Personality and Social Psychology Review
1–29

DOI: 10.1177/1088868317723532
pspr.sagepub.com
SAGE

Sequential Stereotype Priming: A Meta-Analysis

Ciara K. Kidder[1], Katherine R. White[2], Michelle R. Hinojos[1], Mayra Sandoval[1], and Stephen L. Crites Jr.[1]

Abstract

Psychological interest in stereotype measurement has spanned nearly a century, with researchers adopting implicit measures in the 1980s to complement explicit measures. One of the most frequently used implicit measures of stereotypes is the sequential priming paradigm. The current meta-analysis examines stereotype priming, focusing specifically on this paradigm. To contribute to ongoing discussions regarding methodological rigor in social psychology, one primary goal was to identify methodological moderators of the stereotype priming effect—whether priming is due to a relation between the prime and target stimuli, the prime and target response, participant task, stereotype dimension, stimulus onset asynchrony (SOA), and stimuli type. Data from 39 studies yielded 87 individual effect sizes from 5,497 participants. Analyses revealed that stereotype priming is significantly moderated by the presence of prime–response relations, participant task, stereotype dimension, target stimulus type, SOA, and prime repetition. These results carry both practical and theoretical implications for future research on stereotype priming.

Keywords

prejudice/stereotyping, priming, automatic/implicit processes, meta-analysis

Psychological interest in stereotype measurement has spanned nearly a century. In an early study, participants more accurately matched occupations and pictures when the appearance of an individual in a picture was stereotypical of their occupation, than when their appearance was not stereotypical of their occupation (Rice, 1926). Since then, the measurement of stereotypes has diversified. Currently, stereotype measures fall into two major categories—explicit and implicit (Fazio & Olson, 2003). Explicit measures directly ask participants for information regarding the target construct with little effort to be indirect or covert. Generally, explicit stereotype measures ask participants to report their level of agreement toward items of a stereotypic nature. For example, a study of occupational gender stereotypes asked participants to rate their agreements with statements such as "Men are better suited to certain occupations" (Yu, Yang, Lu, & Yan, 2014, p. 148). Although explicit measures have been very useful, they have limitations. Because participants are aware of what these measures aim to capture, participants have the opportunity to change their responses to portray themselves more favorably (Orne, 1962). In addition, explicit measures may fail to properly capture some of the more subtle and automatic forms of social cognition (Greenwald & Banaji, 1995). These limitations have led researchers to develop and use implicit measures to complement explicit measures.

Implicit measures attempt to measure target constructs indirectly by engaging participants in a task where the target construct affects responses in a relatively "automatic" way (De, Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009). Many implicit measures were developed specifically to explore mental associations between concepts that might be quickly activated and influence subsequent thoughts and behaviors (for review, see Gawronski, 2009). These implicit measures generally look at reaction times and/or errors that occur when people make quick judgments about stimuli that are presented in different contexts. Researchers vary the context in which stimuli are presented to see if context activation influences the speed or accuracy of participants' responses to stimuli.

Implicit measures became important and popular because they provided a basis for developing and exploring dual-process theories about stereotypes and other constructs, especially when used in conjunction with explicit measures. These theories outline conditions when behavior is driven by more "automatic" processes versus more controlled, deliberate processes and are now widely used to explain how stereotypes

[1]University of Texas at El Paso, USA
[2]Columbus State University GA, USA

Corresponding Author:
Ciara K. Kidder, Department of Psychology, Marian University, 45 S. National Avenue, Fond du Lac, WI, 54935, USA.
Email: ckkidder08@marianuniversity.edu

and many other constructs guide behavior (Chaiken & Trope, 1999; Smith & DeCoster, 2000). Implicit measures have been critical to these theoretical advances because they have been used to measure the more "automatic" processes outlined in these dual-process theories. Explicit measures, in contrast, are used to measure the more deliberate processes. In fact, implicit measures have become so popular that multiple meta-analyses have examined their relations with explicit measures and various behavioral outcomes (Cameron, Brown-Iannuzzi, & Payne, 2012; Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005).

As the utility of these dual-processes theories and implicit measures has been increasingly demonstrated, the development and employment of implicit measures has surged. Many different implicit measures have been used to examine stereotypes and related constructs, such as sequential priming (Meyer & Schvaneveldt, 1971; Neely, 1977), the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), and the Go/No-Go Association Task (GNAT; Nosek & Banaji, 2001). As the use of implicit measures has grown and they have been used to study dual-process theories, there have been tendencies to view them as interchangeable and homogeneous. However, there is reason to believe that this assumption requires more careful investigation.

Researchers have come to recognize that some implicit measurement procedures are distinct and lead to slightly different outcomes. For example, there has been considerable debate and research examining differences between two of the most popular implicit measures—sequential priming and the IAT (Marsh, Johnson, & Scott-Sheldon, 2001; Olson & Fazio, 2003; Sherman, Presson, Chassin, Rose, & Koch, 2003). The sequential priming paradigm was first employed in 1971 to study semantic associations in memory (Meyer & Schvaneveldt, 1971). In typical sequential priming studies, participants are presented with a series of trials that each contain two stimuli (prime and target). The congruency in meaning between the prime and target is varied across trials with some prime-target trials congruent in meaning and others incongruent. Participants are asked to make some response to target stimuli, and priming occurs when people respond faster to targets that are semantically related with the prime (e.g., BIRD–ROBIN) than to targets that are semantically unrelated with the prime (e.g., ARM–ROBIN). Sequential priming was later employed to study stereotypes as researchers used exemplars of stereotype groups (e.g., male or female names) and items stereotypically associated with the stereotype groups (e.g., occupations, physical traits) as primes and targets (e.g., Banaji & Hardin, 1996; Dovidio, Evans, & Tyler, 1986). By contrast, in an Implicit Association Test (IAT), participants are asked to categorize two distinct types of stimuli along different dimension (e.g., names as male/female and college majors as science/humanities). During critical trials, the response options are either congruent with stereotypes (e.g., left hand used for male names and science majors) or incongruent with stereotypes (e.g., left hand used for male names and humanities majors). Stereotype activation is thought to occur when responses are slower on trials with incongruent response option pairs than congruent response option pairs.

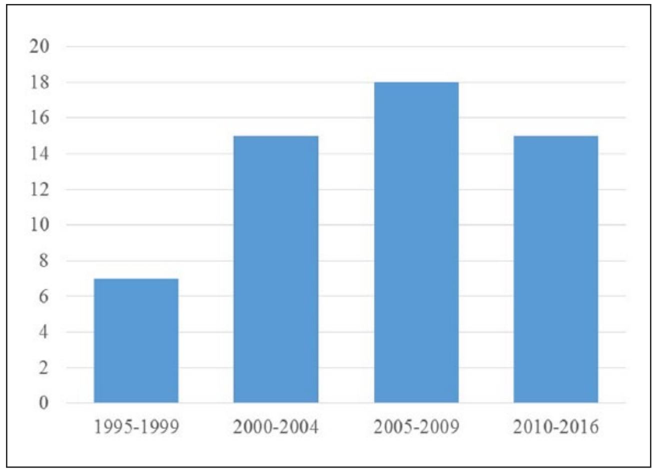

Figure 1. The use of sequential stereotype priming, based on the studies that passed the first round of exclusion criteria.
Note. Not all of the studies represented here were included in the analysis.

Soon after these techniques were widely adopted to study stereotypes and related constructs, researchers began comparing them and recognized that they were not equivalent. When used to examine the same topic, correlations between scores on these two measures have been low or close to zero (Marsh et al., 2001; Sherman et al., 2003), leading some researchers to claim that they may not measure the same constructs (Olson & Fazio, 2003). This spurred further research and theoretical discussion on the nature of these two measures (e.g., Gawronski & Bodenhausen, 2005). What has received considerably less attention, particularly for research on stereotype priming, is how changes within a given procedure (e.g., sequential priming paradigm) may impact findings.

The use of sequential priming to study stereotypes has grown since the mid-1990s (see Figure 1). Researchers have expanded from its initial use examining gender stereotypes to other stereotype dimensions such as race (e.g., Dickter & Kittel, 2012), age (e.g., Wentura & Brandtstädter, 2003), and religion (e.g., Magee, 2011). Researchers have also changed various characteristics about the paradigm, such as the judgment that participants make in response to targets, the types of primes/targets that are employed (e.g., words, pictures), and the temporal distance between prime and target. Although some of these procedural changes may seem relatively inconsequential, they have the potential to change the set of cognitive processes that are activated in sequential priming paradigms, giving rise to different findings (Klauer, Dittrich, Scholtes, & Voss, 2015). For example, Müller and Rothermund (2014) observed a gender stereotype priming when participants categorized each target stimulus as a "male" or "female" name, but there was no priming effect when participants categorized each target stimulus as a "person" or "city" name. Thus, changing the judgment participants made about targets changed the results, despite the use of a sequential priming paradigm for both conditions.

Based on findings from stereotype and evaluative priming research, as well as existing theory, we believe there are different processes at work in the sequential priming paradigm and that grouping or "averaging" across different "types" of sequential priming tasks (SPTs) obscures these differences. This argument has been articulated most clearly in work by Wentura and colleagues (Wentura & Degner, 2010; Wentura & Rothermund, 2014) and directly demonstrated in evaluative priming research. Fazio, Sanbonmatsu, Powell, and Kardes (1986) illustrated that sequential priming could be used to examine attitudes by manipulating the evaluative congruity between primes and targets (e.g., SUNSHINE–HAPPY; DEATH–HAPPY). As evaluative priming research accumulated, researchers observed strong evaluative priming effects when participants made evaluative (i.e., positive/negative) judgments about the target, but weaker (or no) priming effects were found when participants made other judgments about the target (e.g., word/not a word; verbally named the target). This led to the recognition that there are at least two types of relations that can lead to evaluative priming effects: the relation between primes and targets and the relation between primes and *responses* required by targets. Furthermore, the relative contribution of these two relations is dramatically influenced by the judgment people make about the target (Fazio, 2007; Gawronski & Bodenhausen, 2005; Klauer, Musch, & Eder, 2005; Spruyt, Hermans, De Houwer, & Eelen, 2002; Wentura, 1999, 2000).

Although there has been considerable focus on how differences in sequential priming paradigms affect evaluative priming, the same is not necessarily true for stereotype priming. Certainly, researchers have devoted considerable effort to identifying conditions when stereotype priming is more or less likely to occur (for review, see Blair, 2002), but rarely has the focus of these efforts been on how differences within the sequential priming paradigm itself might change results. This is unfortunate because differences among SPTs might help explain a number of inconsistent findings.

