Psychological Bulletin

On the Automatic Activation of Attitudes: A Quarter Century of Evaluative Priming Research

David R. Herring, Katherine R. White, Linsa N. Jabeen, Michelle Hinojos, Gabriela Terrazas, Stephanie M. Reyes, Jennifer H. Taylor, and Stephen L. Crites, Jr.

Online First Publication, January 21, 2013. doi: 10.1037/a0031309

CITATION
Evaluation is a fundamental concept in psychological science. Limitations of self-report measures of evaluation led to an explosion of research on implicit measures of evaluation. One of the oldest and most frequently used implicit measurement paradigms is the evaluative priming paradigm developed by Fazio, Sanbonmatsu, Powell, and Kardes (1986). This paradigm has received extensive attention in psychology and is used to investigate numerous phenomena ranging from prejudice to depression. The current review provides a meta-analysis of a quarter century of evaluative priming research: 73 studies yielding 125 independent effect sizes from 5,367 participants. Because judgments people make in evaluative priming paradigms can be used to tease apart underlying processes, this meta-analysis examined the impact of different judgments to test the classic encoding and response perspectives of evaluative priming. As expected, evidence for automatic evaluation was found, but the results did not exclusively support either of the classic perspectives. Results suggest that both encoding and response processes likely contribute to evaluative priming but are more nuanced than initially conceptualized by the classic perspectives. Additionally, there were a number of unexpected findings that influenced evaluative priming such as segmenting trials into discrete blocks. We argue that many of the findings of this meta-analysis can be explained with 2 recent evaluative priming perspectives: the attentional sensitization/feature-specific attention allocation and evaluation window perspectives.

Keywords: attitudes, meta-analysis, priming, affect, evaluation

The notion that researchers can measure a person’s automatic evaluative reaction in addition to a more deliberative one to the same . . . object has proved to be one of the most transformative in the history of the field. (Briñol & Petty, 2012, p. 303)

Evaluation refers to the operations of differentiating good from bad (Cacioppo & Berntson, 1994; Eagly & Chaiken, 1993). The omnipresent nature of evaluation has led to its recognition as a core psychological process. Early cross-cultural research showed that evaluation is perhaps the single most important dimension of meaning (see Osgood, 1952; Osgood & Tannenbaum, 1955; Triandis & Osgood, 1958). The process of evaluation is closely linked with fundamental behavioral mechanisms to approach and avoid. Evaluation and subsequent approach–avoidance behavior is so
fundamental that it can be found in many circuits across multiple levels of the nervous system (for reviews, see Berntson, Boysen, & Cacioppo, 1993; Norman et al., 2011). Low-level spinal circuits, for example, evaluate a stimulus and then initiate an appropriate approach–avoidance behavior (e.g., stimulation of pain receptors can lead to withdrawal). Evaluation is also a fundamental aspect of affect and attitudes, which arise due to activity at subcortical and cortical areas (Cunningham & Zelazo, 2007; LeDoux, 1995). The evaluations associated with affect and attitudes help guide many complex behaviors such as comforting people who are suffering, deciding which food to order at a restaurant, and voting for a presidential candidate.

Attitudes are conceptualized as evaluative summaries of information about an object. Notably, attitudes have traditionally been assessed with self-report measures (see Krosnick, Judd, & Wittenbrink, 2005, for review). Such measures require respondents to explicitly articulate their evaluation of an object or issue; for example, respondents might be asked to evaluate “legalizing marijuana” using a scale from 1 (strongly disagree) to 7 (strongly agree). The object can be any concrete or abstract entity—President Obama, marijuana, eating pizza, Mexicans, physician-assisted suicide, poverty, et cetera. Multiple meta-analyses have revealed that self-report assessments of attitudes predict a wide variety of behaviors (Glasman & Albarracín, 2006; Kraus, 1995). Although research with self-report measures has been and continues to be important, there are limitations with self-report measures that led researchers to explore other measures of evaluation.

One limitation of self-report evaluation measures is their reliance on explicit and deliberative processes. Respondents are asked to consider an object or item, weigh the various pros and cons associated with that item, and then translate their overall evaluation to a value on a rating scale. The validity of self-report assessments can be compromised if respondents decide that they do not want to report their true evaluation and thus misreport it. For example, an applicant of the FBI who thinks that legalizing marijuana is a good idea might hesitate to reveal her true attitude and instead report the opposite attitude. Another limitation is that most explicit self-report measures require a deliberative process and thus may fail to capture more spontaneous and initial aspects of an evaluation that might be important for guiding certain behaviors. For example, when asked to evaluate President Obama, a respondent may have an immediate positive evaluation (“He is intelligent and articulate”) that is then replaced by a more negative evaluation based on a deliberative consideration of his domestic and foreign policies (e.g., disliking increased involvement in Afghanistan). In the mid-1980s, researchers extended theory and methods from cognitive and memory research to attitudes in an attempt to minimize more deliberative processes that might lead people to misreport their attitude or override initial evaluations.

Fazio, Sanbonmatsu, Powell, and Kardes (1986) made a significant contribution to attitude research by developing an attitude priming paradigm similar to that used in memory research. This advancement led to a paradigm shift because it allowed for more implicit attitude assessments that relied on reaction times. This enabled evaluative processes to be examined that were difficult to study with more explicit and deliberative self-report attitude measures. Fazio et al.’s attitude priming paradigm was adapted from the sequential semantic priming procedure originally used in studies of memory (for reviews, see McNamara, 2005; Neely, 1991). In sequential semantic priming, people view a sequence of stimuli pairs in which the semantic or associative relation between the first (prime) and second (target) stimuli in a pair is manipulated. People respond to the target stimulus (e.g., dog), and their responses are typically faster when the target is semantically or associatively congruent with the prime (e.g., cat) compared to incongruent (e.g., chair). Fazio et al. altered the sequential semantic priming paradigm by varying evaluative congruity between the prime and target instead of semantic and associative congruity. This revealed that people respond more quickly that target adjectives (e.g., repulsive) are good or bad when they were preceded by evaluatively congruent (e.g., spider) compared to incongruent primes (e.g., party). This study was significant because it suggested that attitudes could quickly be brought to mind with little deliberation, potentially biasing the way people view and behave in their surroundings.

Because evaluation on a good/bad dimension is an important aspect of many constructs other than attitudes, there has been an explosion of research using evaluative priming as illustrated in Figure 1. This increase is due in part to evaluative priming being extended to study a wide range of topics including affect, ageism, alexithymia, anorexia, approach/avoidance, attachment, depression, emotion, exercise behaviors, food preferences, ingroups/outgroups, lower back pain, positive schizotypy, prejudice, post-traumatic stress disorder, schizophrenia, stereotypes, and tobacco smokers (see Asgaard, Gilbert, Malpass, Sugai, & Dillon, 2010; Dannowski et al., 2006; Degner & Wentura, 2010; Dovidio, Evans, & Tyler, 1986; Eves, Scott, Hoppé, & French, 2007; Fazio, Jackson, Dunton, & Williams, 1995; Goubert, Crombez, Hermans, & Vanderstraeten, 2003; Kerns, 2005; Milanak & Berenbaum, 2009; Perdue, Dovidio, Gurtman, & Tyler, 1990; Perdue & Gurtman, 1990; Roefs et al., 2005; Scherer & Lambert, 2012; Suslow, Arolt, & Junghanns, 1998; Suslow, Dannowski, Arolt, & Ohmann, 2010; Suslow, Roestel, Droste, & Arolt, 2003; Vandebosch & De Houwer, 2011; Veldhuizen, Oosterhoff, & Kroene, 2010; Vermeulen, Luminet, & Corneille, 2006; Weisbuch & Ambady, 2008).

Although there have been at least 10 narrative reviews of the evaluative priming literature (for reviews, see De Houwer, Teige-Mocigema, Spruyt, & Moors, 2009; Fazio, 2001; Fazio & Olson, 2003; Ferguson & Bargh, 2003; Klauser, 1998; Klauser & Musch,
ON THE AUTOMATIC ACTIVATION OF ATTITUDES

2003; Spruyt, Gast, & Moors, 2011; Wentura & Degner, 2010; Wentura & Rothermund, 2003; Wittenbrink, 2007), there has not been a complete meta-analytic review. There have been four “partial” meta-analyses focused on specific issues using evaluative priming data, but these studies did not aim to synthesize the entire evaluative priming literature (Cameron, Brown-Iannuzzi, & Payne, 2012; Dijksterhuis & Aarts, 2003; Hofmann, De Houwer, Peregini, Baeyens, & Crombez, 2010; Unkelbach, Fielder, Bayer, Stegmüller, & Danner, 2008). Although two of these four studies computed an average effect size for the evaluative priming effect (i.e., difference in reaction time between evaluatively congruent and incongruent targets), these effect sizes are not representative of the evaluative priming literature because these investigations selectively chose certain evaluative priming studies to examine other theoretical issues. In addition, these investigations did not examine a number of important moderator variables of the evaluative priming effect (see below). Thus, little can be gleaned from these partial meta-analyses concerning the overall evaluative priming effect and the key moderators of evaluative priming. Because researchers have proposed different theoretical explanations for evaluative priming, a meta-analysis that carefully explores important moderators of evaluative priming would be useful for clarifying the theoretical underpinnings of evaluative processes. Researchers initially adopted spreading activation theory to explain evaluative priming (see Klauer & Musch, 2003). Spreading activation theory posits that stimuli activate representations in memory and that this activation spreads to related concepts or nodes (Anderson, 1983; Collins & Loftus, 1975), including concepts that are evaluatively related (Bower, 1991). However, as research accrued, researchers realized that spreading activation theory did not provide a good explanation for the entire set of findings (Klauer, Roßnagel, & Musch, 1997). Other theoretical explanations for evaluative priming (e.g., due to response processes as described below) were developed to help better explain the pattern of findings across different studies. The prominence and viability of various theoretical explanations of evaluative priming has waxed and waned as researchers employed slightly different variations of the evaluative priming paradigm and findings accumulated.

The most common variation of the evaluative priming paradigm is the evaluative decision task (EDT), in which participants indicate whether a target stimulus is good or bad (for reviews on structure, see Berthet, Kop, & Kouider, 2011; De Houwer, 2003; Schmitz & Wentura, 2012). One issue with the EDT is that there are two types of congruity on every trial that are confounded: evaluative congruity between the prime and target and congruity between the prime and response. For instance, when the prime and target are evaluatively congruent (e.g., saint-delightful), the prime is evaluatively congruent with both the target and the required response “good” to the target. When the prime and target are evaluatively incongruent (e.g., saint-repulsive), the prime is evaluatively incongruent with both the target and the required response to the target. This raises theoretical questions about what is causing reaction time differences in the EDT. Is the congruity between the prime and target influencing how easily the target can be encoded, and/or is the congruity between the prime and response influencing how quickly a response to the target can be executed?

There are two theoretical perspectives that have been used to account for priming in the EDT. The earliest theoretical perspective on evaluative priming borrowed from semantic priming theory by using an analogy of spreading activation to account for the findings (for review, see Fazio, 2001, 2007; Fazio et al., 1986). According to this encoding perspective, prime stimuli activate object—evaluation associations in memory that make the valence of targets more accessible, facilitating evaluatively priming. Another explanation of evaluative priming in the EDT is that the prime influences the ease with which a person can generate a response to the target, similar to that which occurs in the classic Stroop paradigm. That is, in a classic Stroop task, participants view a word (e.g., blue) written in either the same (blue) or different (green) ink color and state the ink color of the word (MacLeod, 1991; Stroop, 1935). Participants typically respond slower when the ink color does not match the word meaning because the word activates an incompatible response that must be inhibited in order to make the correct response. According to this response perspective of evaluative priming, even though people do not have to respond to the prime, it activates a particular response (e.g., “good”) that is either congruent or incongruent with the response activated by the target (see Klauer & Musch, 2003). When the activated response to the prime is congruent with the target, people respond faster, but when the activated response to the prime is incongruent with the target, they respond slower. In sum, evaluative priming in the EDT task can be explained by an encoding perspective that postulates the prime facilitates encoding of the target and by a response perspective that postulates the prime facilitates or inhibits the response made to the target.

To help disentangle theoretical explanations for evaluative priming in the EDT, researchers have employed other variations of the sequential evaluative priming paradigm that do not require an evaluative response to the target. In these variations, the valence of the prime is irrelevant to the response to the target, so only theoretical perspectives that focus on how prime valence facilitates encoding of the target explain evaluative priming (i.e., the response perspective is eliminated). In a pronunciation or naming task, for instance, participants simply say aloud the name of the target (e.g., delightful) rather than evaluate it as “good” in the EDT. In this task, there is no confound between a response evoked by the prime and that required by the target as in the EDT. That is, in the EDT, a prime might activate the response “good” that can be either congruent with the response evoked by a positive target or incongruent with the response evoked by a negative target. In the pronunciation task, however, any response evoked by the prime (e.g., initiation of pronunciation) is equivalent regardless of whether the target is evaluatively congruent or incongruent with the prime. For example, seeing the word party as a prime will activate the same response when the target delightful or the target hateful is presented. Thus, by changing the judgment people make in response to targets in evaluative priming tasks, researchers have

1 The aim of Dijksterhuis and Aarts (2003) was to test a negativity bias of affective processing by comparing the differences in response times for negative and positive primes. Unkelbach et al. (2008) used evaluative priming data to show a processing advantage for positive evaluative information. Hofmann et al. (2010) examined evaluative priming data from studies in which primes were evaluatively conditioned (see Prime Conditioned section below) to demonstrate a range of evaluative conditioning effects in a meta-analytic review. And Cameron et al. (2012) examined the relation between evaluative priming and behavioral intentions as well as explicit measures of attitudes.
been able to isolate and examine different theoretical explanations for evaluative priming (discussed below).

