants either had an intention to shop (true intention) or perform a mock crime with shopping as the cover story (false intention) if they were intercepted and questioned. After planning to perform one of these actions, participants then carried out a reaction time task in which goal relevant cues (e.g., "receipt", "cash", etc.) were presented. Only participants with a true intention to shop displayed an approach bias to goal-relevant shopping cues via responding more quickly to positive than negative words. These data suggest that reaction times may be useful for differentiating true and false intentions. An extension of this study might be to have participants carryout a CAT to assess their evaluations related to a mock scenario in addition to carrying out the implicit RT-based intention paradigm. If the CAT indicated no negative attitude toward the mock scenario and there was also no approach bias, then this combination of indicators may be more diagnostic than either of the two measures alone. A number of applied researchers have recently used this logic for detecting concealed information and "malintent" by using multiple measures, techniques, or both (Martin, Martin, Webb, & Horgan, 2011; Matsuda et al., 2011).

4.5 Conclusion

The primary goal of the current study was to develop a Concealed Attitude Test (CAT) with a Bayesian approach using the late positive potential of the event-related potential. Altogether, we classified incongruent pictures and context pictures with 76.9% and 84% accuracy, respectively. When participants were truthful, 42/52 (81%) participants' evaluatively incongruent pictures were correctly classified. Whereas when participants concealed, 38/52 (73%) participants' evaluatively incongruent pictures were correctly classified. The percentage of participants' evaluatively incongruent pictures correctly classified during concealment between key pressing (19/27 or 70%) and counting (19/25 or 76%) was comparable. Surprisingly, there was a trend for a greater percentage of participants' evaluatively incongruent pictures correctly classified for truth telling during key pressing (24/27 or

89%) than for truth telling during counting (18/25 or 72%). Greater accuracy during truthful key pressing than truthful counting may be because of a bidirectional influence between attention and task difficulty. It will be important to extend these findings by investigating key pressing tasks that involve neither misreporting nor withholding responses. Caution is warranted when interpreting these data as speculation was made with respect to null as well as marginal findings. Future research would also profit from determining how implicit measures of attitudes and intentions might be used in tandem. The CAT may have national security implications (e.g., screening), but a wealth of basic science is warranted before such conclusions can be drawn. In sum, these data provide a promising starting point for an individual level based approach for detecting concealed attitudes.

References

Acosta, V. W., & Nasman, V. T. (1992). Effect of task decision on P300. *International Journal of Psychophysiology*, *13*(1), 37–44. doi:10.1016/0167-8760(92)90018-7

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, *50*(2), 179–211.

Ajzen, I. (2011). Theory of planned behavior. *Handbook of Theories of Social Psychology: Volume One*, *1*, 438.

Allen, J. J. B. (2002). The role of psychophysiology in clinical assessment: ERPs in the evaluation of memory. *Psychophysiology*, *39*, 261–280.

Allen, J. J. B., & Iacono, W. G. (1997). A comparison of methods for the analysis of event-related potentials in deception detection. *Psychophysiology*, *34*(2), 234–240.

Allen, J. J. B., Iacono, W. G., & Danielson, K. D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: A methodology for prediction in the face of individual differences. *Psychophysiology*, *29*(5), 504–522.

Allen, J. J. B., & Mertens, R. (2009). Limitations to the detection of deception: True and false recollections are poorly distinguished using an event-related potential procedure. *Social Neuroscience*, *4*(6), 473–490. doi:10.1080/17470910802109939

Alqallaf, F., & Gustafson, P. (2001). On cross-validation of Bayesian models. *Canadian Journal of Statistics*, *29*(2), 333–340. doi:10.2307/3316081

Ambach, W., Bursch, S., Stark, R., & Vaitl, D. (2010). A Concealed Information Test with multimodal measurement. *International Journal of Psychophysiology*, *75*(3), 258–267.

American Encephalographic Society. (1994). Guideline thirteen: Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, *11*, 111–113.

Ask, K., Granhag, P. A., Juhlin, F., & Vrij, A. (2013). Intending or Pretending? Automatic Evaluations of Goal Cues Discriminate True and False Intentions. *Applied Cognitive Psychology*, *27*(2), 173–177. doi:10.1002/acp.2893

Barrett, G., Neshige, R., & Shibasaki, H. (1987). Human auditory and somatosensory event-related potentials: effects of response condition and age. *Electroencephalography and Clinical Neurophysiology*, *66*(4), 409–419. doi:10.1016/0013-4694(87)90210-0

Bayes, T. (1763). A Letter from the Late Reverend Mr. Thomas Bayes, F. R. S. to John Canton, M. A. and F. R. S. *Philosophical Transactions (1683-1775)*, *53*, 269–271. doi:10.2307/105732

Ben-Shakhar, G. (2002). A critical review of the Control Questions Test (CQT). In *Handbook of polygraph testing* (pp. 103–126). San Diego, CA, US: Academic Press.

Ben-Shakhar, G. (2012). Current research and potential applications of the concealed information test: an overview. *Frontiers in Cognitive Science*, *3*, 342. doi:10.3389/fpsyg.2012.00342

Ben-Shakhar, G., & Elaad, E. (2002). The Guilty Knowledge Test (GKT) as an application of psychophysiology: Future prospects and obstacles. In *Handbook of polygraph testing* (pp. 87–102). San Diego, CA, US: Academic Press.

Ben-Shakhar, G., & Elaad, E. (2003). The validity of psychophysiological detection of information with the guilty knowledge test: A meta-analytic review. *Journal of Applied Psychology*, *88*, 131–151.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The Self-Assessment Manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*, 49–59. doi:10.1016/0005-7916(94)90063-9

Bradley, M. M., & Lang, P. J. (2007). Emotion and motivation. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (Vol. 3rd, pp. 581–607). New York, NY: Cambridge University Press.