Cameron and colleagues' (2012) recent meta-analysis explored the relation between sequential priming measures and outcome variables (explicit measure or other behavior) and revealed an inconsistency that might be explained by differences among SPTs. An objective of their meta-analysis was to identify conditions when sequential priming measures are and are not strong predictors of scores on explicit measures or behavioral outcomes (for a number of different constructs, including stereotypes). For example, imagine that a researcher manipulates the time participants have to make a judgment about another person. Some participants have to make a very fast judgment and others have more time to weigh different factors before making their judgment. In this example, a sequential priming measure of stereotypes should be a better predictor of the behavior under the "fast" condition than the "deliberative" condition. Thus, time pressure should moderate the relation between the sequential priming stereotype measure and the outcome variable. Cameron and colleagues examined these theoretical moderators in two ways—within-study by using the predictions of the author of each study and between-study based on coding of each study by Cameron and her colleagues. The interesting aspect of the results is that the expected moderation effect was found for the within-study coding but not the between-study coding.

Cameron and colleagues discuss a number of reasons for this discrepancy between the within- and between-study moderators (the prediction was that both the within- and between-study moderator effects would be found). However, one explanation that is not discussed is that the between-study effect compares across different types of SPTs, which invoke different processes. For example, imagine that Study 1 uses a variant of the SPT that invokes cognitive process "A" and Study 2 uses a variant of the SPT that invokes cognitive processes "A" and "B." Comparing moderator effects across these two studies in a meta-analysis may not reveal the predicted moderator effect unless one also codes for the processes invoked by the different SPTs. As discussed above, research on evaluative priming suggests that there are at least two distinct processes that give rise to evaluative priming and whether one or both of these contribute depends on the judgment participants make about the target. We propose this is also the case for stereotype priming effects measured in sequential priming paradigms.[1]

One factor that may influence the magnitude of stereotype priming is the judgment participants make about targets. Researchers have asked participants to make a variety of different judgments on target stimuli in SPTs. One commonly used judgment task asks individuals to categorize targets according to the stereotype dimension of interest. For example, if gender stereotypes are being assessed, participants indicate whether the targets are words that describe males or females (e.g., Banaji & Hardin, 1996, study 1; Kawakami & Dovidio, 2001; Macrae & Cloutier, 2009). Although researchers have long recognized that this type of judgment may not get at the more "automatic" nature of stereotype priming (e.g., Banaji & Hardin, Study 2), it is still commonly used (e.g., Plaza, Boiché, Brunel, & Ruchaud, 2016). Other types of judgments are widely thought to more effectively tap into "automatic" stereotype activation because the judgment about the target is completely unrelated to the stereotype of interest. A lexical decision task (LDT) is one example of this type of judgment task. In a lexical decision version of sequential priming, two types of targets are used—words and nonwords. Participants must indicate whether each target is a word or nonword. Another type of task is the pronunciation or naming task, in which participants are asked to simply read each target out loud. Researchers have also used more focused judgments to examine specific aspects of a stereotype, such as indicating whether the target image is a “gun” or “tool.” Although there is general acknowledgment in the literature that these judgments are not equivalent, there has not been a systematic examination of the different judgment types. As a result, they are all frequently grouped into an “implicit measurement” category and differences are obscured.

There is a theoretical explanation for differences in priming effects across judgment types, suggesting that this is more than a procedural issue. De Houwer (2003) distinguishes between two potential sources of priming, and Wentura and Degner (2010) expand on the framework provided by De Houwer to differentiate among different types of priming elicited in sequential priming paradigms that use different tasks (see also Wentura & Rothermund, 2014). One source of priming identified in this work is an association between the prime and target stimuli. De Houwer (2003) refers to this as a stimulus-stimulus (S-S) relation. Priming when there is an S-S relation is thought to occur because encountering a prime facilitates processing of target stimuli that have some type of association (although see Neely, 1991; Wentura, 2000). The process that underlies priming in S-S relations is believed to be a connection between the mental representation of the prime and the mental representation of the target. Initially, spreading activation models of memory (e.g., Collins & Loftus, 1975) were used to explain how connections among mental representations in memory were activated and lead to priming, but parallel distributed processing models of memory (Rumelhart & McClelland, 1986a; Rumelhart & McClelland, 1986b) are now more commonly used to explain these mental connections and the resulting priming. Wentura and colleagues (Wentura & Degner, 2010; Wentura & Rothermund, 2014) refer to sequential priming paradigms that use these tasks as *semantic* priming paradigms because the priming is due to a semantic relationship between a prime and target.

The second source of priming identified in the works of De Houwer (2003) and Wentura and colleagues (Wentura & Degner, 2010; Wentura & Rothermund, 2014) is the relation between a task-irrelevant prime and the response required to the target. De Houwer (2003) refers to this as an irrelevant stimulus-response (S-R) relation. For instance, consider a SPT in which participants must indicate whether target stimuli are more associated with males or females. People have an explicit goal to classify each target as male–female and generate the appropriate response, but it may be difficult to focus this goal exclusively on targets given the temporal proximity of the prime and target (e.g., Klauer, Teige-Mocigemba, & Spruyt, 2009). When people encounter a congruent trial (FEMALE–NURSE), the prime can activate the same response “female” that is required for target. Alternatively, when people encounter an incongruent trial (MALE–NURSE), the prime can activate the response “male,” which must be inhibited to make the correct response “female” to the target. Thus, irrelevant S-R relations occur when encountering the prime facilitates and/or inhibits the required response to target stimuli. The processes that underlie this type of priming are thought to be similar to those in the classic STROOP task where an irrelevant stimulus feature either facilitates or inhibits the response to the task-relevant stimulus feature. De Houwer (2003) distinguished between two types of S-R tasks—irrelevant S-R and relevant S-R. We coded for both of these but there were no relevant S-R tasks in this meta-analysis. We therefore use just the term *S-R task* from this point forward for simplicity, but this should be viewed as “irrelevant S-R” in De Houwer’s framework.

The presence of S-R relations is determined primarily by the judgment that participants make about targets and can significantly alter the interpretation of observed priming effects. S-R relations occur when people make judgments about targets that can also be applied to primes in a way that varies across the congruent and incongruent trials. This S-R source of priming is absent when participants perform a task like the LDT because the response for targets in both congruent and incongruent trials is identical (e.g., “word” is correct response for congruent and incongruent trials in an LDT). Similarly, S-R relations should be absent for the pronunciation task, because responses to targets are idiosyncratic and therefore do not vary systematically with the prime stimulus categories. This raises questions regarding the proper interpretation of priming effects because, when S-R relations are present, they are typically confounded with S-S relations. For instance, when people make a male–female judgment about targets, the congruent trials (e.g., FEMALE–NURSE) presumably have both facilitative S-S and S-R relations. Alternatively, in the incongruent trials (e.g., MALE–NURSE), there is no S-S relation, and there may be an inhibitory S-R relation because “MALE” can activate a response that must be rejected to correctly indicate that “NURSE” is associated with females. Thus, priming that occurs in S-R tasks is thought to reflect a combination of both S-S and S-R relations. Wentura and colleagues (Wentura & Degner, 2010; Wentura & Rothermund, 2014) refer to any sequential priming paradigms that are likely to elicit both S-S and S-R relations as *response* priming paradigms because the priming is due, in part, to processes associated with response selection and execution. They also propose that priming effects are more robust in response priming paradigms relative to semantic priming paradigms. The extent to which effects observed in these paradigms reflect S-S versus S-R relations, however, is more difficult to determine.

We examine paradigms with only S-S relations and S-S plus S-R relations in this meta-analysis. We expect to find larger effect sizes in paradigms that have an S-R relation relative to those that only have an S-S relation (Wentura & Degner, 2010; Wentura & Rothermund, 2014). Perhaps the more interesting question is whether there is a significant priming effect in tasks that have solely S-S relations because these tasks better reflect whether stereotype activation can occur when people encounter an individual/stimulus and do not have an explicit goal that might activate stereotype processing (see discussion for more on processing goals and “automatic”).

Variables of Interest

Several variables that have been previously studied (presence of S-R relations, judgment task, stereotype dimension, the kinds of primes and targets used, and stimulus onset asynchrony [SOA]) have important theoretical implications for the stereotype priming literature. Each of these variables and how they inform stereotype theory are discussed in the following sections.

Presence of an Irrelevant S-R Relation

As previously discussed, priming effects may occur because of variations across trials in the relation between prime and target stimuli (S-S relation) and variations across trials in the irrelevant relation between a prime stimulus and the response required to the target (S-R relation) (De Houwer, 2003). Variations in the association between primes and targets (S-S relation) are an essential feature of SPTs. For example, there is a stronger association between a prime and target when they are stereotypically associated (e.g., FEMALE–NURTURING) and a weaker (or no) association when they are not stereotypically associated (e.g., MALE–NURTURING). We did not code for S-S relations as these are inherent in the congruent versus incongruent trials that are used to compute the priming effect. The presence of a relation between a prime and response (S-R relation), however, varies across different sequential priming paradigms so we coded for the presence of an S-R relations (yes/no). This resulted in 57 effect sizes with S-R relations and 30 effect sizes with no S-R relations (i.e., S-S relation only). The presence of an S-R relation depends largely on the type of judgment that participants make about targets (in conjunction with the types of targets), so presence of S-R relations is partially confounded with the judgment that participants make about a task. This will be discussed in subsequent sections.

Judgment Task

Judgment task refers to the judgment that people make in response to targets. As discussed above, judgment task has been shown to be important for evaluative priming because different judgments affect the degree to which different cognitive processes are invoked and help to explain evaluative priming. We coded five classifications of judgments used in sequential stereotype priming: Lexical Decision, Stereotype Categorization, Semantic Categorization, Weapon's Identification, and Pronunciation.

A frequently used and historically important judgment task in the literature is the LDT. In the current meta-analysis, 20 effect sizes come from studies using an LDT. During an LDT, participants see a prime (e.g., WOMAN) followed by a target (e.g., CARING) and make a judgment about whether the target is a word or nonword. The LDT includes both relevant trials where the target is a word and irrelevant trials in which the target is a nonword (e.g., RICAGN). The irrelevant trials are included to ensure that participants attend to the task but are excluded during analysis. Stereotype priming effects are examined by comparing response latencies for stereotypically congruent prime-target trials (e.g., WOMAN–CARING) and stereotypically incongruent prime-target trials (e.g., WOMAN–AGGRESSIVE). The expected result is that participants are faster to respond "word" when the prime–target pairing is stereotypically congruent. The use of the LDT for stereotype priming has had mixed results with some studies finding significant priming effects (e.g., Verhaeghen, Aikman, & Van Gulick, 2011) and other studies finding no significant priming effects (e.g., Anderson, 2011).

The second prevalent judgment task used in the literature is a stereotype categorization task. The current meta-analysis includes 24 effect sizes obtained from studies using stereotype classification tasks. In these tasks, participants make stereotype-relevant categorizations of the target. In Banaji and Hardin (1996; Study 1), for example, participants saw a prime that was stereotypically related to men or women (e.g., nurse, doctor) followed by a target pronoun (e.g., she, he) that participants classified as male or female. Significant priming effects were found when the gender stereotype of the prime matched the gender of the target. Later studies used a variety of stereotype classifications, such as classification of faces by gender (e.g., Martin & Macrae, 2007) or race (e.g., Dickter & Kittel, 2012; Stewart, Latu, Kawakami, & Myers, 2010), gendered name classifications (e.g., Blair & Banaji, 1996; Castelli, Macrae, Zogmaister, & Arcuri, 2004), and gendered object classification (e.g., Macrae & Martin, 2007; Study 1b).