The present meta-analysis examines a number of variables, such as the judgment people make to the target, in evaluative priming research that have important theoretical implications. These variables are briefly reviewed below.

**Judgment Task**

As discussed, a principal difference across evaluative priming studies is the judgment people make about targets. The judgment task most widely employed is the EDT in which participants indicate the valence of targets. An EDT was used in Fazio et al.’s (1986) initial demonstration of evaluative priming. In Experiment 1, for instance, primes were idiosyncratic liked and disliked attitude objects that were identified in a preliminary phase of the study, and targets were adjectives that were positive (e.g., appealing, delightful) and negative (e.g., repulsive, awful). Participants indicated the evaluative connotation of the target adjective as quickly and accurately as possible. The EDT has been by far the most common task used in the literature, and the present meta-analysis includes 76 effect sizes (from 72 experiments and 51 publications) that used an EDT. Although the EDT has been widely used, other judgment tasks have also been employed to help elucidate the processes underlying evaluative priming.

As discussed, one reason other judgment tasks have been used is because the inherent confound in the EDT between prime–target and prime–response congruity makes it difficult to discern whether encoding or response processes are responsible for evaluative priming. Another reason is because the EDT does not allow one to examine whether the valence of a stimulus can influence judgments when evaluation is irrelevant to a perceiver’s goals. That is, people have an explicit goal of evaluating each target stimulus in the EDT so evaluation is a central processing goal throughout the experiment. Even though people are not asked to evaluate the prime in a typical EDT, the temporal proximity of the prime and target may engender evaluation of the prime because it is not possible to limit the evaluation to just the target (Klauser, Teige-Mocigemba, & Spruyt, 2009). Thus, researchers have used different judgment tasks in order to (a) better ascertain whether evaluative priming is due to encoding or response processes and (b) explore whether evaluative processing occurs when evaluation is irrelevant to the perceiver’s processing goals. Three judgment tasks have been commonly employed in the literature: the pronunciation task, lexical decision task (LDT), and semantic categorization task.

After the EDT, the pronunciation or naming task is the most commonly used task in evaluative priming research. The present meta-analysis included 37 effect sizes (from 31 experiments and 20 publications) that employed the pronunciation task. In the pronunciation task, participants are asked to pronounce or name the target as quickly as possible by speaking into a microphone while the latency of their voice onset is recorded. For example, Bargh, Chaiken, Raymond, and Hymes (1996, Experiment 1) used a pronunciation task in an experiment that was similar to Fazio et al.’s (1986) original studies. The primes were idiosyncratic attitude objects, targets were adjectives with evaluative connotations, and participants named the target as quickly as possible.

Another judgment task used less frequently to examine the explicit goal of evaluation is the LDT. The present meta-analysis includes six effect sizes (from five experiments and four publications) that employed the LDT. In the LDT, participants indicate whether the target is a word or nonword as quickly as possible. A principal difference between the LDT and other evaluative priming tasks is that the LDT contains relevant trials in which targets are words and irrelevant trials in which targets are nonwords. Evaluative priming is examined by exploring whether the valence of a prime affects the latency with which people indicate a positive or negative target is a word. Nonword trials are irrelevant to theory and are typically not examined in the analyses because the valence of a prime should not facilitate nonword responses. These trials are included to ensure that participants attend to the relevant trials; if only word trials are used, participants could respond “word” after every trial without processing the meaning of the target. For example, Wentura (2000, Experiment 1) used a LDT that was similar to Fazio et al.’s (1986) original study: The primes were positive or negative nouns and the targets were positive or negative adjectives. In addition to these relevant trials, there were irrelevant trials in which either primes or targets were nonwords. Participants indicated whether the target was a word or nonword as quickly and accurately as possible.

The most recent task used to examine evaluative priming is a task in which participants categorize targets along a nonevaluative semantic dimension. The present meta-analysis includes six effect sizes (from five experiments and five publications) that employed semantic categorization tasks. Although the specific judgment tasks used in these studies varied, participants always categorized targets along a dimension of meaning that was nonevaluative. For instance, De Houwer, Hermans, Rothermund, and Wentura (2002) used a semantic categorization task in which both primes and targets were positive or negative nouns. The targets were nouns from one of two semantic categories—people (e.g., friend, snob) and animals (e.g., butterfly, cockroach), and primes were nouns that did not refer to people or animals (e.g., peace, death). Participants’ task was to indicate whether the target referred to a person or animal.

**Prime Strength**

Strength refers to the relative impact of an attitude on cognition and behavior—strong attitudes are more likely to influence behavior relative to weak attitudes (Glaser & Albarracin, 2006). For example, a person with a strong positive attitude toward the Dallas Cowboys is more likely to watch games on television relative to a person with a weak attitude. Similarly, a prime associated with a strong attitude (strong prime) might evoke greater priming relative to a prime associated with a weaker attitude (weak prime). There are various ways that strength has been conceptualized and measured in the attitude literature (Krosnick & Petty, 1995). Two of these conceptualizations, accessibility and extremity, have been used in the evaluative priming literature.

2 Although evaluation is a type of semantics, for simplicity we refer to nonevaluative semantic categorization as “semantic categorization” throughout this article.

3 Attitude accessibility and extremity are integrated because (a) evaluative priming studies manipulating prime strength use both conceptualizations and (b) these constructs are correlated (i.e., $r = -0.69$; Bargh et al., 1992; Krosnick, Boninger, Chuang, Berent, & Carnot, 1993).
refers to how quickly an attitude comes to mind when people encounter an exemplar of the attitude and is typically measured by recording the response latency associated with indicating one’s attitude to an exemplar. Attitude extremity is the magnitude of deviation from neutrality on an attitude scale (see Krosnick & Petty, 1995). Inspired by Fazio, Chen, McDonel, and Sherman’s (1982) perspective of attitudes as object–evaluation associations varying in strength in memory, a number of studies explored whether prime strength moderated the impact of evaluative priming (cf. Bargh, Chaiken, Raymond, & Hymes, 1996; Fazio et al., 1986; for reviews, see Klauer & Musch, 2003; Wittenbrink, 2007). In early evaluative priming reports, evaluative priming only occurred for strong prime attitudes, suggesting conditional evaluative priming consistent with Fazio’s attitude theory (Bargh, Chaiken, Govender, & Pratto, 1992, Experiment 1; Fazio et al., 1986). In later reports, however, evaluative priming occurred regardless of prime strength, suggesting unconditional evaluative priming (Bargh et al., 1992, Experiments 2 and 3; Chaiken & Bargh, 1993; Klauer et al., 2009; Spruyt, Hermans, Pandelaere, De Houwer, & Eelen, 2004). One study even found reverse evaluative priming (i.e., slower responses for evaluatively congruent stimuli) for strong primes (Glaser & Banaji, 1999). Given these disparate findings, a meta-analysis may help discern whether other variables moderate the impact of prime strength. Altogether, eight prime strength studies involving 28 effect sizes were examined in the present meta-analysis.

Stimulus-Onset Asynchrony (SOA)

SOA refers to the time between the onset of one stimulus and the onset of a subsequent stimulus. In evaluative priming studies, SOA is used to refer to the duration between the onset of a prime and the onset of the target. SOA has occasionally been manipulated to explore the impact of controllability on evaluative priming (Spruyt et al., 2011). In Fazio et al.’s (1986) initial article, for example, the SOA was manipulated (300 vs. 1,000 ms), and the findings revealed a congruity effect when the SOA was 300 ms but no effect when the SOA was 1,000 ms. Although subsequent studies have primarily used SOAs of either 250 or 300 ms, this meta-analysis includes studies with SOAs as short as 100 ms and as long as 1,000 ms.

Prime Type and Target Type

Prime type and target type refer to whether primes and targets were verbal or nonverbal (i.e., pictures, symbols). The lexical hypothesis (Glaser, 1992; Glaser & Glaser, 1989) proposes the existence of two internal coding systems: lexical and semantic. Glaser (1992) also suggested that pictures are more effective than words in priming studies because pictures have direct, functional connections to the semantic system, whereas words must first undergo processing in the lexical system before gaining access to the semantic system. Given that evaluative information is stored in the semantic system (Bower, 1981, 1991; De Houwer & Hermans, 1994), it is somewhat surprising to note the near absence of studies systematically examining the impact of pictures or words on evaluative priming. Two exceptions are studies conducted by Spruyt, Hermans, De Houwer, and Eelen (2002) and Zhang, Lawson, Guo, and Jiang (2006). Using a pronunciation task, Spruyt et al. observed significant evaluative priming when primes were pictures (Studies 1–3), nonsignificant priming when primes were words (Studies 2 and 3), and no significant effect for target type on evaluative priming (Study 3). Using an EDT, Zhang et al. found no moderating effect of prime type on evaluative priming. Although this suggests that prime (but not target) type significantly impacts evaluative priming, firm conclusions cannot be reached given that no other studies have systematically examined the impact of prime and target type on evaluative priming. One critical difference between these two lines of research is the type of judgment task that was used: Spruyt et al. used a pronunciation task, whereas Zhang et al. used an EDT. The present review will therefore explore whether judgment task moderates prime type.

With regard to prime type, analyses include 70 effect sizes (from 59 experiments and 37 publications) where prime stimuli were verbal and 49 effect sizes (from 44 experiments and 35 publications) where prime stimuli were nonverbal (pictures or symbols). In addition, there were six effect sizes (from six experiments and three publications) where combinations of both verbal and nonverbal prime stimuli were used. With regard to target type, the meta-analysis included 97 effect sizes (from 85 experiments and 59 publications) that used verbal target stimuli and 26 effect sizes (from 21 experiments and 16 publications) that used nonverbal stimuli.

Prime Conditioning

Prime conditioning refers to whether prime stimuli had an established valence prior to the experimental session or valence was created during the experimental session. In typical evaluative priming studies, prime stimuli have a clear positive or negative valence before the experimental session. However, in some evaluative priming studies, neutral stimuli were evaluatively conditioned prior to the priming session. Evaluative conditioning is a type of classical conditioning that occurs when the valence of a stimulus is formed or changed by pairing it with another valenced stimulus (De Houwer, Thomas, & Baeyens, 2001; Gast, De Houwer, & De Schryver, 2012; Hofmann et al., 2010). In evaluative priming studies that used evaluative conditioning, primes are conditioned stimuli that become positive or negative by pairing with positive or negative unconditioned stimuli. There are no studies we are aware of that examined evaluatively conditioned targets.

De Houwer, Hermans, and Eelen (1998) were the first to examine the effects of evaluatively conditioned primes in a series of studies. In this research, participants engaged in an evaluative conditioning procedure prior to the evaluative priming phase of the experiment. In the evaluative conditioning phase, nonwords were paired with words that were either positive or negative in order to establish a positive or negative valence in the nonwords. These evaluatively conditioned nonwords were then used as primes in subsequent evaluative priming procedures. The results of several studies revealed that evaluative priming occurred with evaluatively conditioned nonword primes. These results have also been replicated with nonverbal stimuli (e.g., pictures; Hermans, Baeyens, Lamote, Spruyt, & Eelen, 2005; Hermans, Spruyt, & Eelen, 2003; Spruyt, Hermans, Pandelaere, et al., 2004). The present review included 19 effect sizes (from 19 experiments and 15 publications) where prime stimuli were evaluatively conditioned.
Total Trials and Blocks

In sequential evaluative priming, a trial is a single prime–target pairing. The evaluative priming effect is examined by comparing evaluatively congruent trials with evaluatively incongruent trials. Because the effect of priming on reaction time is small relative to other variables that influence reaction time (e.g., type and complexity of stimulus, familiarity with stimuli, etc.), priming is detected by conducting multiple trials for the congruent and incongruent conditions and then aggregating across trials in each condition. The process of aggregating helps diminish “noise” associated with extraneous variables so the “signal” produced by priming can be detected. In the present meta-analysis, for example, the average number of total trials was 143 (ranging from 16 to 640). Although some research has manipulated the ratio of congruent to incongruent trials, the overwhelming majority of evaluative priming studies (and all studies in this meta-analysis) have a 50:50 ratio of congruent to incongruent trials (in the LDT, the 50:50 ratio refers only to relevant trials and does not include irrelevant trials in which nonwords are presented as targets). Thus, the number of congruent and incongruent trials that are compared to determine the priming effect are each half the total number of trials. Finally, because sequential priming typically involves many trials, researchers tend to separate the total trials into distinct blocks so participants can have short breaks during the course of the priming procedure. The average number of blocks were three, ranging from one to 10 in the present review.