Brázdil, M., Roman, R., Daniel, P., & Rektor, I. (2003). Intracerebral somatosensory event-related potentials: effect of response type (button pressing versus mental counting) on P3-like potentials within the human brain. *Clinical Neurophysiology*, *114*(8), 1489–1496. doi:10.1016/S1388-2457(03)00135-4

Cacioppo, J. T., & Berntson, G. G. (1994a). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, *115*, 401–423. doi:10.1037/0033-2909.115.3.401

Cacioppo, J. T., & Berntson, G. G. (1994b). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, *115*, 401–423. doi:10.1037/0033-2909.115.3.401

Cacioppo, J. T., Berntson, G. G., Klein, D. J., & Poehlmann, K. M. (1998). Psychophysiology of emotion across the life span. In *Annual review of gerontology and geriatrics* (pp. 27–74). New York, NY: Springer Publishing Co. Retrieved from internal-pdf://Cacioppo et al. 1998-2217890048/Cacioppo et al. 1998.pdf

Cacioppo, J. T., Crites, S. L., Berntson, G. G., & Coles, M. G. (1993). If attitudes affect how stimuli are processed, should they not affect the event-related brain potential? *Psychological Science*, *4*, 108–112. doi:10.1111/j.1467-9280.1993.tb00470.x

Cacioppo, J. T., Crites, S. L., & Gardner, W. L. (1996). Attitudes to the right: Evaluative processing is associated with lateralized late positive event-related brain potentials. *Personality and Social Psychology Bulletin*, *22*, 1205–1219.

Cacioppo, J. T., Crites, S. L., Gardner, W. L., & Berntson, G. G. (1994). Bioelectrical echoes from evaluative categorizations: I. A late positive brain potential that varies as a function of trait negativity and extremity. *Journal of Personality and Social Psychology*, *67*, 115–125. doi:10.1037/0022-3514.67.1.115

Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential Priming Measures of Implicit Social Cognition: A Meta-Analysis of Associations With Behavior and Explicit Attitudes. *Personality and Social Psychology Review*, Advanced online publication. doi:10.1177/1088868312440047

Cohen, J., & Polich, J. (1997). On the number of trials needed for P300. *International Journal of Psychophysiology*, *25*(3), 249–255. doi:10.1016/S0167-8760(96)00743-X

Corral, G. (2009). *Does type of stimulus influence task-irrelevant evaluative categorization processes?* University of Texas at El Paso, El Paso, TX.

Crites, S. L., Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1995). Bioelectrical echoes from evaluative categorization: II. A late positive brain potential that varies as a function of attitude registration rather than attitude report. *Journal of Personality and Social Psychology*, *68*, 997–1013. doi:10.1037/0022-3514.68.6.997

Crites, S. L., Fabrigar, L. R., & Petty, R. E. (1994). Measuring the affective and cognitive properties of attitudes: Conceptual and methodological issues. *Personality and Social Psychology Bulletin*, *20*, 619–634. doi:10.1177/0146167294206001

Crites, S. L., Mojica, A. J., Corral, G., & Taylor, J. H. (2010). An event-related potential paradigm for identifying (rare negative) attitude stimuli that people intentionally misreport. *Psychophysiology*, *47*, 984–988. doi:10.1111/j.1469-8986.2010.01002.x

Croft, R. J., Gonsalvez, C. J., Gabriel, C., & Barry, R. J. (2003). Target-to-target interval versus probability effects on P300 in one- and two-tone tasks. *Psychophysiology*, *40*(3), 322–328. doi:10.1111/1469-8986.00036

Cumming, G. (2008). Replication and p Intervals: p Values Predict the Future Only Vaguely, but Confidence Intervals Do Much Better. *Perspectives on Psychological Science*, *3*(4), 286–300. doi:10.1111/j.1745-6924.2008.00079.x

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9–21.

Dien, J., & Santuzzi, A. M. (2005). Application of repeated measures ANOVA to high-density ERP datasets: A review and tutorial. In T. C. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 57–82). Cambridge, MA: MIT Press.

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating?. *Behavioral and Brain Sciences*, *11*, 357–374.

Donchin, Emanuel. (1981). Surprisep . . . Surprise? *Psychophysiology*, *18*(5), 493–513.

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology*, *14*, 456–467. doi:10.1111/j.1469-8986.1977.tb01312.x

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Orlando, FL: Harcourt Brace Jovanovich College Publishers.

Elaad, E., & Ben-Shakhar, G. (1991). Effects of mental countermeasures on psychophysiological detection in the guilty knowledge test. *International Journal of Psychophysiology*, *11*(2), 99–108. doi:10.1016/0167-8760(91)90001-E

Fabiani, M., Gratton, G., & Federmeier, K. D. (2007). Event-related brain potentials: Methods, theory, and applications. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (Vol. 3rd, pp. 85–119). New York, NY: Cambridge University Press.

Fabiani, M., Karis, D., & Donchin, E. (1986). P300 and recall in an incidental memory paradigm. *Psychophysiology*, *23*(3), 298–308.

Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy (“ lie detection”) with event-related brain potentials. *Psychophysiology*, *28*, 531–547.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition*, *25*, 603–637. doi:10.1521/soco.2007.25.5.603

Glasman, L. R., & Albarracín, D. (2006). Forming attitudes that predict future behavior: A meta-analysis of the attitude-behavior relation. *Psychological bulletin*, *132*, 778. doi:10.1037/0033-2909.132.5.778

Gobbini, M. I., Leibenluft, E., Santiago, N., & Haxby, J. V. (2004). Social and emotional attachment in the neural representation of faces. *NeuroImage*, *22*(4), 1628–1635. doi:10.1016/j.neuroimage.2004.03.049

Gonsalvez, C. J., Gordon, E., Grayson, S., Barry, R. J., Lazzaro, I., & Bahramali, H. (1999). Is the target-to-target interval a critical determinant of P3 amplitude? *Psychophysiology*, *36*(5), 643–654.

Gougler, M., Nelson, R., Handler, M., Krapohl, D., Shaw, P., & Bierman, L. (2011). Meta-Analytic Survey of Criterion Accuracy of Validated Polygraph Techniques. *Polygraph*, *40*, 203.

Granhag, P.-A. (2010). On the psycho-legal study of true and false intentions: Dangerous waters and some stepping stones. *The Open Criminology Journal*, *3*, 37–43.