Another class of judgment tasks are semantic categorization tasks. We use the term *semantic categorization task* to refer to any task that requires participants to classify targets stimuli along a semantic dimension that is (apparently) unrelated to the stereotype of interest in the study. The current meta-analysis includes seven effect sizes from semantic categorization tasks. There were a few different types of judgments used across these tasks such as requiring participants to indicate whether targets were pronoun or nonpronoun (e.g., Banaji & Hardin, 1996; Study 2), high status–low status (Mellott, 2003; Study 3), and high power–low power (Mellott, 2003; Study 4).

The fourth major category of judgment task was the weapon-identification task. In a typical weapon-identification paradigm (see Payne, 2001), participants first see a picture of a white or black face (prime) followed by a target image that is subsequently judged to be a tool or a weapon. Results show standard priming effects with faster responses and fewer errors for stereotypically congruent pairs (BLACK FACE–GUN IMAGE; WHITE FACE–TOOL IMAGE) than for stereotypically incongruent pairs (BLACK FACE–TOOL IMAGE; WHITE FACE–GUN IMAGE). The current meta-analysis includes 30 effect sizes from weapon-identification tasks. We elected to include the weapon task as a separate category because it is in some ways a cross between a stereotype categorization task and a semantic categorization task. It is not an explicit stereotype judgment (e.g., male–female; Black–White) as is typically used in stereotype categorization tasks, but the judgment is related to stereotypes (e.g., Blacks are stereotypically associated with guns) so it may differ from some of the semantic judgment tasks that ask people to make judgments that appear to have no relations to stereotypes (e.g., pronoun or not).

The pronunciation task requires participants verbally pronounce the target. Although pronunciation tasks have been used more extensively to study other constructs, they have not been used as extensively to study stereotypes. The current meta-analysis includes six effect sizes from pronunciation tasks, five of which are from unpublished research (Rothermund, n.d.).

As previously mentioned, the presence of S-R relations is partially confounded with judgment task. Specifically, S-R relations are thought to be present when a stereotype categorization task is used, but absent when the LDT or pronunciation task is used (Wentura & Degner, 2010). The weapon identification task is less straightforward but we categorized all of these as having an S-R relation because the response and prime are confounded, similar to those in the stereotype categorization tasks. There is more on this in the "Discussion" section. Finally, semantic classification tasks are split with about half being classified as S-S and the remainder S-R.

Stereotype Dimension

Stereotype dimension refers to the social group that is associated with particular social beliefs. We can have stereotyped beliefs about any group of people based on nearly any dimension in which we can categorize people as different, including skin color, age, gender, nationality, occupation, and so on. For the most part, researchers examining stereotypes have not distinguished among these different dimensions with regard to the strength or functionality of the stereotype. That is, stereotypes involving gender, for example, are thought to be generally equivalent to those for occupations. Stereotype research, however, has tended to focus on dimensions that apply to nearly everyone and along which we can categorize people based on their appearance—especially race and gender. We were able to find five stereotype dimensions with race and gender being by far the most common; the current meta-analysis includes 44 effect sizes from studies examining race stereotypes and 38 effect sizes from studies examining gender stereotypes. The other dimensions were occupation, vehicle type, and weight. These three stereotype dimensions contributed four effect sizes to the current meta-analysis.

Although stereotype research has not systematically compared different stereotype dimensions, research in other areas suggests that dimension might be an important variable that determines the relative magnitude of stereotype effects. In comparing the most commonly studied social dimensions of race and gender, research has found that race is associated with very early (less than 250 ms) differences in brain activity that do not occur for gender (Ito & Urland, 2003). Such preferential attention to race may increase the likelihood that people will be categorized along the race dimension. This categorization would then activate associated concepts in memory, producing a stereotype priming effect for race that is potentially stronger than for gender or other stereotype dimensions (at least in certain circumstances).

Research examining event-related brain potentials also suggests that the extent to which race and other stereotype dimensions are activated may depend on the judgment task that people perform. Differential brain activity due to race, for example, occurs relatively early and in tasks where people do not direct their attention to social categories whereas effects for gender occur later when people are performing more complex social judgments (Ito & Urland, 2005). Extending these findings to the judgment tasks used in stereotype priming research suggests there might be differential effects for race and gender across judgment tasks. That is, the stereotype categorization tasks in which people categorize target stimuli as being more associated with a particular social category (e.g., Black/Whites or male/female) involve more complex social judgment relative to the LDT in which people categorize target stimuli as being a word or nonword. In the LDT, there might be stronger stereotype priming effects for race relative to gender (and other stereotype dimensions) because race attention/categorization appears to occur relatively automatically. However, more equivalent stereotype priming across dimensions might be obtained in stereotype categorization tasks because these tasks force people to attend to stereotype dimensions and allow more inconspicuous dimensions to be activated.

Stimulus Type

Stimulus type refers to whether the prime and target stimuli are verbal (i.e., words) or nonverbal (i.e., pictures). In the current meta-analysis, 35 effect sizes were obtained from studies that used verbal primes, 49 effect sizes were obtained from studies that used nonverbal primes, and three effect sizes were obtained from studies that used both verbal and nonverbal primes. With regard to target stimuli, 50 effect sizes were obtained from studies that used verbal targets and 37 effect sizes were obtained from studies that used nonverbal targets. Although the impact of prime and target type on stereotype priming in sequential priming paradigms have not been studied, a number of theoretical arguments suggest this dimension might affect stereotype priming. First, Glaser (1992) has proposed that people possess both lexical and semantic encoding systems and verbal and nonverbal stimuli may function somewhat differently in priming paradigms due to the way they are processed in these systems. Pictures are postulated to have direct connections with the semantic encoding system, whereas words have only indirect connections as they must be processed by the lexical encoding system before being sent to the semantic system. As a result, pictures (and other nonverbal stimuli) may be more effective in semantic priming paradigms. Second, Macrae and Bodenhausen (2000) argue that social category activation should only occur following the perception of a person and not a category label because person perception is not symbolically equivalent to words. Third, other person categorization research shows that exemplar categorization, as might occur to category labels (e.g., women), differs from prototype categorization that would occur to pictures of individuals (Smith & Zarate, 1990). These arguments suggest that different kinds of primes may activate different kinds of processes (e.g., a process for persons/images versus a process for categories/words) which in turn could affect the speed of responding, and ultimately the magnitude of the priming effect where we could see greater priming effects with nonverbal primes. Although this has not been explored in stereotype priming, a recent meta-analysis on evaluative priming found that nonverbal stimuli produced larger evaluative priming effect than verbal stimuli (Herring et al., 2013). The current meta-analysis will examine whether these effects are also true for sequential stereotype priming.

Stimulus Content

Stimulus content refers to what the prime and target stimuli depict (e.g., category labels, jobs, traits, etc.). Content is best thought of as a more precise way of examining stimulus type, where we are examining the particular prime and target stimuli in more detail. For example, when coding prime content, there were three types of nonverbal primes (i.e., body/action, faces, or objects) and five types of verbal primes (i.e., categories, external traits, internal traits, names, and pronouns). Although we cannot fully divorce stimulus content and stimulus type, examining prime and target content will tell us more about how different kinds of verbal and nonverbal stimuli affect stereotype priming effects.

Stimulus content has been studied to some extent in the stereotype literature but effects on stereotype priming seem inconsistent. Dovidio et al. (1986) examined stereotypic personality traits, and Banaji and Hardin (1996) examined stereotypic occupations, and both found significant priming effects. However, Blair and Banaji (1996; Study 1) specifically examined content and found significant stereotype priming effects for nontrait attributes (e.g., occupations) but not for personality traits. These results were not replicated in a subsequent experiment (Study 2) which found significant stereotype priming effects across both types of stimuli. With these inconsistencies, a careful examination of stimulus content is warranted. The current meta-analysis will examine both levels of prime and target classification (type and content) to determine whether there are significant differences that may inform both theory and methodology.

SOA

SOA refers to the temporal interval between the onset of the prime and the onset of the target. Because of proposed differences between automatic and controlled processing (e.g., Blair, 2002; Blair & Banaji, 1996; Devine, 1989), we examined SOA as a moderator variable. In semantic priming, longer SOAs result in stronger priming effects (for review, see Neely, 1991). However, some studies show reverse effects, with stronger priming effects at shorter SOAs, and no priming at longer SOAs (e.g., Kawakami, Dion, & Dovidio, 1998). Stronger stereotype priming effects at shorter SOAs is similar to those in evaluative priming (Herring et al., 2013). In the current meta-analysis, SOAs ranged from 70 ms to 2,200 ms.[2] Based on the distribution of SOAs, we categorically coded them into three categories: "short" (<200 ms, $k = 5$), "typical" (200-350 ms; $k = 70$), and "long" (>350 ms; $k = 11$).

Other Variables

In addition to these variables of theoretical interest, several other methodological variables were examined because previous research suggested they might affect the magnitude of priming effects. First, to understand how repetition affects the priming effect, we examined the number of experimental trials completed by participants, the number of trails per block, and the repetition of primes and targets over the course of the experiment. Second, we coded whether or not the study included supplemental priming methods intended to boost the stereotype priming effect or to reduce the stereotype priming effect.[3] The current meta-analysis includes 16 effect sizes from studies using boosting methods, four effect size from studies using reducing methods, and 65 from studies that did not attempt to influence the effect size with additional methodology. Third, we also examined whether presenting the stereotype information in the prime or target position affected stereotype priming.

The Present Meta-Analysis

The goals of the current investigation were to examine the effectiveness, limitations, and theoretical contributions of the sequential stereotype priming paradigm. Specifically, we explore whether the primary variables—presence of S-R relation, judgment task, stereotype dimension, stimulus type, stimulus content, and SOA—affect the reliability and magnitude of the stereotype priming effect. Furthermore, we examine some critical interactions between these primary variables. To accomplish these goals, the current meta-analysis includes 87 independent effect sizes from 39 individual studies.

Method

Literature Review

To gather studies for inclusion in the current meta-analysis, the authors conducted literature searches for published manuscripts. Published journal articles were collected from aggregated searches using EBSCOhost that included the following databases: PsycINFO, the Psychology & Behavioral

Sciences Collection, PsycARTICLES, and the Science and Technology Collection. We conducted two separate Boolean searches of these aggregated databases using different search terms. The first search used the terms "stereotype" AND "priming," and the second search used the terms "automatic" AND "stereotype." Each search was further defined by the following search options: "Select a Field (optional)," "Search Related Terms," and "Exclude Book Reviews." These searches encompassed manuscripts from May 1996 (Banaji & Hardin's initial paper) to October 2016. Published dissertations were collected from ProQuest within the same time frame. We again conducted two separate searches with different terms. The first search used the search terms "automatic" AND "stereotype" OR "prejudice," and the second search used the search terms "stereotype" AND "priming" OR "prime." These searches included the "Anywhere but in Text" option selected.