The number of total trials and blocks is potentially relevant because the magnitude of the evaluative priming effect may change over the course of the experiment. The magnitude of evaluative priming, for instance, might increase due to learning from practice or decrease due to boredom or fatigue. It is plausible that both total trials and blocks might lead to gradual changes in priming over the course of the experiment. For example, performance in evaluative priming may gradually increase over time as a function of the number of total trials. Alternatively, the breaks between blocks might lead to increased performance because they allow consolidation of memories and reduce fatigue (Ariga & Lleras, 2011; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Delaney, Verkoejen, & Spiegel, 2010; Tucker, 2003). It is also possible that the effect of either total trials or blocks is moderated by other variables such as judgment task. For instance, the prime in incongruent trials in the EDT task is a “false signal” that must be ignored in order to initiate a correct response to the target. So, people may have more errors to incongruent trials in the EDT relative to the errors in other judgment tasks such as the LDT or pronunciation task. With practice in the EDT, people may develop strategies to diminish the impact of the prime on their response to the target, which may lead to diminished priming in the EDT across trials.

Prime Repetitions and Target Repetitions

Because many trials are required to examine evaluative priming, this requires either a large set of stimuli if every stimulus is presented minimally or a smaller set of stimuli that are presented multiple times. Prime repetitions and target repetitions refer to the number that any given prime and target stimuli were repeated across the experiment. For example, if a prime stimulus was used only once during the procedure (i.e., every trial had a unique prime stimulus), the number of prime repetitions would be zero. If a given prime stimulus was presented 20 times during the experiment, the number of prime repetitions would be 19. The number of prime repetitions and target repetitions ranged markedly in the present review (from 0 to 127 and from 0 to 20, respectively).

The numbers of prime and target repetitions is potentially relevant because they may impact the magnitude of evaluative priming. As stimuli are repeated and people become more familiar with them, this may affect how quickly a given stimulus is processed that might impact the magnitude of priming.

Other (Secondary) Variables

In addition to the above variables, we explored the effect of nine other variables that appear to have less theoretical relevance. Specifically, we examined whether the magnitude of evaluative priming was influenced by (a) instructions to pay attention to primes (prime instructions—pay attention to primes vs. ignore primes vs. no specific instructions), (b) response instructions to targets (response instructions—paid to respond accurately vs. error messages for incorrect responses vs. no additional instructions), (c) response time transformations (response time analysis—analyzed raw reaction times vs. transformed reaction times to make distribution normal), (d) response time facilitation scores (facilitation scores—neutral baseline vs. no baseline condition), (e) the part of speech of verbal primes (prime part of speech—nouns vs. adjectives vs. combinations of nouns and adjectives vs. verbs), (f) the part of speech of verbal targets (target part of speech—nouns vs. adjectives vs. combinations of nouns and adjectives vs. verbs), (g) the type of participants recruited (population—academics vs. non-academics), (h) whether the study was published (published—published vs. unpublished), and (i) whether prime and target stimulus sets were intermixed (stimulus presentation—same stimulus set used for both primes and targets vs. different stimulus sets for primes and targets). See the Appendix for more information about these variables.

The Present Meta-Analysis

The aims of the present investigation were to (a) examine the relative impact of the different tasks (e.g., EDT vs. nonevaluative tasks) to provide insight into the theoretical mechanisms of evaluative priming, (b) examine other important evaluative priming moderator variables to help resolve unanswered questions (e.g., prime strength), and (c) examine variables that have traditionally received little attention (e.g., blocks, prime and target repetitions), which might reveal effects that could lead to new research questions.

The emphasis of the current review is on potential differences among judgment task effect sizes. This approach is different from the more common approach of emphasizing the test of the weighted average effect size, which is expected to be significant in the present review because evaluative priming is generally considered robust. As discussed, there are multiple theoretical perspectives that can explain evaluative priming. These accounts are not mutually exclusive, and it is very likely that evaluative priming effects are due to multiple processes. Thus, estimating the magnitude of effect sizes across different judgment tasks, which have commonly been used to explore different theoretical perspectives,
is important for disentangling the various processes that underlie evaluative priming. To this end, we meta-analyzed 125 effect sizes from 73 evaluative priming studies (1986–2010) and coded 18 moderator variables.

**Method**

**Literature Review**

PsycINFO and PsycARTICLES via EBSCOhost were used to search for published studies between February 1986 (i.e., year of Fazio’s seminal article) and December 2010. A Boolean search was performed using the terms affective OR evaluative AND priming, with the “Select a Field (optional),” “select for related verbal,” “English,” and “Exclude Book Reviews” options selected. ProQuest was used to search for dissertations and theses with the same time frame and search terms. In ProQuest the search was qualified with the options “citation” and “abstract,” and document language was limited to “English.” For all three search engines, abstracts were first reviewed to determine whether there was clear reason the article should be excluded (see criteria below). If there was doubt whether to include the article, it was obtained for further review. Finally, we did two things to identify unpublished studies. We reviewed major literature reviews and looked for citations of unpublished work. We then wrote to these authors and specifically requested that they send us the unpublished study that was cited. We also contacted many authors to request additional information about their published studies (see below). When we contacted authors, we requested that they send us unpublished studies as well. In total, we contacted 49 authors and asked for unpublished articles; this included nearly all the authors who had been or were currently active evaluative priming researchers.

**Inclusion and Exclusion Criteria**

A set of criteria was devised to select studies for inclusion in the present meta-analysis. A study was defined as a published or unpublished experiment, or collection of experiments, in which an independent effect size was nested (as in Van den Bussche, Van den Noortgate, & Reynvoet, 2009; note that experiments could have multiple effect sizes so long as the effect sizes were independent). Studies were included in the review if at least one experiment in the study met all the criteria below (see Figure 2).

1. An evaluative sequential priming paradigm was used. This requirement eliminated studies that did not use sequential evaluative priming (i.e., prime before target) or used a sequential priming paradigm to examine nonevaluative, semantic priming (e.g., stereotypes). This requirement also removed studies that examined backward priming in which a target requiring a response was presented before a prime not requiring a response (Fockenberg, Koole, & Semin, 2006).

2. Each sequential priming trial contained foveal presentations of a single prime and a single target, neither of which was degraded or obscured.

3. Valence of primes and targets were equated across participants. This requirement removed studies that made inferences about a prime or target stimulus that was unknown in valence or differed in valence across participants. For example, some studies present known evaluative primes or targets with pictures of people from different racial categories to assess individual differences in racial attitudes (Fazio et al., 1995). Other studies present a novel target to examine whether people misattribute valence from the prime to the novel target (Payne, Cheng, Govoron, & Stewart, 2005). To be included, the valence of primes and targets had to be equated across participants prior to the sequential priming procedures by either using evaluatively normed stimuli assumed to be equivalent across people or using idiosyncratic evaluative stimuli that were assessed for each individual prior to the sequential priming paradigm.

4. Primes were presented between 100 and 300 ms, without masking, and required no response. This requirement eliminated paradigms that used very short prime presentations or masking to examine subliminal priming.

5. The SOA was not greater than 1,200 ms.

6. There was no temporal overlap between prime and target presentations. This requirement was implemented to reduce variance in evaluative priming that might arise from limitations with nonpriming related processes that occur when both prime and target were presented simultaneously. In the majority of sequential priming research, the prime is presented and then removed before the target is presented. There have been a few studies, however, in which the prime and target were presented simultaneously (i.e., SOA = 0) or temporally overlapped (Klauer & Musch, 2001; Klauer & Stern, 1992). Because these experiments may introduce other variables that are not present in most sequential priming studies (e.g., divided attention, saccadic eye movement required for foveal viewing of both), we did not include these studies.4

7. Targets required binary responses, and response times were measured. This requirement eliminated paradigms in which response times to targets were not measured or participants could make responses with more than two options (e.g., indicate whether target was positive, neutral, or negative).

8. The proportion of evaluatively congruent and incongruent trials and the proportion of positive and negative stimulus presentations were equivalent. This requirement eliminated studies that used an unequal number of congruent and incongruent trials or an

---

4 A potential problem with this criterion is that it may underestimate the impact of response competition on priming because it eliminates studies that evoke the greatest response competition (i.e., when both prime and target appear simultaneously). That is, very strong response competition effects are found in paradigms such as the Stroop task where the presentation of a single stimulus can activate two responses. The conceptually equivalent condition in priming occurs when two stimuli are presented simultaneously (i.e., SOA = 0) or perhaps temporally overlap. Thus, if response competition decreases as the temporal distance between the prime and target increases in sequential priming, not including studies that use temporally overlapping primes and targets might underestimate the magnitude of response competition. To investigate this possibility, we located studies that had overlapping prime and target presentations (n = 4, k = 7). Of these studies, only one study (k = 2) would have been included in the meta-analysis given our other inclusion criteria (e.g., necessary data not available; see also discussion below about maintaining statistical independence for within-subject designs). Reanalysis of judgment task with these additional effect sizes did not alter the results reported below. Reanalysis of SOA, however, lessened the impact of this moderator variable to marginal significance, Q(1) = 3.36, p = .06, potentially because the two effect sizes added to this analysis came from pronunciation tasks that at lower SOAs yield smaller effects relative to the EDT (e.g., see Figure 5). We thank Dirk Wentura for raising this issue and suggesting the judgment task by stimulus presentation interaction reported below.
 unequal number of positive and negative stimulus presentations. In the majority of evaluative priming research, there are four relevant trials (positive–positive, negative–negative, positive–negative, and negative–positive), and all four types are presented equally. Studies that altered the proportion of these trials (e.g., presented two congruent trial types 70% of the time and the two incongruent types 30% of the time) were not included.

9. Participants were in their natural state. This requirement eliminated studies that attempted to alter a state variable such as mood prior to or during the priming procedure.

10. Participants were not informed about evaluative priming prior to the priming procedure.

11. Participants were healthy adults. This requirement eliminated studies that focused on children, adolescents, or clinical samples (e.g., participants with mood or anxiety disorders).

12. Requisite statistical and methodological information (e.g., SOA duration) was either reported in the study or provided by the corresponding author or could be estimated based on information provided.

13. The article was written in English. We used this criterion for the practical reason that the authors’ language abilities were limited to English. Further, reviews of meta-analyses have revealed that excluding non-English articles has a negligible effect on findings (see Discussion section below).

Data, Study Information, and Unpublished Study Requests

There were several instances in which a key piece of information was not included in the original publication. Some studies, for
example, did not report statistics that could be used to compute the effect size for the evaluative congruity effect (e.g., used shorthand statistical notations such as “$F < 1$” or “$p < .05$”). Other studies omitted important methodological details such as prime duration or sample size. We contacted 49 authors for additional data, study information, or unpublished studies, and 37 replied (75%). Of these 37 authors, 20 (54%) supplied the requested information and/or unpublished data. Altogether, we gained an additional 40 effect sizes for the overall analysis and 10 effect sizes for the prime strength analysis through data and study information requests.

**Coding and Reliability**

The first author trained four coauthors to code the moderator variables using a scoring manual (see Appendix), and studies were randomly assigned to coders. The first author also double-coded 45% of the effect sizes (56) and studies (32). Final values were determined by resolving inconsistencies between the first author and coders through discussion. Studies that involved prime strength manipulations were coded twice: once collapsing over prime strength for the primary analysis and once coding weak and strong prime conditions separately for the prime strength analysis. Reliability was assessed for continuous variables with intraclass correlation coefficients (ICCs) and for categorical variables with kappa coefficients. Across all moderator variables, the reliability between coders for both continuous (ICC = .97) and categorical (κ = .93) variables was highly reliable. Reliability for continuous variables ranged from .92 (blocks) to .99 (SOA), whereas reliability for categorical variables ranged from .71 (prime part of speech) to 1.0 (published, population, prime type, and target type).

**Meta-Analytic Procedures**

SPSS macros were used to compute the weighted average effect size, homogeneity test, analog to analysis of variance (ANOVA), and meta-regression analyses described below (Lipsey & Wilson, 2001; see also http://mason.gmu.edu/~dwilsonb/). Before conducting the meta-analyses, we carried out procedures to maintain statistical independence, and effect size was calculated correcting for repeated measures designs in which the correlation coefficient is unknown (see Lipsey & Wilson, 2001, for review).

**Maintaining statistical independence.** An important aspect of meta-analysis is that the effect sizes are statistically independent. This is problematic when variables are manipulated within subjects because these effect sizes are not independent. Three variables were commonly examined using within-subject designs and thus required special treatment: SOA, judgment task, and prime strength (see below for details on the prime strength analysis). For SOA, there were 10 studies that manipulated SOA within subjects (e.g., Fockenberg et al., 2006; Hermans, De Houwer, & Eelen, 1994; Klauer et al., 2009). One approach for maintaining statistical independence is to collapse over all levels by averaging. However, this approach can mask the true effect of a variable; for example, if strong evaluative priming occurs at 300 ms and becomes nonsignificant by 400 ms, averaging a 300-ms condition with a 800-ms condition may suggest that weak priming occurs at 550 ms. Because the majority (76%) of studies used SOAs of 250 to 300 ms, many effect sizes with these SOAs, but fewer that used longer SOAs, were included. Thus, when possible (seven of the 10 studies that manipulated SOA within subjects), we selected the longest SOA value from the study to introduce variability into the SOA variable. For studies in which the effect size of the longest SOA was not reported in the article (three out of 10 studies), we used the suboptimal strategy of averaging.