Greenwald, A. G., Andrew, T., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*(1), 17–41. doi:10.1037/a0015575

Güereca, Y. M., Crites, S. L., Zambrano-Vasquez, L., Taylor, J. H., Corral, G., Torres, I. D., & Herring, D. R. (2010). Lie to me: Using ERPs to detect concealed attitudes. *Psychophysiology*, *47*(Suppl.1), 79.

Hagemann, D., Naumann, E., & Thayer, J. F. (2001). The quest for the EEG reference revisited: A glance from brain asymmetry research. *Psychophysiology*, *38*, 847–857. doi:10.1017/S0048577201001081

Herring, D. R., Taylor, J. H., White, K. R., & Crites, S. L. (2011). Electrophysiological responses to evaluative priming: The LPP is sensitive to incongruity. *Emotion*, *11*, 794–806. doi:10.1037/a0022804

Herring, D. R., White, K. R., Jabeen, L. N., Hinojos, M., Terrazas, G., Taylor, J. H., … Crites, S. L. (in press). On the automatic activation of attitudes: A quarter-century of evaluative priming research. *Psychological Bulletin*.

Hrubes, D., Ajzen, I., & Daigle, J. (2001). Predicting Hunting Intentions and Behavior: An Application of the Theory of Planned Behavior. *Leisure Sciences*, *23*(3), 165–178. doi:10.1080/014904001316896855

Iacono, W. G. (2007). Detection of deception. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (Vol. 3rd, pp. 688–703). New York, NY: Cambridge University Press. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2007-08652-029&site=ehost-live&scope=site

Iacono, W. G., & Lykken, D. T. (1997). The validity of the lie detector: Two surveys of scientific opinion. *Journal of Applied Psychology*, *82*, 426–433.

Ito, T. A., & Cacioppo, J. T. (2000). Electrophysiological evidence of implicit and explicit categorization processes. *Journal of Experimental Social Psychology*, *36*, 660–676. doi:10.1006/jesp.2000.1430

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology*, *75*, 887–900. doi:10.1037/0022-3514.75.4.887

Johnson Jr, R. (1993). On the neural generators of the P300 component of the event-related potential. *Psychophysiology*, *30*, 90–97.

Johnson, M. M., & Rosenfeld, J. P. (1992). Oddball-evoked P300-based method of deception detection in the laboratory II: Utilization of non-selective activation of relevant knowledge. *International Journal of Psychophysiology*, *12*(3), 289–306. doi:10.1016/0167-8760(92)90067-L

Johnson, R. J. (1986). A Triarchic Model of P300 amplitude. *Psychophysiology*, *23*, 367–384. doi:10.1111/j.1469-8986.1986.tb00649.x

Johnson, R. J., Henkell, H., Simon, E., & Zhu, J. (2008). The self in conflict: The role of executive processes during truthful and deceptive responses about attitudes. *NeuroImage*, *39*(1), 469–482.

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(2), 163–178. doi:10.1111/1469-8986.3720163

Kappenman, E. S., & Luck, S. J. (2012). ERP Components: The Ups and Downs of Brainwave Recordings. In S. J. Luck & E. S. Kappenman (Eds.), *Oxford Handbook of Event-Related Potential Components* (pp. 3–30). New York, NY: Oxford University Press.

Kenny, D. T. (1953). Testing of differences between variances based on correlated variates. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *7*(1), 25–28. doi:10.1037/h0083569

Kiefer, M., Adams, S. C., & Zovko, M. (2012). Attentional sensitization of unconscious visual processing: Top-down influences on masked priming. *Advances in Cognitive Psychology*, *8*, 50–61. doi:10.2478/v10053-008-0102-4

Kircher, J. C., Horowitz, S. W., & Raskin, D. C. (1988). Meta-analysis of mock crime studies of the control question polygraph technique. *Law and Human Behavior*, *12*(1), 79–90. doi:10.1007/BF01064275

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, *36*(14), 1–1.

Kline, R. B. (2004). *Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research*. Amer Psychological Assn.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*(3), 557–577. doi:10.1017/S0048577201990559

Krosnick, J. A., Judd, C. M., & Wittenbrink, B. (2005). The measurement of attitudes. *The handbook of attitudes*, 21–76.

Kruschke, J. K. (2010). *Doing Bayesian data analysis: A tutorial introduction with R and BUGS*. Burlington, MA: Academic Press.

Lang, P. J., & Bradley, M. M. (2007). The International Affective Picture System (IAPS) in the study of emotion and attention. *Handbook of emotion elicitation and assessment*, 29.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8*. University of Florida, Gainesville, FL.

Lench, H. C., Flores, S. A., & Bench, S. W. (2011). Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: A meta-analysis of experimental emotion elicitations. *Psychological bulletin*, *137*(5), 834.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, *35*(03), 121–143.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, *43*(6), 385–388. doi:10.1037/h0046060

Lykken, D. T. (1998). *A tremor in the blood: Uses and abuses of the lie detector* (Vol. xvi). New York, NY, US: Plenum Press.

Makeig, S., & Onton, J. (2012). ERP Features and EEG Dynamics: An ICA Perspective. In *The Oxford Handbook of Event-Related Potential Components* (pp. 51–86). New York, NY: Oxford University Press.

Maris, E. (2012). Statistical testing in electrophysiological studies. *Psychophysiology*, *49*(4), 549–565. doi:10.1111/j.1469-8986.2011.01320.x

Martin, D. J., Martin, J. Z., Webb, A. K., & Horgan, A. J. (2011). MALINTENT: THEORY AND DETECTION IN PRIMARY SCREENING. In *PSYCHOPHYSIOLOGY* (Vol. 48, pp. S10–S10).

Matsuda, I., Nittono, H., & Allen, J. J. B. (2012). The Current and Future Status of the Concealed Information Test for Field Use. *Frontiers in Psychology*, *3*. doi:10.3389/fpsyg.2012.00532

Matsuda, I., Nittono, H., & Allen, J. J. B. (2013). Detection of concealed information by P3 and frontal EEG asymmetry. *Neuroscience Letters*, *537*, 55–59. doi:10.1016/j.neulet.2013.01.029

Matsuda, I., Nittono, H., & Ogawa, T. (2011). Event-related potentials increase the discrimination performance of the autonomic-based concealed information test. *Psychophysiology*.