The four literature searches yielded 776 individual hits.[4] The abstracts of all hits were examined by the authors for any obvious reasons the article should be excluded (see exclusion criteria below). Articles that did not obviously meet the exclusion criteria were retained for more extensive review and coding.

Inclusion/Exclusion Criteria

Due to the broad search terms, further criteria were established to determine inclusion in the current meta-analysis. If all criteria were met, the study was included in the analyses. Figure 2 shows the reasons and numbers for exclusion.

1. Article reported at least one empirical study.
2. A sequential priming paradigm was used. This criterion eliminated similar studies examining automatic stereotypes using nonsequential implicit measures such as the IAT.
3. Study design was symmetrical, meaning that primes and targets included stimuli from both groups being compared in the study. For example, a study on gender stereotype priming should include primes for both men and women (e.g., pictures of men *and* women) plus target stimuli for both men and women (e.g., traits stereotypically associated with men *and* women). An asymmetric design would present stimuli for only one of the groups for either the primes or targets. For example, presenting only male primes paired with male and female target stimuli. In this type of design, response time differences may reflect genuine priming effects. Alternatively, they may simply reflect a main effect of the primes or targets (e.g., participants respond more slowly to stimuli following that category of primes). Without a symmetric design, this confound cannot be eliminated, and thus, studies with an asymmetric design were not included.

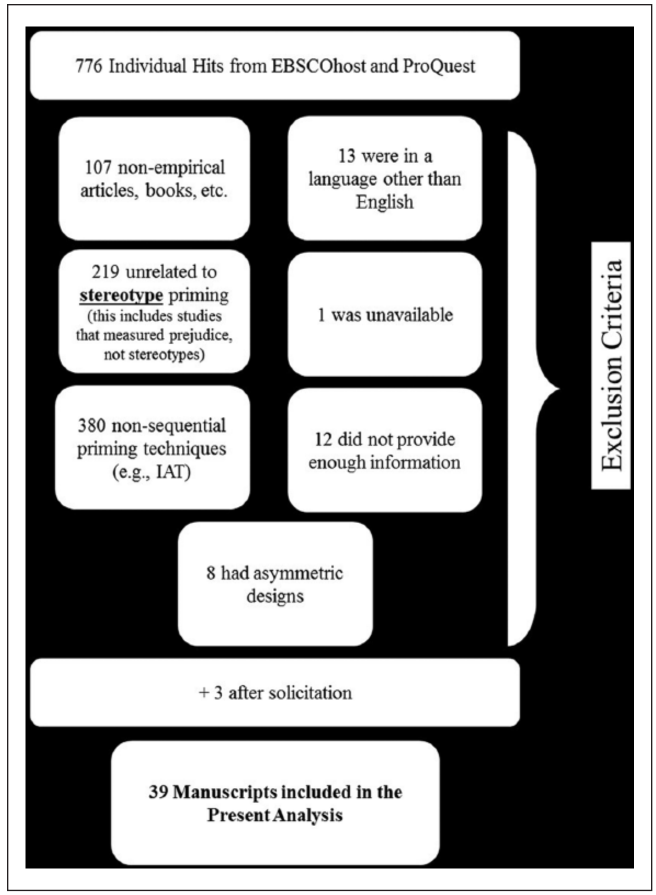

Figure 2. Flow chart of the exclusion of studies.
IAT = Implicit Association Test.

4. Appropriate methodological and statistical information was reported in the manuscript, provided by the corresponding author, or could be estimated based on the information provided.
5. The article was available in English or Spanish as limited by authors' language abilities.

Requests for Statistical and Methodological Information

Several studies met most of the inclusion criteria but were missing information. Some studies were missing the statistics needed to calculate effect sizes for the congruity effect because they used shorthand statistical notations (e.g., "n.s.", "F <1", etc.) or did not report the main effect or interaction that tested the specific congruity effect. Other studies failed to report important methodological details such as prime duration. We contacted 19 corresponding authors for the missing statistical and methodological information. Fourteen authors replied (74%), and of these, 12 (63%) supplied the requested information, allowing for those data to be included in the analysis when appropriate.

Coding and Reliability

The authors coded for potential moderator variables using a scoring manual (Appendix). All of the studies were coded by two of the three coders to ensure accuracy in coding. (Coding for all studies are available to review on the Open Science Framework website.) Any discrepancies were resolved through discussion.

Meta-Analytic Procedures

Effect size calculations. The measure of effect size used in the present meta-analysis is *Cohen's* d_z. A stereotype priming effect is represented by a positive d_z, which reflects people responding more quickly and/or accurately to stereotypically congruent trials than stereotypically incongruent trials. Larger $d_{z's}$ indicate greater priming effects.

Effect sizes were calculated from F and t statistics reported in the included studies. These statistics were derived from either a congruity main effect (incongruent trials–congruent trials) or from prime (e.g., Black and White) by target (e.g., Black and White) interactions. When the exact F or t values were known, effect size was calculated using a formula to help account for the correlation between conditions in repeated measure designs which are used in sequential priming.[5] The d_z calculation below is used when calculating effect sizes where F or t are for repeated measures designs:

$$d_z = \frac{t}{\sqrt{n}} \text{ OR } d_z = \sqrt{\frac{F}{n}},$$

where t and F represent the respective reported statistics for each study and n represents the total number of participants.

When the exact F or t value was unknown, but the authors reported that it was nonsignificant or $p > .05$, an effect size of 0 was entered for that study.[6]

Once an effect size was obtained for each study, we also calculated the variance of d using the following formula (Morris & DeShon, 2002):

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

which includes a correction factor when the correlation between conditions is unknown:

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

Statistical independence. One of the major requirements of meta-analyses is that the effect sizes are statistically independent (Lipsey & Wilson, 2001). Statistical independence is commonly violated when variables are manipulated within subjects. There were a few cases in which effect sizes in the present meta-analysis were not independent, and we selected one effect size to include and the other(s) were excluded. The most prevalent instances of statistical dependence occurred in 16 studies in which both reaction time and error rates were reported. For analysis, the effect size for reaction time was used to keep variability related to dependent variable low (because the majority of effect sizes come from reaction times), unless a response window was used in the task (k = 25). In those cases, the error rates were used, as the RTs would be constricted. Another instance of nonindependent effect sizes occurred with SOA, which was manipulated within subjects for two of the included studies. SOA was averaged, and the effect size of the overall prime-target congruency was reported.

Analytic approach. All meta-analytic models were conducted using the R program "metafor" (Viechtbauer, 2010), using the inverse-variance method. When appropriate, the R program "car" (Fox & Weisberg, 2011) was used to conduct contrasts among variables of interest.

Sequential stereotype paradigms often vary across the literature in a number of ways so several manipulated variables were coded. We divided our analyses of these variables into two sections to reduce the number of analyses as we considered interactions between particular moderators. All moderator analyses were conducted using a mixed-effects approach where the moderator of interest was added to the model as a categorical or continuous moderator.

The primary moderators are the most theoretically relevant variables that were reviewed in the introduction: (a) presence of irrelevant S-R relation (S-S only paradigms or S-R paradigms), (b) judgment task, (c) stereotype dimension, (d) format of stimuli (prime and target type), (e) content of the stimuli (prime and target content), and (f) SOA. In addition to main effects for each of these moderators, we also examined interactions among them. Because presence of S-R relation, judgment task, and stereotype dimension was of particular interest and we predicted certain interactions involving these variables, we ran models to test how these variables interacted with each other and several of the remaining moderators.[7]

The secondary moderators are more exploratory and/or reflect methodology that has been shown to influence effects in other sequential priming paradigms (see Herring et al., 2013). The secondary moderators include (a) the number of trials, (b) number of trials per block, (c) the number of repetitions per prime and target (prime and target repetition), (d) supplemental priming, and (e) location of stereotype information (prime or target). Interaction effects were only examined for the last of these secondary moderators.

To determine how moderators influence priming effects, we ran mixed-effects model with the moderator as depicted in the following example equation:

$$y = judgment + error. \tag{1}$$

Table 1. Nominal-by-Nominal Associations Between Primary Moderators.

		S-R	Judgment	Dimension	Prime type	Target type	Prime content	Target content
S-R relation	*Cramer's V*							
	Sig.							
Judgment task	*Cramer's V*	.583	.928					
	Sig.	<.001	<.001					
Dimension	*Cramer's V*	.394	.556					
	Sig.	.001	<.001					
Prime type	*Cramer's V*	.166	.528	.433				
	Sig.	.311	<.001	<.001				
Target type	*Cramer's V*	.583	.630	.520	.365			
	Sig.	<.001	<.001	<.001	<.001			
Prime content	*Cramer's V*	.506	.545	.617	.777	.441		
	Sig.	.003	<.001	<.001	<.001	.003		
Target content	*Cramer's V*	.695	.659	.743	.654	.893	.631	
	Sig.	<.001	<.001	<.001	<.001	<.001	<.001	
SOA	*Cramer's V*	.489	.366	.264	.081	.217	.246	.297
	Sig.	< .001	.004	.018	.892	.092	.739	.775

Note. SOA = stimulus onset asynchrony.

When assessing interactions, we took a two-step approach. First, we ran a mixed-effects model that included the two moderator variables as depicted in the following example equation:

$$y = classification + judgment + error. \tag{2}$$

This equation allows us to determine how these variables may overlap in accounting for variance. For instance, if task classification accounts for 20% of the heterogeneity in effect sizes and judgment task accounts for 10%, the combined equation should account for 30% if the two moderators do not account for any of the same variance (i.e., the moderators are unrelated). In the second step, we added the interaction of the moderators to the model to determine whether the interaction accounted for any additional variability as depicted in following equation:

$$\begin{aligned} y = classification + judgment + \\ classification \times judgment + error. \end{aligned} \tag{3}$$

This allows us to look at whether the interaction helps to explain variance above and beyond the two moderators alone.

Publication bias. A common problem with meta-analyses is the problem of publication bias, which can lead to a review and conclusions that are not representative of the population of studies (Lipsey & Wilson, 2001). Publication bias assumes that published papers in a given area have larger effects than unpublished papers of the same topic. To prevent publication bias in the present meta-analysis, a number of steps were taken to locate unpublished works. First, when we requested additional statistical/methodological information from authors as discussed above, we asked those authors to send us unpublished information. Second, we wrote to a number of active and prominent authors doing sequential stereotype priming to request unpublished work. Third, we sent out a "call for papers" on the listserv for the Society of Social and Experimental Psychology and the Society for Personality and Social Psychology general forum. Our efforts led to the inclusion of 19 effect sizes from unpublished work.