For judgment task, there was one study that manipulated task within subjects (Storbeck & Clore, 2008). This study used three task manipulations: the EDT, semantic categorization task, and the LDT. Because there were numerous EDT effect sizes in the meta-analysis, we elected not to use the EDT effect size and randomly selected the semantic categorization task effect size for inclusion.

**Effect size computation.** $F$ and $t$ statistics were used to compute Cohen’s $d$. These statistics were extracted from experiments for evaluative congruity main effects (congruent vs. incongruent) or prime (positive vs. negative) by target (positive vs. negative) interactions, both of which produce identical statistical values of the evaluative congruity effect. That is, the $F$ statistic will be the same regardless if all four prime–target combinations (positive–positive, positive–negative, negative–negative, negative–positive) are examined via an interaction or are averaged into two conditions (evaluatively congruent = positive–positive + negative–negative/2; evaluative incongruent = positive–negative + negative–positive/2) and analyzed as a main effect. The evaluative congruity effect is reflected in $d$, such that positive values indicate a larger evaluative congruity effect—longer reaction times for targets evaluatively incongruent with primes (saint–repulsive) compared to targets evaluatively congruent with primes (saint–delightful). When there was a reversed evaluative priming effect (i.e., quicker response to evaluatively incongruent target), a negative value was attached to $d$. Though it is better to use the correlation between conditions when computing $d$ for repeated measures designs (Dunlap, Cortina, Vaslow, & Burke, 1996), the correlation between congruent and incongruent conditions is rarely reported in the evaluative priming literature. Thus, the following formulae were used to compute $d$ from repeated measures designs (see Morris & DeShon, 2002; Rosenthal, 1991) when the correlation between evaluatively congruent and incongruent conditions was unknown:

$$d = \frac{t}{\sqrt{n}} \text{ or } \frac{\sqrt{F}}{\sqrt{n}},$$

where $F$ and $t$ are the respective $F$ and $t$ statistic values and $n$ is the total number of participants in the experiment. The above formulae were used to compute the effect sizes when the exact values for $F$ or $t$ were known ($k = 120$).

When the exact $F$ or $t$ values were not reported and the investigator did not respond to data requests ($k = 5$), alternative procedures were used. When there was sufficient information in an article to compute conservative effect sizes (e.g., article stated “$F < 1$”), two procedures were carried out. First, effect sizes were entered as 0 when investigators reported $F$ values less than one ($k = 2$) or nonsignificant ($k = 2$). Second, an effect size was imputed as the critical value from the $F$ distribution when an investigator mentioned the evaluative congruity effect was statis-

---

5 The initial coding included 16 (of 18) moderator variables as done by Michelle Hinojos, Gabriela Terrazas, and Stephanie M. Reyes. Two additional variables (prime and target part of speech) were subsequently coded by David R. Herring and Stephen L. Crites.
tically significant without reporting the exact $F$ statistic (see Cohn, 1991, for similar procedures).

Variances of $d$s were then computed. As with $d$, the variance of $d$ requires the correlation between conditions with a repeated measure design (Becker, 1988). When the correlation between conditions is unknown, the following formula can be applied (Morris & DeShon, 2002):

$$ SE^2 = \left( \frac{1}{n} \right) \left( \frac{n-1}{n-3} \right) \left( 1 + \frac{n d^2}{c(df)} \right) - \frac{d^2}{c(df)^2}, $$

where $n$ is equal to the total number of participants and $c(df)$ is the bias function:

$$ c(df) = 1 - \frac{3}{4df - 1}. $$

The above formulae were used to compute the variance for each $d$.

**Weighted average effect size, homogeneity, and model assumptions.** Variance weights were created by computing the inverse of the variance before computing the weighted average effect size ($\overline{ES}$). This procedure was carried out to make effect sizes proportional. For instance, sample effect sizes derived from large samples have smaller variances more accurately estimating the population effect size, whereas small samples have larger variances that less accurately estimate the population effect size. Using the inverse of the variance ensures that effect sizes with less sampling variance carry greater weight in computing the $\overline{ES}$.

After computing the effect sizes, the homogeneity statistic Cochrane’s $Q$ was computed to determine whether the $\overline{ES}$ estimated a common population mean, and the descriptive statistic $\overline{F}$ was estimated to provide a proportion of inconsistency among effect sizes (Higgins, Thompson, Deeks, & Altman, 2003). Because the homogeneity test was violated (see below), indicating that the effect sizes did not estimate a common population mean and dispersion of the variance was not due to sampling error alone (see Hedges & Olkin, 1985; Shadish & Haddock, 2009), the fixed-effects model was untenable.

When the homogeneity test is violated as in the present review (see below), one decides whether to use a simple random-effects model or a mixed-effects model. In a simple random-effects model, it is assumed that each effect size is part of a random sample of possible effect sizes (see Borenstein, Hedges, Higgins, & Rothstein, 2010; Hedges & Vevea, 1998; Raudenbush, 2009). This conceptualization is in contrast with the fixed-effects model, which presumes a fixed population effect size. In addition to subject-level sampling error ($\nu$), the random-effects model has random or between-study variance ($\nu$), reflecting uncertainty about the process generating the effect size. Frequently, the random-effects model is too simple because the investigator has a priori moderators in mind. Thus, we used a mixed-effects model because it integrates both the fixed- and random-effects models. That is, study-level characteristics are fixed, but it is acknowledged that there is uncertainty about the effect being generated (i.e., random variance between studies). The mixed effect is linearly modeled as (Raudenbush, 2009)

$$ d_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + K + \beta_p X_{ip} + \mu_i + \epsilon_i, $$

where $\beta_0$ is the model intercept, $X_{i1}, \ldots, X_{ip}$ are coded study characteristics, $\beta_1, \ldots, \beta_p$ reflect associations between study characteristics and the effect size, $\mu_i$ is the random effect of study $i$, and $\epsilon_i$ are errors of estimation. In sum, mixed effects are used to model phenomena that can be explained in part by certain study coded characteristics (e.g., SOA, judgment task), but also reflect a large degree of uncertainty about the process generating the effect (e.g., underlying priming mechanisms). For the present mixed model, the random variance component was estimated with the noniterative method of moments estimation due to ease of implementation and superior reliability (Raudenbush, 2009).

**Moderator analyses.** Because the evaluative priming paradigm involves a number of manipulated variables, we coded a large number of moderator variables. To help reduce the number of overall analyses and focus on theoretically relevant variables, we separated the moderator variables into primary (theoretical) and secondary (exploratory) variables. The primary variables include variables reviewed earlier: (a) the type of task carried out on the target stimulus (judgment task), (b) SOA, (c) the format of prime stimuli (prime type), (d) the format of target stimuli (target type), (e) whether the prime was evaluatively conditioned (prime conditioned), (f) the total number of trials, (g) the total number of blocks, (h) the total number of prime repetitions, and (i) the total number of target repetitions. The secondary variables include (a) prime instructions, (b) the type of target response instructions (response instructions), (c) how response times were analyzed (response time analysis), (d) whether facilitation scores were analyzed (facilitation scores), (e) prime part of speech, (f) target part of speech, (g) the type of participants (population), (h) whether the effect size was from a published article (published), and (i) whether stimuli were intermixed or separated as primes and targets (stimulus presentation).

Hedges and colleagues (Hedges, 1982; Hedges & Olkin, 1985; see Konstantopoulos & Hedges, 2009) established for meta-analysis an analog to ANOVA, which is used for categorical variables, and an analog to multiple regression, which is used for continuous variables. In the analog to ANOVA, the $Q$ statistic is partitioned into between-category heterogeneity ($Q_b$) and within-category heterogeneity ($Q_c$). $Q_b$ reflects the weighted sum of squares of category means about the grand mean, and is equivalent to testing an omnibus $F$ test of group mean differences in a one-way ANOVA. A significant $Q_b$ indicates significant category effect size differences by more than sampling error (see Tables 2 and 3 for $Q_b$ values for each categorical moderator). When categorical variables with more than two categories were significant, simple contrast procedures were used with the chi-square distributed $\chi^2$ (Hedges & Becker, 1986; Rosenthal & Rubin, 1982).

The analog to multiple regression involves weighted least squares regression algorithms to control heteroscedasticity among effect sizes and correction of standard errors inappropriate for meta-analysis (Hedges & Olkin, 1985). This modified regression procedure for meta-analysis is commonly referred to as meta-regression. As with the analog to ANOVA, $Q$ is partitioned into metaregression into two components: heterogeneity due to the regression model ($Q_h$) and heterogeneity unaccounted by the model ($Q_u$). $Q_h$ is comparable to the $F$ test in a regression model and, if significant, explains heterogeneity among effect sizes (see Tables 2 and 3 for $Q_h$ values for each continuous moderator).

Because several moderator variables were examined in the current analysis, we examined the zero-order correlation coefficients
Table 1
Correlations Among Moderator Variables

<table>
<thead>
<tr>
<th>Moderator</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Blocks</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Facilitation scores</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3. Judgment task</td>
<td>.19</td>
<td>.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Population</td>
<td>—</td>
<td>—</td>
<td>.09</td>
<td>.09</td>
<td>.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Prime conditioned</td>
<td>—</td>
<td>—</td>
<td>.13</td>
<td>.08</td>
<td>.25</td>
<td>.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Prime instructions</td>
<td>.41</td>
<td>.62</td>
<td>.19</td>
<td>.12</td>
<td>.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Prime part of speech</td>
<td>.28</td>
<td>.23</td>
<td>.21</td>
<td>.13</td>
<td>.20</td>
<td>.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Prime repetitions</td>
<td>.23</td>
<td>.32</td>
<td>.19</td>
<td>.07</td>
<td>.50</td>
<td>.32</td>
<td>.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Prime type</td>
<td>.07</td>
<td>.25</td>
<td>.21</td>
<td>.10</td>
<td>.30</td>
<td>.22</td>
<td>.27</td>
<td>.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Response instructions</td>
<td>.09</td>
<td>.09</td>
<td>.09</td>
<td>.21</td>
<td>.09</td>
<td>.08</td>
<td>.27</td>
<td>.06</td>
<td>.09</td>
<td>.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Reaction time analysis</td>
<td>.02</td>
<td>.22</td>
<td>.27</td>
<td>.07</td>
<td>.10</td>
<td>.20</td>
<td>.28</td>
<td>.09</td>
<td>.11</td>
<td>.33</td>
<td>.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. SOA</td>
<td>—</td>
<td>.06</td>
<td>.16</td>
<td>.07</td>
<td>.07</td>
<td>.17</td>
<td>.32</td>
<td>.04</td>
<td>.38</td>
<td>.06</td>
<td>.17</td>
<td>.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Stimulus presentation</td>
<td>.13</td>
<td>.09</td>
<td>.27</td>
<td>.01</td>
<td>.19</td>
<td>.22</td>
<td>.52</td>
<td>.16</td>
<td>.10</td>
<td>.11</td>
<td>.12</td>
<td>.15</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Target part of speech</td>
<td>.41</td>
<td>.28</td>
<td>.23</td>
<td>.16</td>
<td>.10</td>
<td>.29</td>
<td>.64</td>
<td>.15</td>
<td>.22</td>
<td>.21</td>
<td>.26</td>
<td>.39</td>
<td>.19</td>
<td>.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Target repetitions</td>
<td>.23</td>
<td>.02</td>
<td>.12</td>
<td>-.10</td>
<td>.21</td>
<td>.17</td>
<td>.33</td>
<td>.56</td>
<td>.13</td>
<td>-.29</td>
<td>.10</td>
<td>.07</td>
<td>-.02</td>
<td>.17</td>
<td>.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Target type</td>
<td>-.14</td>
<td>.12</td>
<td>.08</td>
<td>-.01</td>
<td>-.00</td>
<td>.18</td>
<td>.08</td>
<td>.02</td>
<td>.58</td>
<td>-.18</td>
<td>.12</td>
<td>.12</td>
<td>.22</td>
<td>-.21</td>
<td>.35</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Relations between continuous variables and between continuous variables and dichotomous variables were assessed with Pearson’s r, between continuous variables and polychotomous variables with multiple R, and between dichotomous variables and and dichotomous variables and polychotomous variables with Cramér’s V. Italicized moderators are variables that account for a significant or marginal amount of the evaluative priming effect (d) variance with single moderator analyses. Of these moderators, five strong relations were found (in bold) and considered when running controlled (partial/semipartial) regression analyses. SOA = stimulus-onset asynchrony.

*p < .05. **p < .01. ***p < .001.

among the moderators (see Table 1). When a moderator at least marginally predicted evaluative priming (see Tables 2 and 3) and correlated significantly (p < .001) with another moderator that at least marginally predicted evaluative priming, we conducted follow-up analyses using partial and semipartial correlations to ascertain whether the effect of one moderator remained significant when controlling for the other. For example, blocks predicted the amount of evaluative priming and correlated highly with facilitation scores, prime instructions, and target part of speech. Because only target part of speech also predicted the magnitude of evaluative priming (see Table 3), we examined the effect of blocks controlling target part of speech but not facilitation scores or prime instructions. These analyses and findings are discussed below where appropriate.

Publication bias. A major threat to the validity for any systematic review is that it is not representative of the literature, an issue commonly referred to as publication bias (see Sutton, 2009, for review). Conducting a systematic review that is representative of the literature is challenging because unpublished articles are often difficult to locate, and likely unpublished in the first place because of much smaller effects than published articles (Lipsey & Wilson, 1993). As is recommended for meta-analyses and discussed above, we attempted to identify unpublished work by writing to authors (n = 49) publishing in this area and asking them to send us unpublished work. We also systematically examined the studies in this meta-analyses using three methods to help identify publication bias and, if present, the extent of the bias.