Matsuda, I., Nittono, H., & Ogawa, T. (2012). Central and autonomic responses that reflect the intention to conceal in a concealed information test. *International Journal of Psychophysiology*, *85*(3), 325.

Meijer, E., & Ben-Shakhar, G. (2012). Skin Conductance, Respiration, Heart Rate, and P300 in The Concealed Information Test: A Meta Analysis. *Psychophysiology*, *49*, S113.

Mertens, R., & Allen, J. J. B. (2008). The role of psychophysiology in forensic assessments: Deception detection, ERPs, and virtual reality mock crime scenarios. *Psychophysiology*, *45*, 286–298.

Mertens, R., & Polich, J. (1997). P300 from a single-stimulus paradigm: passive versus active tasks and stimulus modality. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *104*(6), 488–497. doi:10.1016/S0168-5597(97)00041-5

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, *48*(2), 229–240. doi:10.1111/j.1469-8986.2010.01061.x

National Research Counsel, C. to R. the S. E. on the. (2003). *The Polygraph and Lie Detection*. Washington, DC: National Academies Press.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the Locus Coeruleus-Norepinephrine System. *Psychological Bulletin*, *131*, 510–532. doi:10.1037/0033-2909.131.4.510

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113. doi:10.1016/0028-3932(71)90067-4

Osgood, C. E., Suci, C. J., & Tannenbaum, P. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.

Perrin, F., Pernier, J., Bertnard, O., Giard, M. ., & Echallier, J. . (1987). Mapping of scalp potentials by surface spline interpolation. *Electroencephalography and Clinical Neurophysiology*, *66*(1), 75–81. doi:10.1016/0013-4694(87)90141-6

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, *9*, 456–479.

Pizzagalli, D. (2007). Electroencephalography and High-Density Electrophysiological Source Localization. In *Handbook of Psychophysiology* (pp. 56–84). New York, NY: Cambridge University Press.

Polich, J. (1987). Response mode and P300 from auditory stimuli. *Biological Psychology*, *25*(1), 61–71. doi:10.1016/0301-0511(87)90067-6

Polich, J. (2012). Neuropsychology of P300. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford Handbook of Event-Related Potential Components* (pp. 159–188). New York, NY: Oxford University Press.

Potts, G. F. (2004). An ERP index of task relevance evaluation of visual stimuli. *Brain and Cognition*, *56*(1), 5–13. doi:10.1016/j.bandc.2004.03.006

Pritchard, W. S. (1981). Psychophysiology of P300. *Psychological Bulletin*, *89*(3), 506–540.

Raskin, D. C., & Honts, C. R. (2002). The comparison question test. *Handbook of polygraph testing*, 1–47.

Rosenfeld, J. P. (2011). P300 in detecting concealed information. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory Detection: Theory and application of the concealed information test*. Cambridge, MA: University Press.

Rosenfeld, J. P., Angell, A., Johnson, M., & Qian, J. H. . (1991). An ERP-Based, Control-Question Lie Detector Analog: Algorithms for Discriminating Effects Within Individuals' Average Waveforms. *Psychophysiology*, *28*, 319–335. doi:10.1111/j.1469-8986.1991.tb02202.x

Rosenfeld, J. P., Soskins, M., Bosh, G., & Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology*, *41*, 205–219.

RStudio: Intergrated development evironment for R. (2012). Boston, MA. Retrieved from http://www.rstudio.org/

Ruscio, J., Haslam, N., & Ruscio, A. M. (2006). *Introduction to the Taxometric Method: A Practical Guide*. Psychology Press.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, *110*(1), 145.

Russell, J. A., & Feldman Barrett, L. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*, 805–819. doi:10.1037/0022-3514.76.5.805

Salisbury, D. F., Rutherford, B., Shenton, M. E., & McCarley, R. W. (2001). Button-pressing affects P300 amplitude and scalp topography. *Clinical Neurophysiology*, *112*(9), 1676–1684. doi:10.1016/S1388-2457(01)00607-1

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools, Inc.

Sheeran, P., & Taylor, S. (1999). Predicting Intentions to Use Condoms: A Meta-Analysis and Comparison of the Theories of Reasoned Action and Planned Behavior1. *Journal of Applied Social Psychology*, *29*(8), 1624–1675. doi:10.1111/j.1559-1816.1999.tb02045.x

Sokolov, E. N. (1963). *Perception and the conditioned reflex*. Oxford, New York: Pergamon Press.

Spruyt, A., Houwer, J. D., & Hermans, D. (2009). Modulation of automatic semantic priming by feature-specific attention allocation. *Journal of Memory and Language*, *61*, 37–54. doi:10.1016/j.jml.2009.03.004

Stern, R. M., Ray, W. J., & Quigley, K. S. (2001). *Psychophysiological Recording* (Vol. 2nd). New York: Oxford University Press.

Stone, J. V. (2002). Independent component analysis: an introduction. *Trends in Cognitive Sciences*, *6*(2), 59–64. doi:10.1016/S1364-6613(00)01813-1

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics*. Pearson Education, Limited.

Taylor, J. H. (2010). Examination of the cognitive mechanisms underlying evaluative and semantic priming effects by varying task instructions: An ERP study. Unpublished dissertation. University of Texas at El Paso, El Paso, TX.

Verschuere, B., Ben-Shakhar, G., & Meijer, E. (Eds.). (2011). *Memory detection: Theory and Application of the concealed information test*. New York, NY: Cambridge University Press.