In addition to our active attempts to find unpublished data, we also conducted analyses on the effect sizes included in this meta-analysis to explore whether publication bias might be present. We assessed publication bias for the data in the present meta-analysis by first examining the funnel plot, which shows the distribution of individual effect sizes around the estimated "true" effect size (Sterne, Becker, & Egger, 2005). Meta-analytic data that do not have publication bias will be symmetrically distributed whereas meta-analytic data with publication bias will show an uneven distribution. Second, we assessed the extent of any publication bias using the trim and fill approach (Duval, 2005). This approach uses an iterative procedure to estimate the number of studies (and their estimated effect sizes) required to make the funnel plot symmetric and calculates a new overall effect size that includes those studies. These analyses and their interpretations are reported below.

Results

The current meta-analysis includes data from 39 studies with 87 individual effect sizes (N = 5,497). The predicted heterogeneity of effect sizes was examined in the context of the primary moderators of interest. The results of those analyses are described in the section below. To assess the correlations between the primary moderators, nominal-by-nominal nonparametric tests were run. The results of these tests identify significant correlations among all of the moderators (see Table 1 for values).[8] Because these variables are significantly correlated, the variables are potential confounds of one another so we conducted interaction models among the primary moderators. Additional exploratory models were run on secondary moderators and are described thereafter.

Primary Moderators

Presence of an irrelevant S-R relation. As described in the introduction and "Method" section, sequential stereotype priming paradigms can be classified in two ways, those where priming effects result from only relationships between the prime and target stimuli (S-S relation) and those in which the prime might have an independent effect on the response that is required to the target (S-R relation). To determine how these different kinds of paradigms influence priming effects, we ran a mixed-effects model with S-R relation (yes/no) as the categorical moderator (as depicted in Equation 1). S-R relation accounted for 40.06% of the heterogeneity of effect sizes (k = 87; $Q_M(1) = 39.08$, $p < .001$). Significant stereotype priming effects were found for both paradigms. Studies where tasks were S-S only (k = 30, d = 0.10, SE = 0.05, p = .032, 95% CI = [0.01, 0.20]) produced significantly smaller priming effects than studies where tasks were also S-R (k = 57, d = 0.48, SE = 0.04, p < .001, 95% CI = [0.41, 0.55]); $\chi^2 = 39.08$, $p < .001$. Figure 3 shows forest plots for the S-S only and S-R paradigms. These plots show the weighted individual effect sizes compared with the average weighted effect size of each paradigm, as described above.

Judgment task. To get a more nuanced understanding of how the different tasks affected priming effects, we also ran a mixed-effects model with judgment task as the categorical moderator. Across the studies, there were five different judgment tasks used: lexical decision, pronunciation, semantic categorization, stereotype classification, and weapons tasks. Judgment task accounted for 40.94% of the heterogeneity of effect sizes (k = 87; $Q_M(5) = 43.96$, $p < .001$). Two of the judgment tasks, stereotype classification and the weapons task, are tasks that are classified as S-R tasks and significant stereotype priming was found in both, stereotype classification task (k = 24, d = 0.52, SE = 0.06, p < .001, 95% CI = [0.40, 0.63]) and the weapons task (k = 30, d = 0.46, SE = 0.05, p < .001, 95% CI = [0.36, 0.56]). Furthermore, the size of priming effects in these two tasks did not differ significantly ($\chi^2 = 0.52$, $p = .470$). The LDT (k = 20, d = 0.16, SE = 0.09, p = .069, 95% CI = [−0.01, 0.34]) and the pronunciation task (k = 6, d = 0.02, SE = 0.11, p = .839, 95% CI = [−0.20 0.24]) which are S-S only tasks, were not associated with significant stereotype priming effects, and were also not significantly different from one another ($\chi^2 = 0.20$, $p = .652$). Both S-R tasks had significantly larger stereotype priming effects than the S-S only tasks (χ^2s > 11.00, ps < .001). The semantic categorization tasks are interesting because approximately half of these were classified as S-S tasks and half were classified as S-R tasks. The stereotype priming in this task was significant (k = 7, d = 0.37, SE = 0.10, p < .001, 95% CI = [0.18, 0.57]). Even with the split S-S and S-R classification, stereotype priming in the semantic categorization tasks overall did not significantly differ from the stereotype priming in the stereotype or weapons S-R tasks (χ^2s = 1.57 and 0.63, ps = .210, .428, respectively) and was significantly larger than the stereotype priming in the lexical decision and pronunciation S-S tasks (χ^2s = 6.48 and 5.31, ps = .011, .038, respectively).

S-R Relation × Judgment. To further explore the unique nature of the semantic classification tasks, we examined an interaction between S-R relation and judgment task. We would not expect large changes in the heterogeneity accounted for in these models because classification and task are confounded, with only semantic tasks classifiable as S-S or S-R. This is confirmed by a significant nominal-by-nominal association between task classification and judgment task (see Table 1; φ = .928, *Cramer's V* = .928, ps = .001).

The process for the interaction analysis is described earlier. The dual-moderator model (the two moderators entered singularly) accounted for 39.77% of the heterogeneity of effect sizes (k = 87; $Q_M(6) = 43.22$, $p < .001$). This model accounted for slightly less heterogeneity than the judgment task model and task classification model. We next ran the interaction model (adding the interaction between classification and judgment to the previous model). This model accounted for no new heterogeneity, with significant stereotype priming effects in semantic S-R studies (k = 3, d = 0.41, SE = 0.15, p = .006, 95% CI = [0.12, 0.69]) and in the semantic S-S stereotype priming effects (k = 4, d = 0.34, SE = 0.14, p = .012, 95% CI = [0.08, 0.61]). It is important to note that there are few studies in each classification and there is high variance in effect sizes within the categories. This may explain why no significant difference was found even though a larger effect size was found for S-R semantic studies than for S-S only semantic studies ($\chi^2 = 0.10$, $p = .756$).

Stereotype dimension. Six stereotype dimensions (race, gender, health, mixed, occupation, and vehicles) were examined by the studies in the present meta-analysis. Of these, all but five of the studies examined either race stereotypes (k = 43) or gender stereotypes (k = 39). The other dimensions were aggregated into an "other" category. A mixed-effects model with stereotype dimension as the moderator variable was run using a similar equation as those described above. Stereotype dimension accounted for 13.82% of the heterogeneity among the effect sizes (k = 87; $Q_M(2) = 11.42$, $p = .003$). Studies examining race stereotypes (d = 0.44, SE = 0.05, p < .001, 95% CI = [0.35, 0.53]) produced significantly larger priming effects than studies examining gender stereotypes (d = 0.29, SE = 0.05, p < .001, 95% CI = [0.20, 0.39]), $\chi^2 = 4.62$, p = .032. Other dimensions did not produce significant priming effects (d = 0.00, SE = 0.14, p = 1.000).

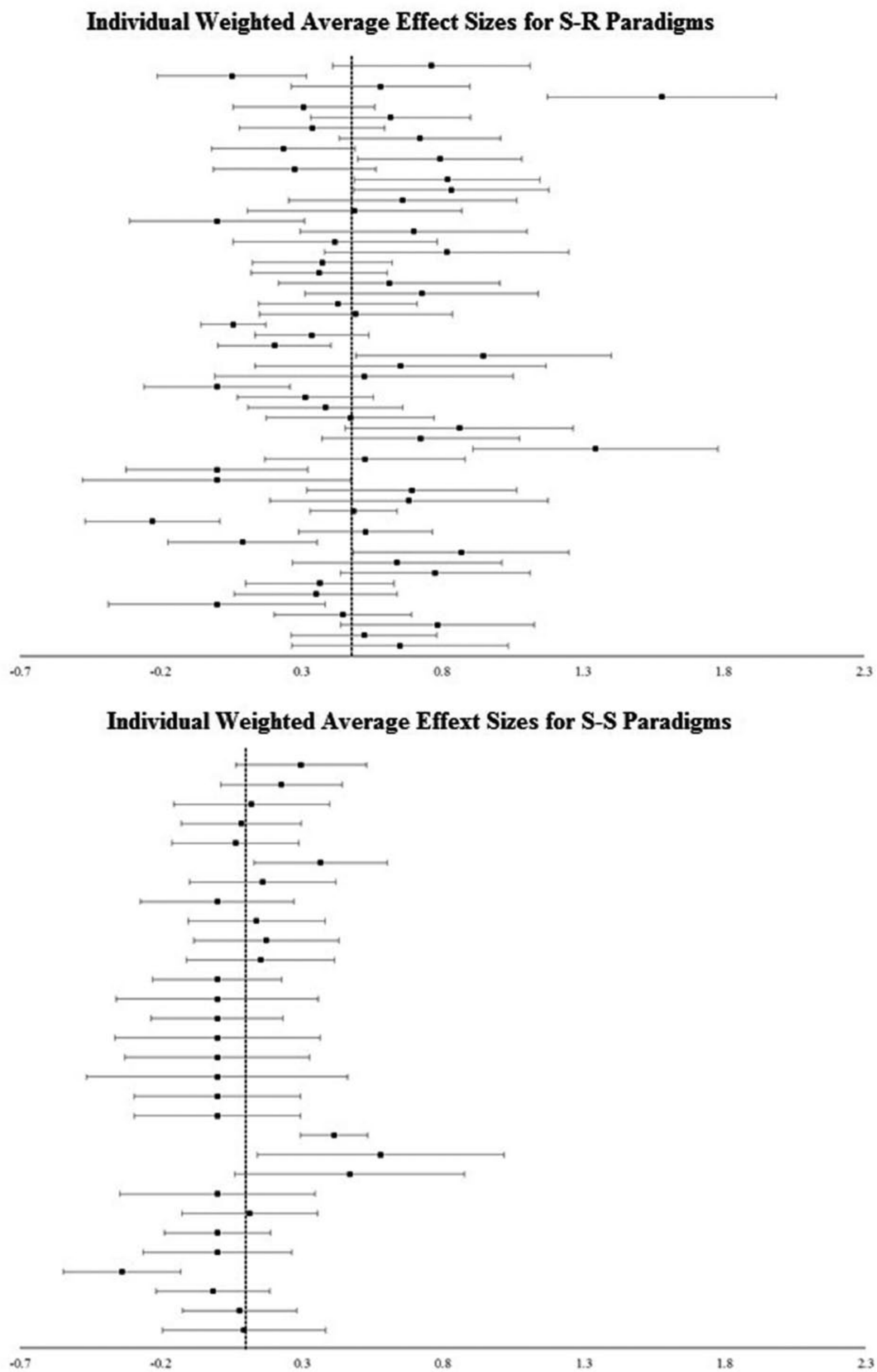

Figure 3. Sequential stereotype priming effects by S-R relation.

Note. The forest plots for studies with S-R relation components (top panel) and S-S only studies (bottom panel). The weighted average effect size of the included studies (black squares), their 95% confidence intervals (bars), and the overall effect for each paradigm (dotted lines).