There are three commonly used methods for identifying publication bias. One method is simply to code whether studies were published and include this as a moderator variable, but this method requires a large set of unpublished studies (Lipsey & Wilson, 2001). A second approach for identifying publication bias is with a funnel plot (see Light & Pillemer, 1984; Sterne, Becker, & Egger, 2005). A funnel plot is a scatterplot that is typically devised by having the effect size plotted along the x-axis and some measure of precision (e.g., sample size, variance, standard error) plotted against the y-axis. The logic of this approach is that large samples with stronger precision will cluster near the base of the funnel about the true effect size, whereas smaller samples with less precision will scatter (via sampling error) from the base of the funnel (see Figure 8). A systematic review free of publication bias should have effect sizes dispersed symmetrically throughout the funnel, whereas a review with publication bias may have asymmetric dispersion of effect sizes throughout the funnel. Finally, a third approach for identifying publication bias is the Egger regression test (Egger, Smith, Schneider, & Minder, 1997; Sterne & Egger, 2005). Funnel plot asymmetry is tested by regressing the standard normal deviate (effect size in this case, d/SE) on precision (1/SE). When providing a regression line through the funnel plot, a symmetric funnel plot will have a line that passes through the origin of the standard normal deviate zero, whereas an asymmetric funnel plot will not pass through the origin. Thus, the degree of asymmetry can be ascertained by testing the intercept of the regression model, where a significant intercept indicates a funnel plot asymmetry. In the present meta-analysis, we used all three approaches for assessing publication bias.

When publication bias is present, one of the most common procedures to examine the extent of the bias is the trim and fill method (see Duval, 2005). The trim and fill method is applied when the effect sizes are asymmetrically displaced to either of the lower or upper portions of the funnel plot. This iterative

6 We thank Karl Klauer for recommending this analysis.
procedure estimates the number of effect sizes that would fill the funnel plot to make it symmetric and provides an updated $ES$. The present review used the trim and fill method to reestimate the $ES$ by filling the lower left portion of the funnel plot using $L_0$ estimation.

**Results**

**Analytic Approach**

The results are divided into sections: (a) an overall analysis of evaluative priming and the homogeneity test, (b) an analysis of primary moderators, (c) an analysis of secondary moderators, (d) a prime strength analysis, and (e) a publication bias analysis.

**Overall Analysis**

Prior to running the overall analysis, one outlier below 3 standard deviations from the $ES$ was replaced with the 3 standard deviation value below the $ES$ (Lipsey & Wilson, 2001). The $ES$ was computed with mixed-effects modeling on 125 effect sizes ($N = 5,367$) ranging from 0.71 to 1.36 (see Figure 3). The mixed-effects model yielded a $ES = 0.37$ ($SE = 0.03$, 95% CI [0.31, 0.43], $z = 12.55$, $p < .001$). Thus, significant evaluative priming was found with mixed-effects modeling. Heterogeneity among effect sizes was confirmed with both the $Q$ and $I^2$ statistics, $Q(124) = 458.97$, $p < .001$, $I^2 = 73\%$ (95% CI [68%, 77%]), indicating substantial heterogeneity among effect sizes. Therefore, we carried out moderator analyses to account for this marked heterogeneity with variables selected in advance for theoretical purposes (primary moderators) and then for variables selected for exploratory purposes (secondary moderators).

**Primary Analyses**

**Descriptive analysis.** Experiments used mostly EDTs (61%), averaged an SOA of 321 ms, used predominately verbal primes (56%) and targets (79%), rarely used evaluatively conditioned primes (15%), and averaged 143 trials, three blocks, eight prime repetitions, and four target repetitions (see Appendix).

**Moderator analysis.** A series of ANOVA analogs and meta-regressions were run on the nine primary moderator variables individually (as in Van den Bussche et al., 2009). Because total trials was conceptually similar to blocks and there was no effect of the total trials variable, we do not discuss total trials further. The statistics for the remaining eight primary moderator variables are presented below (see Table 2). Certain categorical moderators (stimulus presentation, prime part of speech, target type, and prime instructions) contained a level with less than three effect sizes, and accordingly these levels were removed (as in Hofmann et al., 2010), making these variables dichotomous.

**Blocks.** Effect sizes were regressed on blocks. There was a significant relation between blocks and evaluative priming such that increases in blocks were associated with greater evaluative priming ($\beta = .27$, $k = 125$). Because target part of speech was significantly correlated with blocks and also predicted effect size (see Tables 1 and 3), we conducted follow-up analyses that controlled for this high correlation. The effect of blocks remained significant in this analysis ($\beta = .38$, $z = 3.89$, $p < .001$, $k = 89$). Blocks accounted for 7.2% of the variance.

---

**Figure 3.** Forest plot of the 125 effect sizes (black squares) with 95% confidence intervals (horizontal lines). The $ES$ (0.37) is represented by the bottommost square.
Judgment task. There was a significant effect of judgment task, accounting for 9.6% of the variance. The effect size of the lexical decision task; SOA stimulus-onset asynchrony.

ON THE AUTOMATIC ACTIVATION OF ATTITUDES

remained significant when controlling SOA (was significantly different from zero. The effect of prime type significantly from verbal and nonverbal prime conditions but not in semantic categorization tasks or LDTs. The effect for primes that contained both nonverbal and verbal stimulations significantly differed from zero. Note that there was no condition involving a combination of both verbal and nonverbal stimulli for target type.


declinations of judgment task, two-way interactions between judgment task and all other primary variables were examined first, revealing no significant interactions (all zs < 1.88). Although these overall analyses revealed no significant interactions, three interactions involving judgment task were examined with follow-up analyses based on theory.

The first theoretically relevant follow-up analysis was between prime types for each judgment task (see Figure 4). Although the overall interaction between prime type and judgment task was not significant, $Q_{\text{change}}(3) = 3.85, p = .28$. We conducted simple comparisons between prime types for each judgment task to examine theory described earlier regarding nonverbal stimuli and priming (Spruyt et al., 2002; Zhang et al., 2006). For the pronunciation task, there was a strong trend indicating evaluative priming was larger with nonverbal primes ($M = 0.46$) than verbal primes.

SOA. As expected, decreases in SOA were associated with increases in evaluative priming ($\beta = -.18, k = 125$). The effect of SOA remained significant when controlling prime type ($\beta = -.25, z = 2.90, p = .003$), which was correlated with SOA (see Table 1). SOA accounted for 3.3% of the variance. Because the majority of effect sizes (76%) involved either 250- or 300-ms SOAs, we ran a supplemental analysis comparing these SOAs. The contrast between 250- and 300-ms SOAs was not significant ($Q = 0.72$).

Two-way interactions. Given the strong theoretical implications of judgment task, two-way interactions between judgment task and all other primary variables were examined first, revealing no significant interactions (all $zs < 1.88$). Although these overall analyses revealed no significant interactions, three interactions involving judgment task were examined with follow-up analyses based on theory.

The first theoretically relevant follow-up analysis was between prime types for each judgment task (see Figure 4). Although the overall interaction between prime type and judgment task was not significant, $Q_{\text{change}}(3) = 3.85, p = .28$. We conducted simple comparisons between prime types for each judgment task to examine theory described earlier regarding nonverbal stimuli and priming (Spruyt et al., 2002; Zhang et al., 2006). For the pronunciation task, there was a strong trend indicating evaluative priming was larger with nonverbal primes ($M = 0.46$) than verbal primes.

7 Up to this point, standard meta-regressions have been used. This approach, however, is not ideal for testing the overall interaction between prime type and judgment task because it is theoretically unimportant to compare the pronunciation task to another task (e.g., EDT) testing simple slopes because our prediction is specific to differences between nonverbal and verbal stimuli for the pronunciation task. We used the statistically equivalent hierarchical meta-regression with the $Q_{\text{change}}$ statistic (Konstantopoulos & Hedges, 2009). Judgment task was coded with effects coding, which yields comparable Type III sums of squares as an ANOVA model (Cohen, Cohen, West, & Aiken, 2003).

<table>
<thead>
<tr>
<th>Blocks</th>
<th>$d\beta$</th>
<th>95% CI</th>
<th>SE</th>
<th>C</th>
<th>$Q_{d/Q_{k}}$ (df)</th>
<th>$R^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judgment task</td>
<td>.27</td>
<td>[.14, .40]</td>
<td>.07</td>
<td>125</td>
<td>10.75 (1)</td>
<td>.07</td>
<td>.001</td>
</tr>
<tr>
<td>EDT</td>
<td>.45</td>
<td>[.37, .52]</td>
<td>.04</td>
<td>a</td>
<td>76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic</td>
<td>.20</td>
<td>[-.05, .44]</td>
<td>.12</td>
<td>b</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pronunciation</td>
<td>.29</td>
<td>[-.19, .39]</td>
<td>.05</td>
<td>b</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDT</td>
<td>.06</td>
<td>[-.19, .31]</td>
<td>.13</td>
<td>b</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prime conditioned</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>.41</td>
<td>[.27, .55]</td>
<td>.07</td>
<td>a</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>.36</td>
<td>[.29, .42]</td>
<td>.03</td>
<td>a</td>
<td>106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prime repetitions</td>
<td>.02</td>
<td>[-.16, .19]</td>
<td>.09</td>
<td></td>
<td>125</td>
<td>0.07 (1)</td>
<td>.00</td>
</tr>
<tr>
<td>Prime type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonverbal</td>
<td>.48</td>
<td>[.34, .52]</td>
<td>.05</td>
<td>a</td>
<td>49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal</td>
<td>.31</td>
<td>[.23, .38]</td>
<td>.04</td>
<td>b</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both</td>
<td>.53</td>
<td>[.29, .77]</td>
<td>.12</td>
<td>a, b</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOA</td>
<td>-.18</td>
<td>[-.40, -.04]</td>
<td>.11</td>
<td></td>
<td>125</td>
<td>4.69 (1)</td>
<td>.03</td>
</tr>
<tr>
<td>Target type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonverbal</td>
<td>.36</td>
<td>[.23, .49]</td>
<td>.07</td>
<td>a</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal</td>
<td>.37</td>
<td>[.30, .45]</td>
<td>.03</td>
<td>a</td>
<td>97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target repetitions</td>
<td>.12</td>
<td>[-.04, .27]</td>
<td>.08</td>
<td>a</td>
<td>123</td>
<td>1.87 (1)</td>
<td>.01</td>
</tr>
</tbody>
</table>

Note. Different subscripts indicate $p < .05$. $d\beta$ = Cohen’s $d$ beta coefficient; CI = confidence interval; C = contrast index; EDT = evaluative decision task; LDT = lexical decision task; SOA = stimulus-onset asynchrony.

Table 2
Primary Moderator Variable Statistics
(M = 0.23), Q(1) = 3.09, p = .07. However, for the other three judgment tasks there was no difference in prime types (Qs < 2.33).

The second theoretically relevant follow-up analysis involving judgment task was the interaction with SOA coded categorically (250 ms vs. 300 ms; see above), as opposed to the primary SOA analysis in which SOA was coded continuously. The interaction between categorical SOA and judgment task was significant (β = .28, z = 2.00, p = .045; see Figure 5). For the EDT, simple slopes analysis revealed that the 300-ms SOA (relative to 250 ms) was significantly associated with decreases in evaluative priming (β = −.37, z = −2.71, p = .007). This pattern was not significant for the pronunciation task (β = .04, z = 0.36, p = .714). Thus, increasing SOA significantly reduces the evaluative priming effect for the EDT, but evaluative priming is stable across the pronunciation task.

The third theoretically relevant follow-up was between judgment task and stimulus presentation, which we categorized as a secondary variable (see Klauer, Eder, Greenwald, & Abrams, 2007, for review). Stimulus presentation refers to whether primes and targets were selected from distinct stimulus pools (i.e., primes never seen as targets and vice versa) versus from the same pool (i.e., same stimulus could serve as prime on some trials and target on others). Because there were less than three effect sizes for one of the levels, the three nonevaluative judgment tasks (pronunciation, lexical decision, semantic categorization) were combined and contrasted with the EDT. As would be predicted (D. Wentura, personal communication, August 9, 2012), the interaction between stimulus presentation and judgment task was significant (β = −.69, z = −3.04, p = .002). For nonevaluative judgment tasks, separated stimulus sets were significantly associated with larger evaluative priming relative to intermixed sets (β = .48, z = 3.68, p < .001). The simple slope for the EDT was nonsignificant (β = −.01, z = −0.09, p = .46). The finding for nonevaluative tasks may be due to “response competition” that occurs when a stimulus first serves as a target and then later occurs as a prime in nonevaluative tasks. For example, after saying the word happy when it is a target in a pronunciation task, seeing it as a prime may partially activate the process of naming it, which conflicts with the required response to the target.

The two-way interactions not involving judgment task were then examined. All significant interactions involving continuous variables were examined plotting 1 standard deviation above and below the mean (see Appendix for means and standard deviations of moderators; Aiken & West, 1991).