Wang, J., Kitayama, S., & Han, S. (2011). Sex difference in the processing of task-relevant and task-irrelevant social information: An event-related potential study of familiar face recognition. *Brain Research*, *1408*, 41–51. doi:10.1016/j.brainres.2011.05.060

Weymar, M., Bradley, M. M., El-Hinnawi, N., & Lang, P. J. (2013). Explicit and Spontaneous Retrieval of Emotional Scenes: Electrophysiological Correlates. *Emotion*, No Pagination Specified. doi:10.1037/a0033109

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482. doi:10.1002/cne.920180503

Appendix A

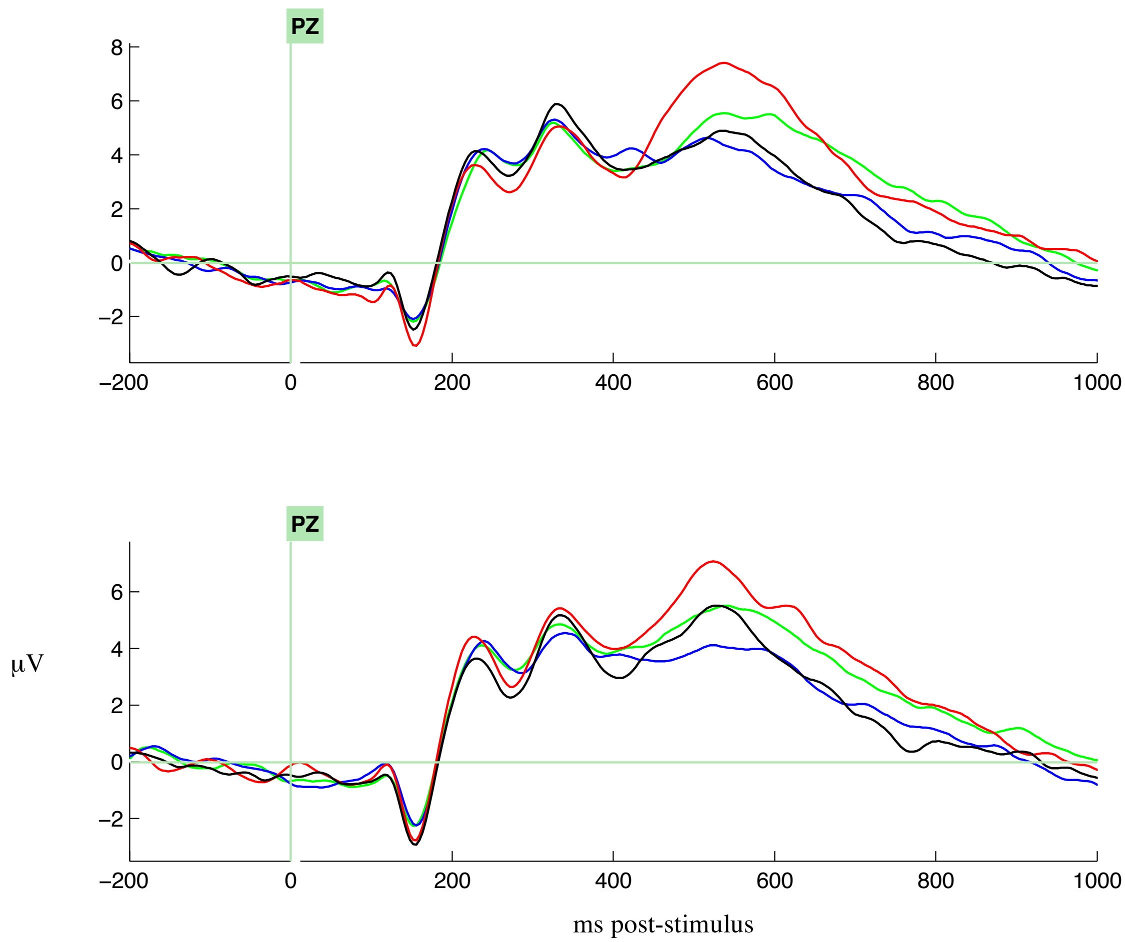


Figure A1: Grand average reflecting a larger evaluative congruity effect for key pressing than counting during control trials (top). There was no difference between response tasks for probe trials (bottom).

Appendix B

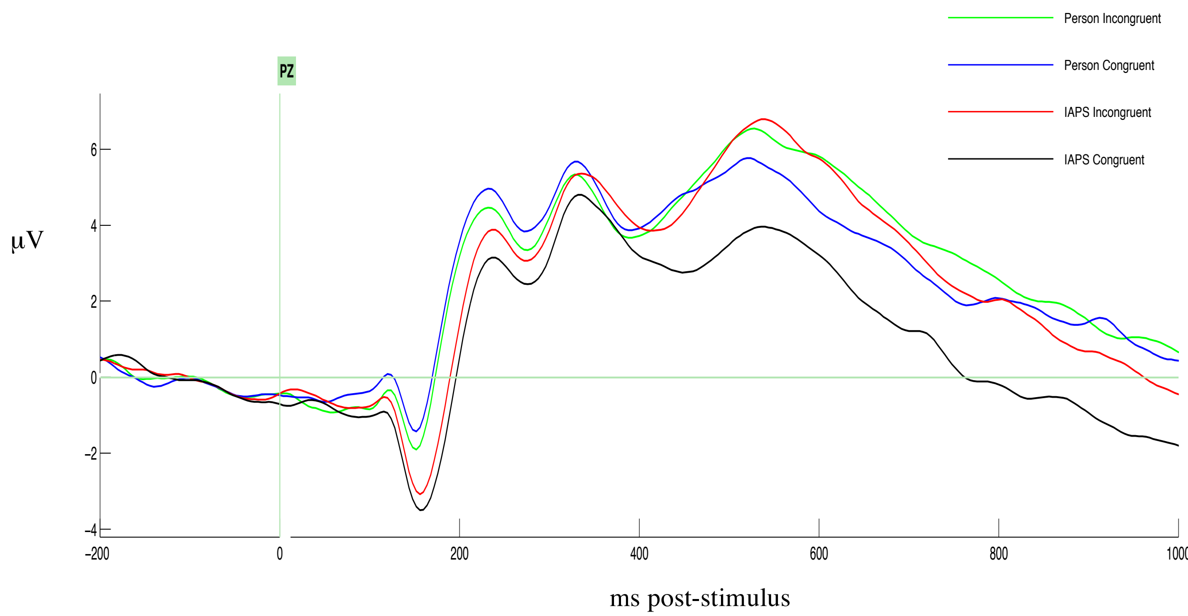


Figure B1: Grand average (Pz) reflecting a larger evaluative congruity effect for IAPS than celebrities.