Dimension interactions. We ran models similar to those described above to determine whether S-R relation and dimension interacted. Adding dimension to the S-R relation model decreased the understanding of heterogeneity of effect sizes by less than half of 1%. An analysis of the interaction between S-R relation and dimension revealed no significant differences in the effect sizes of the S-R relation between race and gender. There was a significant nominal-by-nominal association between S-R relation and stereotype dimension such that S-R tasks were more often used to assess race stereotypes whereas S-S tasks were more often used to assess gender stereotypes (see Table 1; $\varphi = .394$, *Cramer's V* = .394, *p*s = .001).

Stimulus type. Among the included studies, primes and targets were coded as either images, words, or both. We examined how these different kinds of primes and targets affected the size of the priming effect. A mixed-effects model with prime type as the moderator variable was run using a similar equation as those described above. Primes that included both images and words ($k = 3$, $d = 0.57$), images only ($k = 49$, $d = 0.36$), and words only ($k = 35$, $d = 0.31$) all produced significant priming effects (*p*s < .003). Prime type did not significantly account for any variance in the effect sizes across studies ($k = 87$; $Q_M(2) = 1.80$, $p = .406$, $R^2 = 0.00\%$). We also ran several interactions for prime type (with S-R relation, judgment, and dimension). We found a significant interaction between prime type and S-R relation, with the interaction model accounting for 49.19% of the heterogeneity in the model. The interaction's effects are similar to those found in the S-R relation model and the prime type model, with significant differences based on S-S or S-R relations, but not between prime type. Because of those findings, it seems that the significant interaction model is being driven by a very large effect size for an S-R study that uses both images and words which is significantly larger than any of the other average effect size combinations. Because this is based on one effect size, it is difficult to know whether the interaction is reliable.[9]

Another mixed-effects model with target type as the moderator variable was run using a similar equation as those described above. Target type accounted for 5.81% of heterogeneity in the effect sizes across studies ($k = 87$; $Q_M(2) = 11.26$, $p = .004$). Image targets ($k = 36$, $d = 0.48$, $SE = 0.05$, $p < .001$, 95% CI = [0.37, 0.58]) produced significantly larger priming effects than word targets ($k = 50$, $d = 0.26$, $SE = 0.04$, $p < .001$, 95% CI = [0.17, 0.34]); $\chi^2 = 10.58$, $p = .001$. Only one study used targets with words and images. We also ran several interactions for target type (with classification, judgment, and dimension); however, there were significant nominal-by-nominal correlations between these variables which would explain why these interactions did not account for any additional heterogeneity than previous models.

Stimulus content. Among the various stimulus types, the content of those stimuli also varies. We examined how these different kinds of primes and targets affected the size of the priming effect. A mixed-effects model with prime content as the moderator variable was run using a similar equation as those described above. As with prime type, prime content did not significantly account for any variance in the effect sizes across studies ($k = 84$; $Q_M(6) = 11.03$, $p = .14$, $R^2 = 4.55\%$). Internal traits ($k = 4$, $d = 0.73$), mixed content ($k = 7$, $d = 0.41$), faces ($k = 47$, $d = 0.38$), names ($k = 8$, $d = 0.30$), and "other" content ($k = 8$, $d = 0.33$) produced significant priming effects (all *p*s < .010), but there were no differences among them.

Another mixed-effects model with target content as the moderator variable was run using a similar equation as those described above. Target content accounted for 15.10% of the heterogeneity in the effect sizes ($k = 85$; $Q_M(9) = 19.95$, $p = .02$). Faces ($k = 4$, $d = 0.63$) and objects ($k = 31$, $d = 0.47$), both of which constitute image targets, produced significant priming effects (*p*s < .001) but did not significantly differ from one another. Name targets ($k = 9$, $d = 0.45$), targets that were mixed in content ($k = 19$, $d = 0.23$), and internal trait targets ($k = 7$, $d = 0.27$), which constitute word targets also produced significant priming effects (*p*s < .02) but did not significantly differ from one another.

SOA. SOA ranged from 70 ms to 2,200 ms across the included studies. Although SOA could be examined as a continuous variable, the number of effect sizes is not distributed evenly across this continuum. Priming research has generally used "short" SOAs or "long" SOAs. Relatively few studies have SOAs falling between 400 and 1,000 ms; to confirm this, we examined the distribution of SOAs and found three distinct clusters. The largest grouping consisted of studies using SOAs between 200 ms and 350 ms ($k = 70$), a second grouping of SOAs 500 ms and larger ($k = 11$), and a third grouping of SOAs less than 200 ms ($k = 5$). So we categorized SOA into those three clusters called short (SOAs < 200 ms), typical (SOAs = 200 to 350 ms), and long (SOAs ≥ 350 ms).[10]

A mixed-effects model with SOA as the categorical moderator variable was run using a similar equation as those described above. SOA accounted for 10.26% of the heterogeneity of effect sizes across studies ($k = 86$; $Q_M(2) = 8.73$, $p = .013$). Studies using typical SOAs ($d = 0.37$, $SE = 0.03$, $p < .001$, 95% CI = [0.30, 0.45]) and short SOAs ($d = 0.48$, $SE = 0.14$, $p < .001$, 95% CI = [0.20, 0.77]) both produced significantly larger priming effects than studies using long SOAs ($d = 0.10$, $SE = 0.09$, $p = .270$, 95% CI = [−0.08, 0.27]), χ^2s = 7.68 and 5.13, *p*s = .006 and .020, respectively. Short SOAs and typical SOAs were not significantly different from each other ($\chi^2 = 0.57$, $p = .450$).

SOA × S-R Relation. To explore how SOA influences priming effects in the different paradigms, we examined an interaction between S-R relation and SOA. We first ran a mixed-effects model that included S-R relation and SOA as moderator variables similar to the equations described above.

The dual-moderator model accounted for 38.87% of the heterogeneity of effect sizes (k = 86; $Q_M(3) = 38.74$, $p < .001$). This model accounted for less heterogeneity than the singular SOA model and task classification model. Adding the interaction component further decreased the heterogeneity accounted for (37.54%).

Secondary Moderators

Repetition. Priming paradigms are often susceptible to practice effects. Practice effects can be caused by repetition of judgments made by the participants which can be examined in a variety of ways. We coded possible contributors of practice effects by recording the number of experimental trials completed by participants, the number of trials per block, the number of times primes were repeated, and the number of times targets were repeated. We ran mixed-effects models for each of these continuous variables. Prime repetition was the only significant model and accounted for 5.44% of heterogeneity in the model (k = 78, $Q_M(1) = 4.02$, $p = .045$). The stereotype priming effect increases by 0.0031 for every prime repetition.

Supplemental priming. A number of studies measuring stereotype expression attempt to determine the conditions under which priming effects increase or are diminished. We coded studies based on whether they include methods intended to boost priming effects (e.g., "use race" instructions), methods intended to reduce priming effects (e.g., creativity mind-set), or no supplemental priming. We ran a mixed-effects model with supplemental priming as a categorical moderator. Supplemental priming did not account for heterogeneity of effect sizes across studies (k = 78; $Q_M(3) = 3.52$, $p = .318$). This means that boosting methods (k = 16, d = 0.50, SE = 0.08, $p < .001$, 95% CI = [0.34, 0.66]) did not, on average, increase priming effects compared with studies not using supplemental priming (k = 65, d = 0.33, SE = 0.04, $p < .001$, 95% CI = [0.25, 0.42]; $\chi^2 = 3.21$, $p = .073$), nor did reduction methods (k = 4, d = 0.41, SE = 0.16, $p = .09$, 95% CI = [0.10, 0.73]) decrease priming effects compared with studies that did not use supplemental priming; $\chi^2 = 0.22$, $p = .639$.

Stereotype location. Sequential stereotype priming tasks have presented stereotypic information in either the prime (k = 22) or the target (k = 56) locations. We ran a mixed-effects model with stereotype location as a categorical moderator. Stereotype location did not account for heterogeneity of effect sizes across studies (k = 78; $Q_M(1) = 0.37$, $p = .541$). Significant stereotype priming was found when stereotypical information presented as primes (d = 0.42, $p < .001$) or targets (d = 0.36, $p < .001$). We also examined the interaction of stereotype location and S-R relation. There was an increase in heterogeneity with the interaction component (R^2 = 32.01%), compared with the dual-moderator approach (R^2 = 30.94%). Neither the S-S studies with stereotypes as primes (k = 3, d = .28, SE = .16, $p = .068$, 95% CI = [−0.02, 0.60]) nor those with stereotypes with targets (k = 20, d = .11, SE = .06, $p = .079$, 95% CI = [−0.01, 0.22]) showed any significant stereotype priming. Both S-R studies with stereotypes as primes (k = 19, d = .44, SE = .08, $p < .001$, 95% CI = [0.29, 0.59]) and those with stereotypes with targets (k = 36, d = .49, SE = .05, $p < .001$, 95% CI = [0.41, 0.58]) revealed significant stereotype priming; however, there was no significant difference between them ($\chi^2 = 0.40$, $p = .526$).

Publication Bias Analysis

Because the focus of the present meta-analysis was the distinction between paradigms with or without S-R relation, we examined publication bias for these paradigms separately. For each, we first examined publication bias by creating a funnel plot where effect sizes were plotted along the horizontal axis and standard errors were plotted along the vertical axis (Sterne et al., 2005). The resulting funnel plots are depicted in Figure 4. For the S-R paradigms, it appears like there are missing studies in the lower left quadrant, suggesting that there may be missing studies with small effect sizes and large standard error. For the S-S only paradigms, the funnel plot appears fairly symmetrical, which may suggest slight or no publication bias. This method of assessing publication bias is quite subjective, so a more objective approach is warranted.

To assess the potential asymmetry observed from the funnel plot, the trim and fill method (Duval, 2005) was used to identify whether there were any missing studies for each of the different paradigms, and if so, the number of studies needed to balance out the distribution. The method also provides an adjusted estimate of the effect. For the S-R paradigms, the trim and fill method confirmed the assessment of the funnel plot and identified 18 (SE = 6.13) possible missing studies. Figure 5a depicts the funnel plot with estimated missing studies filled in, and as described earlier, many are in the lower left quadrant, suggesting that studies with small effect sizes have been suppressed. Thus, the adjusted stereotype priming effect size for studies using tasks with S-R related components is smaller (d = 0.24, SE = 0.04, $p < .001$, 95% CI = [0.16, 0.31]) than the previously estimated priming effect (d = 0.48). This analysis suggests that if publication bias is the cause of the unsymmetrical distribution of effect sizes, then the stereotype priming effect in S-R related paradigms is smaller than this meta-analysis predicts.