Two significant interactions involving blocks were found. There was a significant interaction between blocks (centered) and prime type (dummy coded; β = −.28, z = −2.67, p = .008; see Figure 6A). For verbal primes, simple slopes analysis revealed that high number of blocks (relative to low) were significantly associated with increases in evaluative priming (β = .47, z = 4.28, p < .001). This pattern was not significant for nonverbal stimuli (β = .05, z = 0.44, p = .663). The interaction between blocks and prime type, however, was nonsignificant (β = −.11, z = −1.15, p = .24, k = 87) when controlling target part of speech and SOA, which were correlated with these variables (see Table 1). There was also a significant interaction between blocks and target type (dummy coded; β = −.32, z = −3.48, p < .001; see Figure 6B). For verbal targets, simple slopes analysis revealed that high blocks (relative to low) were significantly associated with increases in evaluative priming (β = .41, z = 4.63, p < .001). For nonverbal stimuli, there was no significant relation (β = −.30, z = −1.65, p = .099). Thus, these two interactions suggest that increasing the number of blocks increases the evaluative priming effect when primes or targets are verbal stimuli, but has no effect for nonverbal primes or targets.

There were two significant interactions involving prime repetitions. First, there was a significant interaction between prime repetitions and blocks (β = −.35, z = −2.62, p = .009; see Figure 7). For low prime repetitions, follow-up analyses revealed larger priming associated with high blocks (β = .51, z = 4.39, p < .001) relative to low blocks. For high prime repetitions, simple slopes analysis was not significant (β = .09, z = 0.85, p = .394). Second, there was a significant interaction between prime repetitions and prime conditioned (β = −.59, z = −2.54, p = .011). For unconditioned primes, follow-up analyses revealed that high prime repetitions (vs. low prime repetitions) were associated with larger evaluative priming (β = .55, z = 2.29, p = .022). For conditioned primes, the simple slope was not significantly associated with

(M = 0.23), Q(1) = 3.09, p = .07. However, for the other three judgment tasks there was no difference in prime types (Qs < 2.33).

The second theoretically relevant follow-up analysis involving judgment task was the interaction with SOA coded categorically (250 ms vs. 300 ms; see above), as opposed to the primary SOA analysis in which SOA was coded continuously. The interaction between categorical SOA and judgment task was significant (β = .28, z = 2.00, p = .045; see Figure 5). For the EDT, simple slopes analysis revealed that the 300-ms SOA (relative to 250 ms) was significantly associated with decreases in evaluative priming (β = −.37, z = −2.71, p = .007). This pattern was not significant for the pronunciation task (β = .04, z = 0.36, p = .714). Thus, increasing SOA significantly reduces the evaluative priming effect for the EDT, but evaluative priming is stable across the pronunciation task.

The third theoretically relevant follow-up was between judgment task and stimulus presentation, which we categorized as a secondary variable (see Klauer, Eder, Greenwald, & Abrams, 2007, for review). Stimulus presentation refers to whether primes and targets were selected from distinct stimulus pools (i.e., primes never seen as targets and vice versa) versus from the same pool (i.e., same stimulus could serve as prime on some trials and target on others). Because there were less than three effect sizes for one of the levels, the three nonevaluative judgment tasks (pronunciation, lexical decision, semantic categorization) were combined and contrasted with the EDT. As would be predicted (D. Wentura, personal communication, August 9, 2012), the interaction between stimulus presentation and judgment task was significant (β = −.69, z = −3.04, p = .002). For nonevaluative judgment tasks, separated stimulus sets were significantly associated with larger evaluative priming relative to intermixed sets (β = .48, z = 3.68, p < .001). The simple slope for the EDT was nonsignificant (β = −.01, z = −0.09, p = .46). The finding for nonevaluative tasks may be due to “response competition” that occurs when a stimulus first serves as a target and then later occurs as a prime in nonevaluative tasks. For example, after saying the word happy when it is a target in a pronunciation task, seeing it as a prime may partially activate the process of naming it, which conflicts with the required response to the target.

The two-way interactions not involving judgment task were then examined. All significant interactions involving continuous variables were examined plotting 1 standard deviation above and below the mean (see Appendix for means and standard deviations of moderators; Aiken & West, 1991).

Two significant interactions involving blocks were found. There was a significant interaction between blocks (centered) and prime type (dummy coded; β = −.28, z = −2.67, p = .008; see Figure 6A). For verbal primes, simple slopes analysis revealed that high number of blocks (relative to low) were significantly associated with increases in evaluative priming (β = .47, z = 4.28, p < .001). This pattern was not significant for nonverbal stimuli (β = .05, z = 0.44, p = .663). The interaction between blocks and prime type, however, was nonsignificant (β = −.11, z = −1.15, p = .24, k = 87) when controlling target part of speech and SOA, which were correlated with these variables (see Table 1). There was also a significant interaction between blocks and target type (dummy coded; β = −.32, z = −3.48, p < .001; see Figure 6B). For verbal targets, simple slopes analysis revealed that high blocks (relative to low) were significantly associated with increases in evaluative priming (β = .41, z = 4.63, p < .001). For nonverbal stimuli, there was no significant relation (β = −.30, z = −1.65, p = .099). Thus, these two interactions suggest that increasing the number of blocks increases the evaluative priming effect when primes or targets are verbal stimuli, but has no effect for nonverbal primes or targets.

There were two significant interactions involving prime repetitions. First, there was a significant interaction between prime repetitions and blocks (β = −.35, z = −2.62, p = .009; see Figure 7). For low prime repetitions, follow-up analyses revealed larger priming associated with high blocks (β = .51, z = 4.39, p < .001) relative to low blocks. For high prime repetitions, simple slopes analysis was not significant (β = .09, z = 0.85, p = .394). Second, there was a significant interaction between prime repetitions and prime conditioned (β = −.59, z = −2.54, p = .011). For unconditioned primes, follow-up analyses revealed that high prime repetitions (vs. low prime repetitions) were associated with larger evaluative priming (β = .55, z = 2.29, p = .022). For conditioned primes, the simple slope was not significantly associated with
In sum, greater evaluative priming occurs for low prime repetitions with more blocks and unconditioned primes with more prime repetitions.

It was not possible to examine the interaction between prime type and target type because verbal primes never preceded nonverbal targets. Most studies had verbal primes that preceded verbal targets ($k = 68$). When nonverbal primes were used, there was about an equal number of effect sizes for nonverbal targets ($k = 25$) and verbal targets ($k = 24$). We compared these three possible conditions: consistent verbal, consistent nonverbal, and inconsistent nonverbal–verbal prime–target pairs. This analysis was significant, $Q(2) = 6.80$, $p = .033$, and follow-up contrasts revealed a significant difference between inconsistent nonverbal–verbal ($M = 0.51, k = 24$) and consistent verbal ($M = 0.31, k = 70, z^2 = 6.79, p = .009$) and a marginal difference between inconsistent nonverbal–verbal and consistent nonverbal ($M = 0.35, k = 25, z^2 = 2.81, p = .094$). No difference was found between consistent nonverbal and consistent verbal ($z^2 = .04, p = .851$). Thus, strongest evaluative priming occurs when a prime is nonverbal followed by a verbal target.

Additionally, two marginal interactions were uncovered involving blocks, one with prime conditioned ($\beta = -.17, z = -1.88, p = .06$) and one with SOA ($\beta = -.15, z = -1.87, p = .061$). The simple slope for unconditioned primes was significant ($\beta = .33, z = 3.87, p < .001$), such that high blocks (vs. low blocks) were associated with larger evaluative priming, whereas the simple slope for conditioned primes was not significantly associated with blocks ($\beta = -.11, z = -0.49, p = .622$). Also, the simple slope for low SOA was positively associated with blocks ($\beta = .45, z = 3.60, p < .001$), such that high blocks (vs. low blocks) were associated with larger evaluative priming. The simple slope for high SOA was not significantly associated with blocks ($\beta = .05, z = 0.36, p = .717$).
Secondary Analyses

Descriptive analysis. In the majority of studies, participants were instructed either to ignore or to attend to primes (73%) and were not given additional response instructions (82%). Primarily, raw response times (72%) were analyzed, and rarely were facilitation scores computed (10%). Most verbal primes were nouns (67%), and verbal targets almost always were either nouns or adjectives (89%). Most studies used an academic population (94%), and most effect sizes were from published articles (86%). Last, most stimuli were separated into distinct prime and target subgroups (83%).

Moderator analysis. A series of ANOVA analogs and meta-regressions were run on the nine secondary moderator variables individually, as above. The statistics for the secondary moderator variables are presented below (see Table 3). Only three moderator variables approached significance: response instructions, stimulus presentation, and target part of speech. There was a trend for target response instructions on evaluative priming such that no additional instruction \((M = 0.40, k = 103)\) was larger than feedback \((M = 0.24, k = 19)\), and neither differed from paid instructions \((M = 0.22, k = 3)\). Also, there was a strong trend for stimulus presentation such that there were larger effect sizes for stimuli separated \((M = 0.39, k = 102)\) than stimuli intermixed \((M = 0.24, k = 21)\). However, the effect of stimulus presentation on the evaluative priming effect was nonsignificant when either prime part of speech \((\beta = .05, z = 0.36, p = .71)\) or target part of speech \((\beta = .05, z = 0.46, p = .64)\) was controlled. There was a strong trend toward significance for target part of speech such that noun targets \((M = 0.42, k = 40)\) were larger than noun–adjective combinations \((M = 0.05, k = 4)\) as well as combinations involving verbs \((M = 0.07, k = 5)\), which did not differ from one another. There was no difference between noun and adjective targets \((M = 0.36, k = 40)\). When controlling blocks, the effect of target part of speech on the evaluative priming effect remained significant for the noun–verb comparison \((\beta = -.25, z = -2.26, p = .02)\), the noun–noun and adjective combination comparison was no longer significant \((\beta = \ldots)\).
Note. Different subscripts indicate $p < .05$. $d\beta$ = Cohen’s $d$beta coefficient; CI = confidence interval; $C$ = contrast index; log = logarithmic transformation.

$-1.49, z = -1.49, p = .14$, and the noun–adjective comparison became significant ($\beta = -0.20, z = -2.01, p = .04$). None of the target part of speech comparisons remained significant when either prime part of speech ($Bs < .46, zs < 1.56, ps > .11$) or stimulus presentation ($Bs < .43, zs < 1.71, ps > .08$) was controlled. Last, two-way interactions among the secondary variables were not explored because these variables were atheoretical and a number of two-way interactions were already examined.

### Prime Strength Analysis

As discussed, prime strength is typically manipulated within subjects and thus problematic for meta-analysis because each participant will have effect sizes for both weak and strong stimuli for an experiment. Unlike SOA and judgment task in which procedures were devised to avoid using multiple effect sizes from the same experiment (see above), the very nature of the prime strength analysis involves comparing two statistically dependent effect sizes (i.e., strong vs. weak).

To correct the statistical dependency of prime strength, we ran meta-regressions with an SPSS macro (http://www.ipr.northwestern.edu/qcenter) using robust variance estimation for dependent effect sizes (Hedges, Tipton, & Johnson, 2010). Ideally, one would have knowledge of the covariance structure involving the dependent effect sizes; however, this is rarely attainable. Instead, Hedges et al.’s (2010) robust variance estimation theorem uses the residual cross products of the dependent effect sizes as a crude estimate of the covariance structure. Note two additional facets about this theorem: (a) an approximation of the between-effect/within-study correlation still must be provided ($\rho = 1$ was used to be conservative in the present analysis), and (b) the meta-regressions’ coefficients are $t$ distributed with $m$ (study clusters) minus $p$ (predictors) degrees of freedom.

The prime strength analysis involved eight studies yielding 28 effect sizes (14 strong, 14 weak). There was no effect of prime strength, $t(6) = 0.77, p = .465$, nor did any of the primary variables interact with prime strength, all $ts < 1.56$ or $> -0.35$. We also tested the interaction between prime instructions and prime strength because the prime instructions variable has been implicated (Simmons & Prentice, 2006), but there was no interaction effect, $ts(4) < 1.20$.

---

9 An additional six studies could not be used because the corresponding authors did not respond to our data requests.
Publication Bias Analysis

Three techniques were used to assess potential publication bias. First, a funnel plot was constructed by plotting standard errors along the vertical axis and effect sizes along the horizontal axis (Light & Pillemer, 1984; Sterne et al., 2005). The funnel plot (see Figure 8) revealed asymmetry with displacement to the lower right of $ES$ (0.37), leaving a gap in the lower left portion where negative (i.e., reversed) priming study effects would be expected. Second, the asymmetry in the funnel plot was confirmed with the Egger test (Egger et al., 1997; Sterne & Egger, 2005; $a = 2.43$, 95% CI [1.44, 3.42], $t = 4.86$, $p < .001$). Because the judgment task variable has important theoretical implications, we ran separate Egger tests for the evaluative decision and pronunciation tasks (Sutton, 2009). Funnel plot asymmetry remained evident for both the evaluative decision ($a = 2.31$, 95% CI [1.02, 3.60], $t = 3.47$, $p = .001$) and pronunciation tasks ($a = 2.11$, 95% CI [0.51, 3.71], $t = 2.56$, $p = .02$). Third, whether the experiment was published or unpublished was examined as a moderator of effect size, and this analysis revealed no significant effect (see Table 3). This finding is somewhat surprising considering publication bias is indicated by the other two methods; however, the smaller number of unpublished studies ($k = 107$) compared to published ones ($k = 125$) in the present review may be partly responsible for this nonsignificant effect (Lipsey & Wilson, 2001).