Appendix C

The Bayesian approach adapted to the present CAT is presented in general, and an intra-individual approach originally proposed by Allen and colleagues (1992) is presented in particular. Note that the Allen et al. intra-individual approach was applied to the current CAT, however, because this method was suboptimal we report above (see p. 32) an alternative approach that uses additional electrode sites as multiple indicators as opposed to the Allen et al. (1992) approach presented below.

Bayesian Classification

Because the ANOVA approach is intended for making group comparisons and is thus inappropriate for questions at the individual level, Bayes' rule was used to determine whether participants' ERPs were evaluatively incongruent with context pictures (Bayes, 1763; Kline, 2004; Kruschke, 2010; Maris, 2012). In short, Bayes' rule determines the probability of a belief (i.e., posterior; I|L) given a set of conditions: the prior (I), evidence (L), and likelihood (L|I). The basic form of Bayes' rule is:

$$p(I|L) = \frac{p(L|I)p(I)}{p\ (L)} \quad (1)\ .$$

Suppose an evaluative oddball paradigm is used in which evaluatively incongruent pictures are 20% probable and evaluatively congruent (i.e., context and context congruent) pictures are 80% probable. Further, suppose it is known that 90% of the time incongruent pictures produce "large" LPPs (i.e., the sensitivity) and the overall probability of a large LPP is 25%. The *prior* (I = 20%) indicates that without any LPP data the probability of an incongruent picture is 20%. The *likelihood* indicates the probability of a large LPP given an incongruent trial is 90%; multiplied by the prior this numerator is the probability of ERPs to incongruent pictures that produce large LPPs (.90 * .20 = .18). Weighted by the *evidence* or

all ERPs that display a large LPP (.25), the posterior belief (or posterior probability) that an incongruent LPP is large is 72%. Bayes' rule in this example is thus expressed[14]:

$$p(I|L) = \frac{p(.90)p(.20)}{p(.25)} = .72 \qquad (2).$$

To compute the posterior probability, ERP averages submitted to Bayes' rule must accurately reflect the prior (or base rate). As seen in Table 1 (see column 4), the present study has 10 conditions—context and context congruent trials consist of 83.3% of trials (504+24*4 = 600/720) and context incongruent trials consist of 16.6% of trials (24*5 = 120/720). Considering these priors, possible ratios of incongruent to congruent (i.e., context + context congruent) averages submitted to Bayes' rule could be 1:6, 2:12, 4:24, et cetera. For the present CAT, two incongruent averages, one control and one probe, and 10 context averages were used. (Note that this is the reason for 10 columns in Tables 3 and C1 for the context.) Each average reflected the same number trials (48), stimuli (4), and repetitions (12). In this way, differences among posteriors are not likely attributable to differences among the averages' signal-to-noise ratios.

The virtue of the Bayesian approach is the inclusion of multiple indicators. The above example of Bayes' rule included only a single characteristic of the LPP (e.g., peak amplitude) as an indicator for determining the probability of an incongruent picture given a "large" LPP. Ideally, however, one would have multiple indicators for determining whether a picture was incongruent because increasing the number of indicators increases the posterior probability. In addition to LPP peak amplitude, we initially used four additional empirically validated indicators (Allen et al., 1992) that captured unique aspects of each participant's ERPs: 1) the entire waveform (i.e., 451 time points from epochs -300-1500 ms), 2) the first derivative or slope of each time point, 3) the second derivative or rate of change in the slope at each

---

14 Alternatively, one could rework this formula to its extended form with knowledge that context and context congruent pictures have a complement of the specificity equal to 8.75% (1-likelihood of 91.25%):

$$(pI|L) = \frac{p(.90)p(.20)}{p(.90)p(.20) + p(.0875)p(.80)}.$$

Note that knowledge of both the sensitivity (.90) and the complement of specificity (.0875) multiplied by their respective priors (.20 & .80) give rise to L (= .25).

time point, and 4) the deviation or grand mean of the 12 variables subtracted from each individual average. With the addition of these four indicators to LPP amplitude, Bayes' rule can be extended to ask what the probability is an ERP is in response to an incongruent picture given a combination of indicators? Greater consistency among indicators achieves greater posterior probabilities.

Because each of the latter four indicators involved 451 time points by 12 averages per participant, we used an unrotated principal component analysis (PCA; see Tabachnick & Fidell, 2007) to reduce dimensionality and maximize the variance explained for each of the three a priori components (i.e., control incongruent, probe incongruent, and congruent). Once each of the four indicator's dimensionality was reduced for each participant, the loading matrix or relations among the observed variables and extracted components were used to distinguish incongruent from congruent trials. Specifically, were used the sum of the squared loadings (i.e., communalities) of the first three principal components to discriminate incongruent and congruent trials.

Finally, prior to computing the posterior probability, cutpoints for classifying "large" ERPs and likelihood estimates (i.e., sensitivities and specificities) were obtained for each indicator. The cutpoint was obtained by *z* scoring each participant's five indicators (i.e., four communalities and one LPP peak amplitude at Pz) to remove between-subject differences in ERPs. An algorithm was then used that optimized the *z* score cutpoints to define large ERPs so the sensitivity was maximized over the specificity at the same cutpoint. These estimates were obtained through cross-validation; half the sample was randomly drawn and used as a training sample so these sensitivities, specificities, and z cutpoints could then be used to compute the posterior probabilities of the whole (validation) sample (Alqallaf & Gustafson, 2001).

The results of the Allen et al. (1992) intra-individual approach applied to the current CAT are presented in Tables C1 and C2. Posteriors fell either above 64% or below 55%. Considering a posterior greater than 64% to be evidence that an ERP was in response to an incongruent picture, across indicators

the sensitivity was 37.5% (39/104) and specificity was 94.8% (493/520). Because this approach was deemed unsuccessful for differentiating incongruent pictures from the context, we report in the main text an approach that makes use of multiple electrodes instead of using the entire waveform and transformations thereof (i.e., the derivatives and deviation).

Table C1: Posterior Probabilities by Participant, Response Condition, Evaluatively Incongruent Controls and Probes, and Context (across Allen et al.'s indicators).