For the S-S paradigms, the trim fill method identified five (SE = 3.65) possible missing studies. Unlike the S-R studies, the studies are missing from the right side of the plot. Figure 5b depicts the funnel plot with the estimated missing studies filled in and suggests that studies with larger effect sizes are missing. Here, the adjusted stereotype priming effect size for studies using tasks with only S-S components is slightly larger (d = 0.14, SE = 0.03, $p < .001$, 95% CI = [0.07, 0.21]) than the previously estimated priming effect (d = 0.10). This

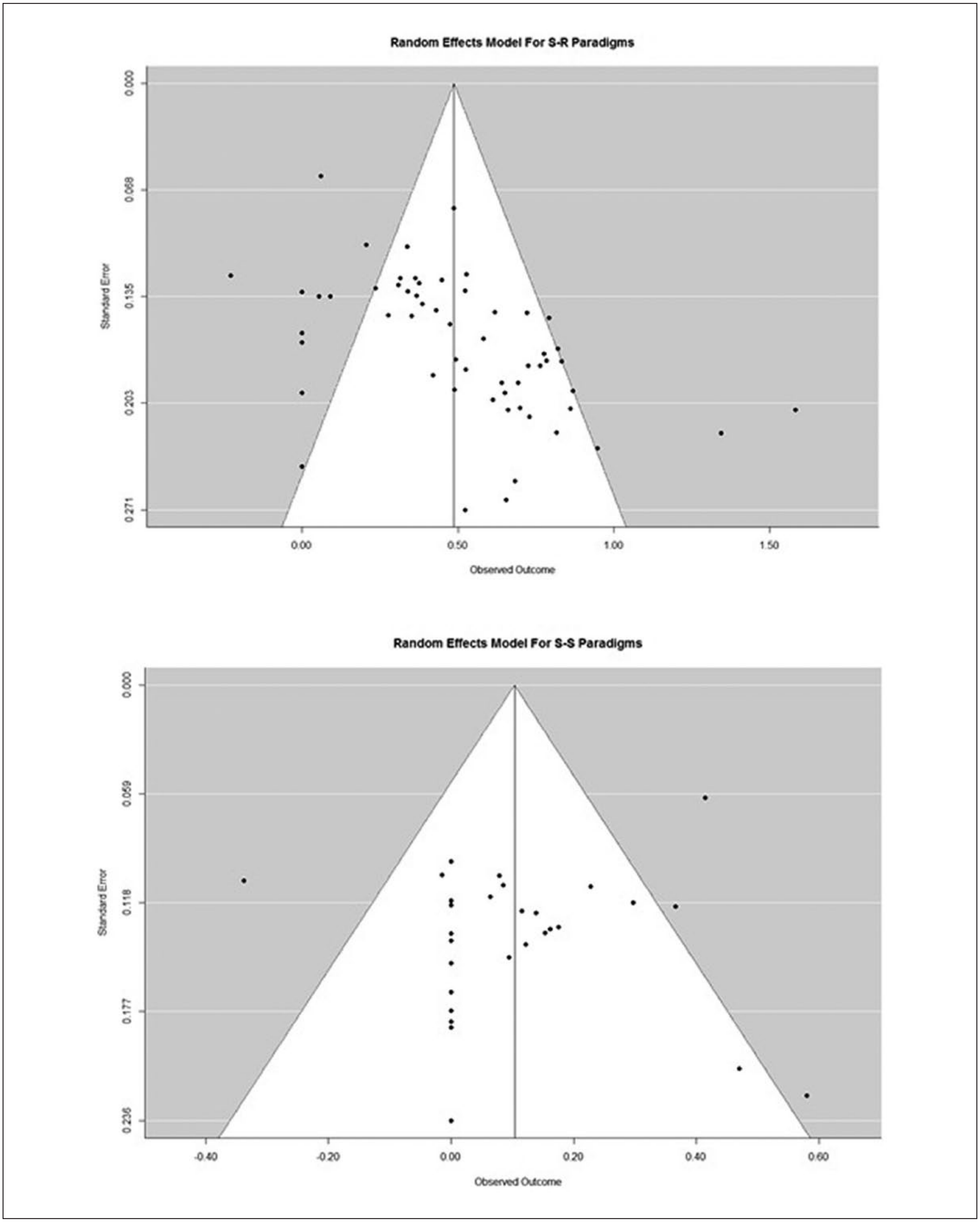

Figure 4. Funnel plot of the weighted average effect sizes for S-R relation studies (top panel) and for S-S only studies (bottom panel).

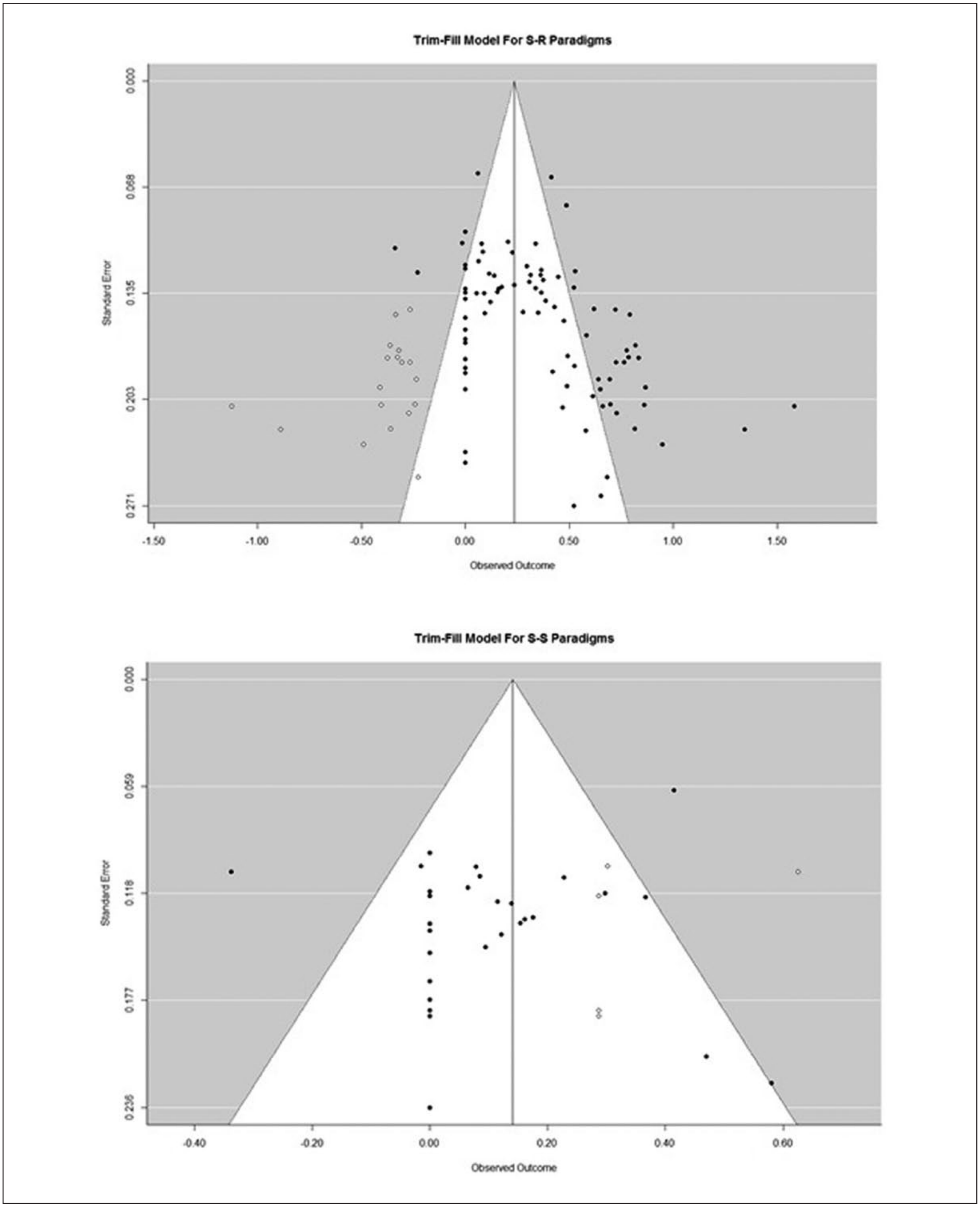

Figure 5. Funnel plot of the adjusted weighted average effect sizes for S-R relation studies (top panel) and for S-S only studies (bottom panel).

Note. The studies included in the current meta-analysis are shown as black circles, the estimated missing studies are shown as white circles.

analysis suggests that if publication bias is the cause of the unsymmetrical distribution of effect sizes, then the stereotype priming effect in S-S related paradigms is slightly larger than what this meta-analysis predicts.

Discussion

The impetus for initiating this meta-analysis was the increasing number of inconsistent results from our own lab that used SPT to explore stereotypes. In some studies, we found robust and significant stereotype priming; in other studies, significant results were ephemeral. In looking for answers in the stereotype priming literature, there seemed to be a general awareness of differences between various implicit measures, but not necessarily awareness of differences that may exist within a single implicit measure. The purpose of this meta-analysis was to identify the primary ways SPTs differ from one another and to provide a theoretical framework for understanding these differences. The present meta-analysis examined 39 studies that contained 87 effect sizes and spanned 20 years of research. As expected, the analysis revealed significant heterogeneity in the size of the stereotype priming effect across the different studies.

One way that SPTs differ from one another is the specific cognitive processes that give rise to observed priming effects. We focused on two ways of classifying SPTs that might capture these distinctions. The first classification was based on whether the SPT allows for a direct association between the prime and the *response* required to the target, called an S-R relation in addition to an association between the prime and target, called an S-S relation. The second, and related, classification was based on the type of judgment that people make to the target. These distinctions revealed a pattern of results that suggest the body of work using the SPT to examine stereotypes is comprised of two distinct types of SPTs—one that evokes strong stereotype priming and another that is associated with significantly weaker stereotype priming.

The present meta-analysis revealed the expected significant heterogeneity in the size of the stereotype priming effect across the two primary categories of SPTs. The first category includes tasks in which priming is thought to primarily reflect variations in the mental association between prime and target stimuli (S-S relation). The second includes tasks in which the priming may also be due to variations in the association between a prime and the *response* to the target (S-R relation) in addition to the S-S relation between the prime and target (see De Houwer, 2003; Wentura & Degner, 2010; Wentura & Rothermund, 2014; see also Note 1). A critical variable for determining whether an SPT is exclusively S-S or adds an S-R relation is the judgment that individuals make about the target (hereafter, we refer to these as S-S tasks or S-R tasks for simplicity, but reader should realize that S-R tasks include S-S and S-R relations). The present meta-analysis coded for five distinct types of judgments—two that result in an S-S classification (LDT and pronunciation), two that result in an S-R classification (stereotype categorization and weapons task), and one (semantic classification) that has some S-S and some S-R tasks, depending on the nature of the judgment. Analyses revealed significant stereotype priming for both the S-S and S-R tasks, but the priming effect was significantly stronger for S-R tasks (stereotype classification and weapons). These results are not too surprising as they largely mirror those found in the evaluative priming literature (Herring et al., 2013). However, they had not been directly demonstrated or examined for stereotype priming until now. As discussed previously, one potential explanation for stronger priming in judgment tasks that have both S-S and S-R relations is that these tasks have two underlying processes that contribute to the priming effect. One process is the connection between the mental representations of the prime and target that allows the target to be more easily recognized and processed in congruent trials (Collins & Loftus, 1975; Rumelhart & McClelland, 1986a, 1986b, 2014; see Ratcliff & McKoon, 1988, for another account). The second process is facilitation or inhibition of the response to the target that occurs because primes activate a response that is either consistent with and facilitates the required response to the target or is incongruent with the required response to the target and must be inhibited to make the correct response (Gawronski & Bodenhausen, 2005; Klauer et al., 2005; Spruyt et al., 2002; although see Neely & Keefe, 1989).