The trim and fill method (Duval, 2005) was used to identify the number of missing studies that would fill the distribution of effect sizes to remove asymmetry and thus remove potential publication bias. After several iterations with $L_0$ estimation, 24 missing studies’ ($k_0$) effect sizes were estimated and the $ES$ was recalculated with mixed-effects modeling, $ES = 0.24$ ($SE = 0.03$, 95% CI [0.17, 0.30], $z = 7.13$, $p < .001$). We also ran separate trim and fill analyses for the EDT and pronunciation task. For the EDT, 16 missing effect sizes ($k_0$) were estimated and the recalculated $ES = 0.29$ ($SE = 0.04$, 95% CI [0.20, 0.37], $z = 6.55$, $p < .001$). For the pronunciation task, seven missing effect sizes ($k_0$) were estimated and the recalculated $ES = 0.15$ ($SE = 0.06$, 95% CI [0.03, 0.27], $z = 2.47$, $p = .01$).

In sum, if publication bias drives distributional asymmetry, then it considerably diminishes the $ES$ and poses a moderate threat to validity, albeit the $ES$ remains statistically significant and contains even greater heterogeneity ($I^2 = 82\%$).

Discussion

The present meta-analysis covered a quarter century of evaluative priming studies ($k = 125$) and found strong support for Fazio et al.’s (1986) seminal finding that prime valence influences the speed with which people respond to a valenced target. The judgment made to targets, format of primes, and SOA predictably influenced evaluative priming: Evaluative priming was strongest when people evaluated targets as good or bad, when pictures and symbols were primes, and with shorter SOAs. Unexpectedly, evaluative priming was stronger as the number of discrete blocks increased and reliable with the pronunciation task but not the lexical decision or semantic categorization task. The major findings of the present meta-analysis are consistent with recent research and theory suggesting that evaluative priming can be due to
multiple processes that impact encoding and/or response mechanisms, depending upon variables in the priming procedure. Although the findings provide some support for the notion that evaluative priming can occur automatically (Bargh, 1994), this support is limited and suggests more nuanced theoretical perspectives are needed. A few such perspectives have been proposed recently, and the findings of this meta-analysis provide some support for these nonexclusive perspectives. Specifically, some findings are consistent with the attentional sensitization/feature-specific attention allocation perspective that provides insight into how attention impacts memory encoding and priming (Spruyt, De Houwer, Everaert, & Hermans, 2012; Spruyt, De Houwer, Hermans, & Eelen, 2007; Spruyt, De Houwer, & Hermans, 2009). Similarly, some findings are consistent with the evaluation window perspective that details how judgment processes can lead to normal and even reversed priming.

A primary conclusion is that neither classic perspective of evaluative priming can completely account for the results from this meta-analysis. The encoding perspective holds that primes activate associations in memory that make it easier to evaluatively encode targets with similar valence. Thus, evaluative priming should occur regardless of the response requirements to targets because assessing the meaning of targets occurs before response processes are initiated. The results of this meta-analysis, however, failed to support an exclusive version of the encoding perspective because evaluative priming did not occur in the semantic categorization task or the LDT. Thus, encountering a valenced prime does not unconditionally facilitate the evaluative encoding of a similarly valenced target. Facilitated encoding may help explain evaluative priming in the evaluative decision and pronunciation tasks, but theoretical formulations are needed that can explain why encoding occurs during some judgment tasks and not others.

As research on evaluative priming accumulated, a response perspective of evaluative priming was proposed because researchers realized that evaluative priming was more reliable and greater in the EDT relative to the other judgment tasks. This second classic perspective of evaluative priming holds that (a) both primes and targets activate response tendencies and (b) responses to targets are facilitated (faster) when the response tendency activated by the prime is congruent with that activated by the target and/or inhibited (slower) when incongruent (Klauer, 1998; Klauer & Musch, 2003). If evaluative priming is due solely to processes associated with selecting and executing responses, it should only occur during the EDT because this is the only task requiring an explicit evaluative response to targets. The present results demonstrate that response processes are not solely responsible for evaluative priming because significant priming was found in the pronunciation task. Response processes, however, may help explain why evaluative priming is significantly stronger in the EDT compared to the pronunciation task. In sum, the present results suggest the need for more nuanced theoretical perspectives that can incorporate both encoding and response processes and also integrate other constructs to explain when and how these (and perhaps other) processes are active.

One attempt to introduce new moderating constructs for evaluative priming comes from theorists recently focusing on the role of attention (Kiefer, Adams, & Zovko, 2012; Kiefer & Martens, 2010; Spruyt, De Houwer, et al., 2007; Spruyt et al., 2009). Spruyt and colleagues (Everaert, Spruyt, & De Houwer, 2011, 2012; Spruyt et al., 2012, 2009) proposed a feature-specific attention allocation model to help explain when priming occurs and demonstrated the important role attention plays in both nonevaluative semantic and evaluative priming. For instance, Spruyt et al. (2009) had participants perform a task in which 25% of the trials were a pronunciation task and 75% of trials were either an EDT or semantic categorization task (object vs. person). The decision participants made in the other task moderated priming in the pronunciation task. When people made evaluative decisions in 75% of the trials, evaluative (but not semantic) priming was found in the 25% of pronunciation trials. When people made semantic decisions in 75% of the trials, semantic (but not evaluative) priming was found in the 25% of pronunciation trials. Thus, evaluative and semantic priming is greater when the task (judgment people make, trial structure, type of stimuli) focuses attention on and sensitizes cognitive processes (“attentional sensitization”) that are specific to the priming being explored (see also Kiefer, 2007; Kiefer et al., 2012).

Theory discussing the importance of attentional sensitization of cognitive processes is new, so the research reviewed in this meta-analysis did not examine attention per se. Nevertheless, there are a number of findings in the meta-analysis that may involve attentional sensitization. A primary finding that might be explained by attention is significant differences across the judgments—significant evaluative priming in the evaluative decision and pronunciation tasks but nonsignificant priming in lexical decision and semantic categorization tasks. Because the EDT focuses attention on evaluation by requiring an evaluative judgment for every trial, an attentional sensitization perspective may help explain why relatively large priming effects occur in the EDT. Can attentional sensitization explain why evaluative priming occurs in the pronunciation task but not the other two tasks that involve nonevaluative responses to the target (semantic categorization and LDT)? Spruyt et al. (2011) have suggested that the pronunciation task produces a neutral processing mindset: Participants do not have to make a semantic judgment and can simply initiate a “relatively automatic” behavioral response. This suggests that evaluative priming occurs during the pronunciation task when participants are able to focus more on evaluation (e.g., evaluating 75% of the time, as described above). In contrast, the semantic categorization tasks and LDTs may engender attentional focus on nonevaluative semantic aspects of stimuli. This nonevaluative attentional focus may prevent evaluative activation sufficient for an evaluative priming effect in “standard” semantic categorization and LDTs (Spruyt et al., 2009). Hence, attentional sensitization may help explain priming differences across judgment tasks employed in evaluative priming. Finally, although significant evaluative priming was not found in the semantic categorization task in the present meta-analysis, no special procedures were used to make evaluation salient in the majority of semantic categorization tasks included in this meta-analysis. If aspects of the task are changed to make evaluation salient, evaluative priming may occur in semantic categorization tasks and LDTs (see, e.g., Spruyt, De Houwer, et al., 2007).

One way to increase sensitization to evaluation might be to reduce variability along other dimensions because this would allow more focused sensitization of the cognitive processes associated with evaluation and less sensitization of cognitive processes associated with other dimensions of meaning. The findings involving evaluatively conditioned versus unconditioned primes might
be interpreted in this way and thus be explained by an attentional sensitization account. That is, evaluatively conditioned primes had equivalent effect sizes regardless of whether they were presented few or many times; however, the effect size for primes with an a priori valence increased as the number of prime repetitions increased. One significant difference between these two prime types (other than when or how valence was acquired) is that conditioned primes had no other semantic meaning because semantically meaningless stimuli are typically selected for conditioning so that valence can be easily acquired. Because conditioned primes likely had no semantic associations other than evaluation, strong evaluative priming may have occurred with low numbers of repetitions because evaluation was the only dimension of meaning that would be activated when these stimuli were encountered. On the other hand, evaluative priming for primes with multiple dimensions of meaning may have increased as the number of repetitions increased because the context and structure of the task gradually focused attention on evaluation.

Two other findings might also be interpreted as due to increased attention as a result of reduced variability. First, priming was larger when the targets were all nouns or all adjectives compared to when targets were noun and adjective combinations or contained verbs. Increased variability in the meaning of words associated with different parts of speech may reduce evaluative priming because attention is taken away from evaluation. Second, there was also a trend for greater priming when primes and targets were separate (some stimuli were always presented as primes and others always presented as targets) compared to when they were intermixed (primes and targets drawn from same pool). However, these two variables were highly correlated (see Table 1), and the effects for each were nonsignificant when the other was controlled. So, these findings must be viewed with caution and will require additional research in which these variables are orthogonally manipulated.

We have argued that the attentional sensitization perspective may explain a number of important findings in this meta-analysis including the different priming across judgment tasks, but care is needed using this perspective, as it is not fully developed in evaluative priming research. For instance, the attentional sensitization perspective might predict larger evaluative priming in the EDT relative to the pronunciation task because the EDT focuses participants on evaluation. The attentional sensitization perspective on priming, however, does not explicitly address the extent to which attentional sensitization can magnify priming. That is, the research has mainly focused on making people aware of evaluation involving priming tasks and illustrating that this awareness leads to noticeable evaluative priming. It is not yet clear whether increasing the “focus” on evaluation continues to increase the magnitude of evaluative priming once a critical level of attentional sensitization has been obtained. Similarly, we suggested that reduced variability along dimensions may facilitate increased attention on the evaluative nature of the task and thus increase evaluative priming. This is a post hoc explanation in which we attempted to tie together various findings with existing theory and thus must be examined with additional research. Nevertheless, these findings and speculations provide avenues for future research. It should be possible, for instance, for researchers to manipulate attention within a task to explore whether attention can magnify the amount of priming. This could be done by intermixing evaluative trials with nonevaluative ones (e.g., pronunciation or semantic decisions), as by Spruyt and colleagues (Spruyt, De Houwer, et al., 2007; Spruyt et al., 2009).

A few findings of this meta-analysis are also consistent with a recent response perspective (Klauser et al., 2009). The evaluation window perspective was proposed primarily to account for instances within the EDT that lead to reverse (contrast) evaluative priming; that is, when people respond more slowly to targets evaluatively congruent versus evaluatively incongruent with primes. According to the evaluation window perspective, independent valence counters detect increases in stimulus activation during an evaluation window relative to initial stimulus activation before the window (Cacioppo & Berntson, 1994; Norris, Gollan, Berntson, & Cacioppo, 2010). The strength and direction of evaluative priming is determined by the temporal location of the prime stimulus during the evaluation window. Normal (assimilation) occurs when the prime is included in the evaluation window of the target, whereas reverse (contrast) priming occurs when the prime is excluded from the evaluation window. With short SOAs the prime and the evaluation window are tightly linked so prime information is captured and assimilative priming occurs; however, as the SOA increases the prime and evaluation window connection breaks down, excluding prime information from the evaluation window, and contrast priming occurs. Thus, the evaluation window perspective critically extends earlier encoding and response perspectives by providing a mechanism explaining both normal and reversed priming.

Results of the present meta-analysis involving SOA are consistent with Klauser et al.’s (2009) evaluation window perspective of the EDT. As predicted by this perspective, the strength of evaluative priming was moderated by SOA—greater evaluative priming occurred with short SOAs. To more thoroughly examine the evaluation window perspective, we conducted follow-up analyses and examined the magnitude of evaluative priming in the pronunciation tasks and EDTs that used SOAs of either 250 or 300 ms (i.e., 76% of studies used one of these two SOA values). According to the evaluation window perspective, there should be greater evaluative priming in the EDT at the short (250 ms) SOA but not in the pronunciation task because there is no response competition in this task. The findings are consistent with this prediction, as there was a significant decrease in evaluative priming from 250 to 300 ms in the EDT but no effect for the pronunciation task. It is not clear how an attentional sensitization account would explain this differential priming effect between EDT and pronunciation, so this finding suggests that response processes likely contribute to evaluative priming in the EDT.

The response instruction findings are also consistent with the evaluation window perspective. Specifically, evaluative priming was only significant when no additional instruction (e.g., “Just respond as quick and accurate as possible”) or feedback (e.g., message that participant made an error) was provided and was not

---

10 Recall that there was an interaction with stimulus presentation and judgment task. Specifically, separated stimulus sets yielded larger evaluative priming compared to intermixed stimulus sets for nonevaluative judgment tasks but not for the EDT (see above). This interaction between stimulus presentation and judgment task remained when target part of speech was controlled ($\beta = -.73, z = -2.50, p = .01$). See Klauser et al. (2007) for a review of this issue.
significant when participants were paid to boost performance. In addition, evaluative priming was weaker when feedback was provided relative to no additional instruction. Feedback and paid instructions are similar because both place an emphasis on accuracy (feedback by heightening awareness of mistakes and paid instructions by rewarding correct responses). The finding that weaker evaluative priming occurs when accuracy is emphasized is in line with previous research (e.g., Glaser, 2003) and can be explained with Klauer et al.’s (2009) evaluation window perspective (see Assumption 4 of the evaluation window perspective).