Sub	RT	Con	Pro	Context 1	Context 2	Context 3	Context 4	Context 5	Context 6	Context 7	Context 8	Context 9	Context 10
1	Counting	0.646	0.036	0.031	0.017	0.02	0.268	0.017	0.094	0.268	0.036	0.094	0.017
2	Counting	0.681	0.507	0.02	0.036	0.02	0.031	0.017	0.094	0.646	0.02	0.055	0.094
3	Key Press	0.507	0.507	0.02	0.064	0.036	0.02	0.064	0.094	0.094	0.031	0.031	0.017
4	Key Press	0.394	0.108	0.017	0.017	0.268	0.017	0.094	0.055	0.094	0.02	0.036	0.268
5	Counting	0.239	0.239	0.055	0.031	0.02	0.017	0.094	0.108	0.094	0.02	0.036	0.036
6	Counting	0.545	0.681	0.017	0.02	0.017	0.239	0.268	0.545	0.055	0.094	0.036	0.108
7	Counting	0.646	0.646	0.094	0.017	0.02	0.02	0.268	0.358	0.094	0.239	0.02	0.02
8	Key Press	0.268	0.239	0.02	0.031	0.02	0.394	0.017	0.108	0.055	0.031	0.031	0.094
9	Counting	0.646	0.108	0.02	0.031	0.02	0.108	0.108	0.02	0.545	0.036	0.094	0.031
10	Key Press	0.646	0.681	0.017	0.268	0.064	0.02	0.036	0.02	0.094	0.394	0.055	0.055
11	Counting	0.646	0.681	0.108	0.017	0.036	0.239	0.094	0.646	0.108	0.017	0.02	0.02
12	Key Press	0.681	0.646	0.02	0.017	0.031	0.02	0.108	0.646	0.507	0.02	0.036	0.02
13	Key Press	0.681	0.646	0.017	0.031	0.055	0.02	0.017	0.064	0.646	0.036	0.055	0.036
14	Counting	0.108	0.02	0.646	0.017	0.02	0.031	0.507	0.545	0.268	0.094	0.094	0.094
15	Counting	0.02	0.394	0.646	0.036	0.055	0.394	0.358	0.02	0.031	0.239	0.017	0.094
16	Key Press	0.681	0.02	0.268	0.031	0.02	0.017	0.055	0.108	0.094	0.036	0.064	0.055
17	Counting	0.507	0.681	0.031	0.055	0.02	0.646	0.02	0.358	0.031	0.036	0.064	0.268
20	Key Press	0.646	0.268	0.02	0.108	0.036	0.017	0.017	0.017	0.094	0.02	0.017	0.646

21	Counting	0.268	0.268	0.646	0.681	0.031	0.031	0.507	0.094	0.094	0.036	0.017	0.036
22	Key Press	0.239	0.646	0.094	0.017	0.02	0.064	0.681	0.507	0.358	0.02	0.031	0.02
23	Key Press	0.646	0.239	0.108	0.094	0.02	0.507	0.02	0.108	0.507	0.02	0.02	0.036
24	Counting	0.036	0.681	0.545	0.094	0.094	0.239	0.064	0.017	0.031	0.02	0.02	0.036
25	Key Press	0.268	0.239	0.017	0.02	0.036	0.094	0.108	0.108	0.646	0.02	0.017	0.017
26	Counting	0.681	0.064	0.064	0.02	0.036	0.055	0.646	0.031	0.017	0.02	0.036	0.02
27	Key Press	0.646	0.646	0.094	0.055	0.02	0.02	0.02	0.036	0.02	0.268	0.094	0.031
30	Counting	0.507	0.094	0.358	0.031	0.017	0.017	0.268	0.358	0.646	0.02	0.02	0.064
31	Counting	0.094	0.268	0.239	0.017	0.094	0.094	0.507	0.268	0.036	0.094	0.108	0.02
32	Key Press	0.646	0.094	0.239	0.268	0.017	0.031	0.031	0.681	0.02	0.108	0.02	0.064
33	Key Press	0.108	0.545	0.031	0.017	0.646	0.064	0.036	0.055	0.036	0.02	0.094	0.064
34	Key Press	0.094	0.108	0.036	0.02	0.064	0.239	0.239	0.646	0.394	0.108	0.02	0.055
35	Counting	0.545	0.681	0.646	0.094	0.017	0.239	0.239	0.094	0.02	0.031	0.064	0.02
36	Key Press	0.239	0.239	0.268	0.036	0.02	0.094	0.094	0.02	0.108	0.036	0.031	0.036
37	Counting	0.02	0.646	0.031	0.02	0.108	0.017	0.094	0.108	0.108	0.02	0.017	0.02
39	Counting	0.358	0.394	0.239	0.02	0.02	0.094	0.055	0.02	0.094	0.064	0.031	0.017
40	Key Press	0.681	0.108	0.017	0.031	0.094	0.646	0.545	0.02	0.02	0.017	0.017	0.394
41	Counting	0.268	0.268	0.031	0.055	0.055	0.108	0.02	0.02	0.646	0.094	0.036	0.02
43	Counting	0.545	0.545	0.031	0.239	0.094	0.031	0.02	0.036	0.681	0.036	0.239	0.094
44	Counting	0.094	0.094	0.02	0.017	0.031	0.507	0.239	0.036	0.268	0.02	0.02	0.02
45	Key Press	0.646	0.268	0.064	0.017	0.394	0.017	0.036	0.646	0.031	0.017	0.02	0.094

46	Key Press	0.646	0.268	0.031	0.268	0.02	0.031	0.02	0.017	0.055	0.036	0.02	0.094
48	Key Press	0.545	0.646	0.02	0.017	0.036	0.239	0.02	0.108	0.036	0.094	0.036	0.108
49	Counting	0.094	0.055	0.358	0.02	0.036	0.017	0.017	0.681	0.545	0.268	0.02	0.094
50	Counting	0.646	0.545	0.031	0.031	0.031	0.268	0.02	0.108	0.017	0.017	0.017	0.017
51	Key Press	0.681	0.646	0.681	0.036	0.02	0.036	0.055	0.268	0.02	0.031	0.017	0.358
52	Counting	0.358	0.507	0.108	0.02	0.017	0.681	0.02	0.094	0.017	0.094	0.02	0.036
53	Counting	0.394	0.394	0.02	0.017	0.036	0.017	0.094	0.108	0.108	0.268	0.031	0.055
54	Key Press	0.646	0.646	0.268	0.036	0.017	0.031	0.094	0.108	0.02	0.02	0.036	0.017
55	Key Press	0.545	0.02	0.394	0.02	0.02	0.031	0.094	0.358	0.036	0.02	0.017	0.094
56	Key Press	0.681	0.268	0.02	0.064	0.094	0.02	0.055	0.108	0.031	0.02	0.02	0.646
57	Key Press	0.094	0.239	0.02	0.031	0.036	0.036	0.094	0.681	0.094	0.545	0.02	0.036
58	Key Press	0.681	0.646	0.017	0.017	0.017	0.036	0.239	0.394	0.268	0.036	0.094	0.108
59	Key Press	0.681	0.268	0.094	0.094	0.017	0.036	0.017	0.268	0.108	0.036	0.017	0.017

Note. There are 10 context averages to accurately reflect the base rate of these trials (i.e., 10/12 or 83.3%). See above appendix for details. Original participant numbers are displayed above (e.g., 59), but only the 52 accepted participants' data are presented above and were analyzed. Sub = subject; RT = Response task; Con = Controls; Pro = Probes.

Table C2: Sensitivities and Specificities by Indicators (Allen et al.'s indicators).

Indicator	Sensitivity	Specificity	Z Cutpoint
Original (*h^2*)	.66	.65	.084
Pz	.77	.80	.56
1st Derivative (*h^2*)	.47	.50	-.086
2nd Derivative (*h^2*)	.49	.47	-.215
Deviation (*h^2*)	.59	.53	-.21

Note. Percentage of evaluatively incongruent averages classified correctly (i.e., sensitivity) and percentage of context averages classified correctly (i.e., specificity) by electrode indicators. Z cutpoint is the *z*-score that maximally separated evaluatively incongruent from context averages (see above appendix for details).

Appendix D

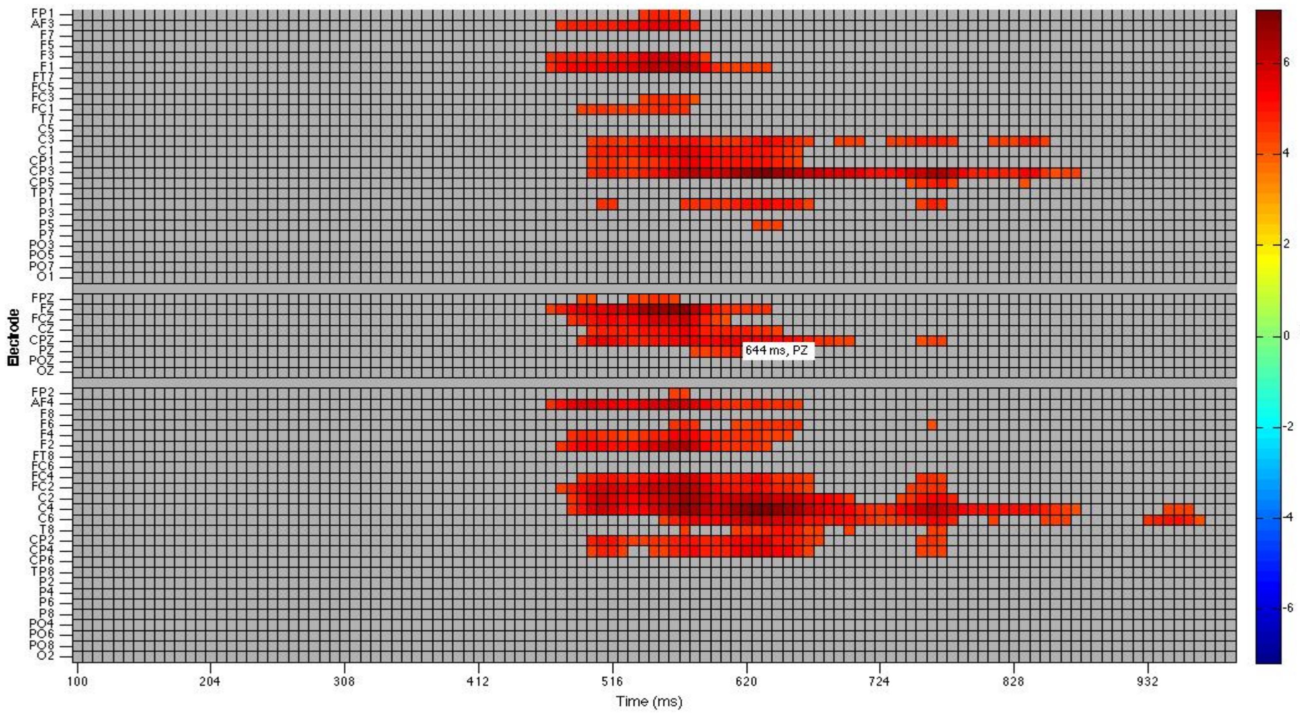


Figure D1: Raster diagram across 60 electrodes and 113 time points. Colored boxes indicate statistically significant evaluative congruity effects (family-wise error = .05). The scale to the right represents *t* values.

Vita

David R. Herring was born in Columbus, Georgia April 18th 1983. David graduated from Paradise Valley High School in 2001 (Phoenix, AZ) and with a B.A. in Psychology in 2007 from Arizona State University (ASU). David worked under the direction of Mary H. Burleson at ASU, publishing his honors thesis with Burleson in the *International Journal of Psychophysiology*. Under the mentorship of Stephen L. Crites, David first authored two articles, one in *Emotion* and the other in *Psychological Bulletin*. David also received a Department of Homeland Security Fellowship, Society for Psychophysiological Research Training Fellowship, and National Science Foundation Dissertation Grant while working with Crites. David is now a Postdoctoral Association under the mentorship of Peter J. Lang, Margaret M. Bradley, and Andreas Keil at the Center for the Study of Emotion and Attention at the University of Florida.

Permanent address: Center for the Study of Emotion & Attention PO Box 112766
Gainesville, FL 32608.

This thesis/dissertation was typed by David R. Herring.