In addition to the attentional sensitization and evaluation window perspectives, a recent perspective by Schmitz and Wentura (2012) provides a possible explanation of how an encoding perspective might lead to priming in the pronunciation task but not in a semantic categorization task. According to this perspective, activation between prime and target is bidirectional. Primes make it easier to identify and encode targets congruent with primes, and targets help maintain or prolong activation of a congruent prime. Using a semantic categorization task to examine evaluative priming might lead to slow response times when primes and targets are evaluatively congruent and semantically incongruent. That is, evaluative congruency between a target and prime helps maintain activation of the prime. Because the prime and target are semantically incongruent, however, the prime activation, which occurs because of its evaluative congruity with the target, makes it difficult to initiate the semantic categorization of the target. This perspective might also predict relatively fast response times when the prime and target are both evaluatively and semantically congruent. The problem is that the two evaluatively congruent conditions (i.e., evaluatively congruent and semantically congruent; evaluatively congruent and semantically incongruent) are frequently combined and contrasted with the two evaluatively incongruent conditions. Thus, this perspective illustrates that it is important in future research not to collapse across conditions, as is common in the literature (see Herring, Taylor, White, & Crites, 2011, Experiment 3; Storbeck & Robinson, 2004). For example, researchers have frequently combined the two congruent priming conditions (positive–positive and negative–negative) and compared them to the two incongruent priming conditions (positive–negative and negative–positive). As more sophisticated theoretical models are developed, information in distinct conditions will be important. Thus, we recommend that future research examine the prime valence by target valence interaction and not collapse these into a congruity main effect.

The significant findings involving the number of discrete blocks were unexpected though potentially interesting findings of this meta-analysis. The findings revealed that the magnitude of priming increased as the number of blocks increased and that this blocking advantage was specific to verbal but not nonverbal stimuli. That is, increasing the number of blocks increased priming when either the primes or targets were words but had no effect on the magnitude of priming when primes or targets were pictures or symbols. Because almost no researchers have systematically manipulated blocks or considered them in theoretical accounts, discussion about possible causes of these effects is speculative. That said, one way to think about these effects involves attention to the task. Sequential priming tasks tend to be fairly monotonous and take some time to complete. We examined blocks and total trials because we thought there might be gradual changes across trials due to practice or fatigue or boredom. The significant effects involving blocks, in conjunction with no significant effects involving total trials, may occur because blocks provide a short break that helps people “recover” from the monotonous nature of the task and refocus attention on subsequent trials. One explanation for the absence of block effects for nonverbal stimuli may be that pictures have privileged access to the semantic system where evaluative information is stored (Bower, 1981, 1991; De Houwer & Hermans, 1994; Glaser, 1992). Thus, this privileged access is sufficient to allow priming even when people have fewer breaks provided by blocking. Verbal stimuli, on the other hand, may require more sustained attention and thus benefit from periodic breaks. Another potential explanation for this effect might be differences in bottom-up compared to top-down attention (see Kiefer, 2007; Kiefer et al., 2012). That is, some stimuli are significant enough that their presence is sufficient to capture attention and sensitize evaluative processing (i.e., attention driven by bottom-up processing of stimuli), whereas others are less salient and may require more top-down attentional focus. The majority of the nonverbal stimuli were pictures (as opposed to symbols), and many of these pictures were graphic (e.g., depict mutilation), which may better capture attention. Verbal stimuli, on the other hand, may require more top-down attentional focus benefiting from periodic breaks. This bottom-up versus top-down perspective on attention may also help explain a significant interaction between prime repetition and blocks. This interaction revealed that primes that were repeated infrequently benefited from increasing blocks, whereas primes that were repeated more frequently did not. Because primes are presented briefly in sequential priming, deciphering evaluation of primes may require top-down sustained attention, especially when they are not repeated frequently. When primes are repeated, however, less attention may be necessary to decipher the meaning of the prime.

There are a few potential limitations to the present meta-analysis. First, there was evidence of publication bias regardless of the judgment task, and all the effect sizes estimated missing were reversed (or contrast) priming effect sizes (i.e., to the leftmost of the funnel plot). This estimation is plausible because contrast findings have only recently received interest (Klauser et al., 2009), so many reversed priming effects estimated missing are probably because of publication bias. If these missing studies were located and included in the review, reanalysis of the data suggests that the weighted average effect size would reduce but remain significant and heterogeneous, suggesting that potential moderators would still exist. Because there is no way of knowing the particulars of these missing studies, it is unclear how the addition of these studies would influence the moderator analyses.

Relatively, a limitation of this review is the exclusion of articles in languages other than English. Because the language abilities of the authors were limited to English, we chose this exclusion criterion. Undoubtedly researchers should include as many languages other than English in systematic reviews when possible; however, reviews of language bias suggest that exclusion of languages other than English may not critically affect
meta-analysis (see Sutton, 2005, for review). For instance, one review of meta-analyses \((n = 79)\) found no significant difference between meta-analyses including languages other than English and meta-analyses excluding languages other than English (Moher et al., 2000). Another review of meta-analyses \((n = 50)\) showed that exclusion of studies with languages other than English would only change the overall effects by 5% in most cases (Jüni, Holenstein, Sterne, Bartlett, & Egger, 2002). In short, it is unlikely that the exclusion of articles published in languages other than English dramatically altered the results of the present review.

It is also important to note that conclusions with regard to the effect of the lexical decision and semantic categorizations tasks should be taken with more caution than those involving evaluative decision and pronunciation tasks because there were fewer effect sizes for the former tasks \((k = 6\) for each task). There are a couple of implications for the smaller number of effect sizes. One is that effects involving the LDT and semantic categorization task are much more susceptible to “file drawer” effects associated with unpublished studies. Although neither was associated with significant evaluative priming, the observed effect size in this meta-analysis may actually overestimate the true effect size given that unpublished studies are more likely to contain nonsignificant findings. Another implication is that the low number of studies made it difficult to examine variables that moderate evaluative priming in these two tasks. That is, it is conceivable that evaluative priming can occur in the LDT and semantic categorization task under certain circumstances. As discussed above, recent research suggests this may occur in the semantic categorization task (Spruyt, De Houwer, et al., 2007). Similarly, Klauer and Stern (1992) and also Wentura (2000) provide some evidence that evaluative priming can occur in the LDT and is moderated by other factors. In fact, Klauer and Stern (see also Wentura, 2000) outline a theoretical perspective to account for instances of evaluative priming in the LDT. This model is an alternative to the encoding model and provides a “response type” explanation of why evaluative priming occurs in certain judgment tasks such as the LDT and perhaps the semantic categorization task.

Similarly, the prime strength analysis included only a limited number of studies (see Footnote 9) and collapsed over different measures of prime strength. The present finding that prime strength does not moderate evaluative priming may be premature, as only one other study (Bargh et al., 1996, Experiment 1) used Fazio’s approach for assessing attitude strength (i.e., idiosyncratic attitude accessibility). In studies using Fazio’s approach, participants first undergo an attitude assessment phase in which all potential prime stimuli are presented individually and the speed with which each participant responds is used to determine which attitudes are accessible for that person. The other techniques that researchers used involved either normed reaction time data or extremity ratings for determining prime strength. To our knowledge there has been no systematic comparison of the prime strength effect between normed and idiosyncratic methods in evaluative priming.

In conclusion, the findings of this meta-analysis are consistent with past research in suggesting that priming effects are likely the result of multiple processes. That is, aspects of the findings are consistent with both encoding and response perspective, and neither perspective can completely account for the findings. As discussed above, a number of the findings are consistent with recent theoretical perspectives such as the feature-specific attention allocation (attentional sensitization) and the evaluation window, which are not mutually exclusive. Some of the application of these perspectives, however, was post hoc, and additional research will be required. A goal of future research should be to further explore and attempt to integrate these and perhaps other perspectives into a single theoretical framework of evaluative priming.

References

References marked with an asterisk indicate studies included in the meta-analysis.


*Berthet, V., Kop, J. L., & Kouider, S. (2011). Response interference in compatibility tasks: Effects of target strength in affective priming and
ON THE AUTOMATIC ACTIVATION OF ATTITUDES


(Appendix follows)
### Appendix

**Descriptions of the Primary and Secondary Moderator Variables**

<table>
<thead>
<tr>
<th>Moderator</th>
<th>Code</th>
<th>Coding description</th>
<th>Descriptive statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td></td>
<td>Total number of blocks</td>
<td>k = 125, M = 2.90, SD = 1.90;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>range = 1–10</td>
</tr>
<tr>
<td>Facilitation scores</td>
<td>1 = yes, 2 = no</td>
<td>Whether facilitation scores were analyzed</td>
<td>k = 125; yes = 113, no = 112</td>
</tr>
<tr>
<td>Judgment task</td>
<td>1 = evaluative decision task, 2 = semantic categorization, 3 = pronunciation/naming, 4 = lexical decision task</td>
<td>Judgment made on targets</td>
<td>k = 125; evaluative decision task = 76, semantic classification task = 6, pronunciation task = 37, lexical decision task = 6</td>
</tr>
<tr>
<td>Population</td>
<td>1 = academics, 2 = nonacademics</td>
<td>Whether the participants were academics (i.e., students, professors, etc.)</td>
<td>k = 123; academics = 116, nonacademics = 7</td>
</tr>
<tr>
<td>Prime conditioned</td>
<td>1 = yes, 2 = no</td>
<td>Whether primes were evaletively conditioned (i.e., made positive or negative)</td>
<td>k = 125; yes = 19, no = 106</td>
</tr>
<tr>
<td>Prime instructions</td>
<td>1 = paid attention to the prime, 2 = ignored the prime, 3 = neither told to ignore nor to pay attention</td>
<td>Primes stimulus instructions</td>
<td>k = 124; pay attention to prime = 14, ignore the prime = 20, neither = 90</td>
</tr>
<tr>
<td>Prime part of speech</td>
<td>1 = nouns, 2 = adjectives, 3 = nouns and adjectives, 4 = verbs (alone or combined)</td>
<td>The part of speech of the verbal prime</td>
<td>k = 67; nouns = 45, adjectives = 12, nouns and adjectives = 5, verbs = 5</td>
</tr>
<tr>
<td>Prime type</td>
<td>1 = symbols, 2 = words, 3 = both</td>
<td>Whether primes were nonverbal (i.e., pictures, shapes, etc.), verbal, or a combination of nonverbal and verbal</td>
<td>k = 125; symbols = 49, words = 70, both = 6</td>
</tr>
<tr>
<td>Published</td>
<td>1 = yes, 2 = no</td>
<td>Whether the article was published</td>
<td>k = 125; yes = 107, no = 18</td>
</tr>
<tr>
<td>Response instructions</td>
<td>1 = paid to respond more accurately or quickly, 2 = received error feedback, 3 = responded quickly and accurately (standard instructions)</td>
<td>Target stimulus instructions</td>
<td>k = 125; participants paid = 3, received feedback = 19, no additional instructions = 103</td>
</tr>
<tr>
<td>Reaction time analysis</td>
<td>1 = uncorrected (raw reaction times), 2 = logarithmic (log) transformation, 3 = reciprocal transformation</td>
<td>How the reaction times were analyzed</td>
<td>k = 124; reaction time data = 90, log transform = 28, reciprocal transformation = 6</td>
</tr>
<tr>
<td>Stimulus-onset asynchrony</td>
<td></td>
<td>Duration of the stimulus-onset asynchrony (ms)</td>
<td>k = 125, M = 321.48, SD = 196.49; range = 100–1,000</td>
</tr>
<tr>
<td>Stimulus presentation</td>
<td>1 = yes, 2 = no</td>
<td>Whether stimuli were kept separate or intermixed as primes and targets</td>
<td>k = 123; yes = 102, no = 21</td>
</tr>
<tr>
<td>Target part of speech</td>
<td>1 = nouns, 2 = adjectives, 3 = nouns and adjectives, 4 = verbs (alone or combined)</td>
<td>The part of speech of the verbal target</td>
<td>k = 89; nouns = 40, adjectives = 40, nouns and adjectives = 4, verbs = 5</td>
</tr>
<tr>
<td>Target type</td>
<td>1 = symbols, 2 = words</td>
<td>Whether targets were nonverbal (i.e., pictures, shapes, etc.), verbal, or a combination of nonverbal and verbal</td>
<td>k = 123; symbols = 26, words = 97</td>
</tr>
<tr>
<td>Prime repetitions</td>
<td></td>
<td>Prime repetitions (total trials/ target set = 1 [.5 if stimulus presentation = 2])</td>
<td>k = 125, M = 7.65, SD = 15.20; range = −0.5 to 127</td>
</tr>
<tr>
<td>Target repetitions</td>
<td></td>
<td>Target presentations (total trials/ target set = 1 [.5 if stimulus presentation = 2])</td>
<td>k = 123, M = 4.19, SD = 5.31; range = −0.5 to 20.33</td>
</tr>
</tbody>
</table>

Received June 11, 2012
Revision received October 23, 2012
Accepted November 6, 2012