Another explanation for the increased priming that is observed in S-R tasks is that the judgments people make about targets in these tasks strengthen the S-S relation between prime and target. Spruyt and colleagues (Spruyt, De Houwer, & Hermans, 2009; Spruyt, De, Houwer, Hermans, & Eelen, 2007), for instance, have demonstrated that priming in S-S paradigms can be enhanced when characteristics of task focus more attention on a specific dimension that is being examined (feature-specific attention allocation). One characteristic of S-R tasks is that people make a judgment about the target that is related to stereotypes, and this may keep attention focused on stereotypes throughout the experiment. In stereotype categorizations tasks, for instance, people make male/female (or Black/White) judgments about targets throughout the experiment. This attentional focus on gender (or race) may enhance the mental associations between the prime and target stimuli (i.e., S-S relation). Thus, the enhanced priming in tasks that have S-R relations may be partly due to increased S-S connections in these tasks relative to other S-S tasks that do not focus attention on the stereotype dimension (e.g., LDT or pronunciation). Although feature-specific attention allocation may increase the strength of S-S relations in S-R tasks, one finding from this meta-analysis raises some slight concerns about the impact of this attentional focus. Specifically, we coded whether studies used supplemental procedures in an attempt to either boost or reduce priming. This analysis revealed that these supplemental procedures did not, on average, have any effect. Presumably, the boosting supplemental procedures should have drawn attention to the stereotype dimension of interest and increased priming. The lack of findings from these supplemental procedures, however, may not be the best way of examining whether feature-specific attention increases the strength of S-S relations because these supplemental procedures are often not integrated with the SPT in a way that keeps attentional focus on the stereotype dimension throughout the task. For example, one technique that has been used to boost stereotype priming is to have participants to categorize stimuli along a stereotype dimension prior to doing an S-S task such as the LDT. This may create a temporary attentional focus that quickly dissipates when participants engage in the LDT. Thus, although boosting techniques in theory should focus attention to the stereotype dimension that is being examined, the extent to which the manipulations engage attentional focus may not be sufficient for an adequate test of this theoretical idea.

In addition to feature-specific attention, the judgments that people make in S-R tasks may also increase the S-S relation between the prime and target by forcing people to more deeply process the semantic meaning of the target and prime stimuli. The findings revealed small but significant findings in the collection of S-S tasks, but the follow-up analyses hint that there are differences across S-S tasks. Specifically, the analyses revealed that stereotype priming was significant in semantic S-S tasks but not in the lexical decision or pronunciation S-S tasks. This finding is in line with previous research suggesting that semantic processing of stimuli in a LDT may be too shallow to elicit significant priming (e.g., Shulman & Davidson, 1977; Stone & Van Orden, 1992 although see Lupker & Pexman, 2010). One explanation for the significant effects in semantic classification tasks, relative to LDT and pronunciation, is that the latter two tasks do not require individuals to process the meaning of a stimulus and priming is stronger when individuals access the meaning of a stimulus. Tasks that require people to access the semantic meaning of stimuli may result in slightly better semantic priming relative to those that do not. Because S-R tasks require participants to semantically process the targets, this may also result in slightly stronger S-S relations in these tasks. Thus, the increase in priming for tasks that have an S-R relation, relative to those with only an S-S relation may be explained by at least three distinct theoretical processes—(a) facilitation/inhibition of the response to the target due to a connection with the prime, (b) greater feature-specific attentional focus that highlights the mental connection between prime and target, and/or (c) deeper semantic processing of primes and targets that helps activate the mental connection between them. These explanations are not particularly new to the literature, but they do have important implications for the proper interpretation of stereotype priming effects, particularly with regard to "automaticity."

There has been considerable theoretical discussion over the years as to the nature of automaticity, and several perspectives conceptualize it as a collection of features (e.g., Bargh, 1994; Moors & De Houwer, 2006). The results of the present meta-analysis have implications for at least two of these features as they apply to stereotype activation. One important feature of automaticity for stereotype research is unintentionality because whether people activate and use stereotypes when they do not intend to do so is of great theoretical and practical interest. Stereotype priming that occurs in both S-S and S-R SPTs may be considered unintentional in the sense that people do not have a goal to evaluate and use the prime when making a judgment about the target. The prime is essentially a distractor because it typically provides no information that can help individuals respond to the target (nearly all SPTs have a 50/50 ratio of congruent and incongruent prime–target pairs for the critical trials). However, intentionality is a function of an individual's goals, and the goals that people have when engaged in typical SPTs differ across S-S and S-R tasks. In prototypical S-S tasks, people have a goal that is *not* related to stereotypes, such as making word/nonword decisions or pronouncing the target. In contrast, typical S-R tasks involve a goal of making an explicit judgment that is directly related to the stereotype of interest such as indicating whether target is male/female for a gender stereotype task (or Black/White for a race stereotype task). The amount of stereotype priming is significantly greater in S-R tasks in which people have an explicit stereotype-relevant goal relative to S-S tasks in which the goal is unrelated to stereotypes. This raises important questions about the activation and use of stereotypes. These results suggest that stereotype activation may not be unconditionally automatic, which has been argued for quite some time (e.g., Blair, 2002). Instead, they suggest that processing goals are a crucial factor. This raises several questions. What types of contexts and/or judgments are likely to lead to stereotype-relevant goals? Are there individual differences in the propensity to have underlying stereotype-processing goals? The findings of this meta-analysis show that we need to focus more processing goals and other variables that can affect stereotype activation.

A potentially interesting outcome of this meta-analysis concerns the semantic categorization tasks in which people make a semantic judgment about targets that is not directly related to the stereotype dimension of interest. The semantic judgment tasks are interesting because approximately half were categorized as S-R tasks and half as solely S-S tasks. Firm conclusions regarding these tasks cannot be made based on the present meta-analysis because (a) there were relatively few effect sizes from these tasks (four from S-S tasks and three from S-R tasks), and (b) there was considerable overlap in the methods of the S-R tasks because two of the three S-R tasks came from a single dissertation (Mellott, 2003). Nevertheless, there are two reasons the findings suggest that semantic tasks should receive increased research attention in future studies. First, the overall difference between S-S and S-R semantic judgment tasks, though not significant, was consist with the larger pattern of findings that S-R tasks evoke greater stereotype priming relative to

S-S tasks. This raises the possibility that semantic judgment tasks can be created to further probe the various theoretical processes that have been proposed to explain stereotype priming. A very subtle change in the semantic judgment and/or the target stimuli, for example, may be able to switch a task from an S-S type to an S-R task. This might allow for the relative contributions of the memory activation and response compatibility processes that drive priming to be better estimated. Similarly, changing aspects of these tasks could allow the contributions of feature-specific attention allocation to be probed. Second, the semantic priming task was the only S-S task that revealed significant stereotype priming when the three distinct tasks that comprise the S-S tasks were examined individually. That is, stereotype priming was not significant in either the lexical decision or pronunciation tasks even though these two tasks have been much more widely used. As discussed above, the genesis for this meta-analysis was our own research in which we attempted to use the LDT to examine stereotype priming and found largely nonsignificant results. It is important that researchers realize this and use tasks that are more likely to be effective. The S-S type semantic tasks provide one potential option, with the important caveat that there were relatively few of these tasks in the current meta-analysis, rendering this conclusion somewhat preliminary.

The weapon-identification task is another interesting case that deserves more discussion. In the standard weapon task (Payne, 2001), people see a picture of a White or Black male face (prime) followed by a target image that must be categorized as either a tool or weapon. We considered this task separately because the weapon-identification task was the most prevalent type of SPT and appears to share features with both the stereotype and semantic categorization tasks. The stereotype categorization tasks are typically ones in which people make a very explicit stereotype judgment about the target that could also be applied to the prime (e.g., judgment about targets is male/female and primes can also be categorized as male/female). This is not the case for the weapon-identification task because the judgment about the target (e.g., gun/tool) is not applicable to the primes, which are typically pictures of Black and White males. The gun/tool judgment, however, is an aspect of the race stereotypes, as guns are stereotypically associated with Black men and tools are presumably more stereotypically associated with White men. The results revealed that the effect sizes for the weapon-identification task are comparable with those from the stereotype categorization tasks ($d = 0.46$ and 0.52, respectively) and are significantly greater than those from the S-S tasks (d = 0.10 for combined S-S tasks). This pattern of results has implications for the nature of the processes that underlie the weapon-identification task.

One possibility is that participants conceptualize the judgment about the target in a weapon-identification task in such a way that the pictures of White or Black males can activate a response that is consistent with the response required to the target. For example, a way to think about the tool/weapon judgment is that weapons are dangerous so a component of the weapon/tool judgment may be "dangerous/safe." This same type of judgment could also be applied to individuals, and participants may see Black males as "dangerous" and White males as "safe" (e.g., Wilson, Hugenberg, & Rule, 2017). This type of framing may allow variations in the S-R relation between the prime and response that are similar to those in the stereotype categorization studies. In fact, Todd, Thiem, and Neel (2016) did this in a recent study where they mimicked a traditional weapon task but instead used targets that were words implying threat or safety and found evidence for significant stereotype priming (this study was coded as a semantic S-R task in the present meta-analysis). We categorized the weapon task as an S-R task because the design allowed for the possibility that there was an irrelevant relation between the prime and the response required to the target, but the nature of this relation was not immediately obvious (as it is in stereotype categorization tasks). The dangerous/safe task is one example of how an S-R relation may exist, but there are other ones (e.g., positive/negative). More research will be necessary on the weapon tasks to see whether there is some type of irrelevant S-R relation that is producing the strong priming in this task or whether there is something else about the weapon task that evokes the strong stereotype priming.

Another factor besides an S-R relation that might account for the strong stereotype priming in this task is that it focuses on a very specific racial stereotype that Blacks are more associated with guns, and perhaps danger, relative to Whites. This makes the weapon task different from the other stereotype priming tasks which attempt to assess broader stereotype content (e.g., many different types of traits). This difference may change the nature of the task and may activate processes other than memory encoding. The weapon-identification task is similar to visual detection tasks in which people have to indicate whether a certain stimulus is present or not (e.g., "is there a gun in the image"). The threshold that people use to make their judgments affects performance in these types of tasks (Correl, Park, Judd, & Wittenbrink, 2002). Certain variables may lead people to adopt a conservative standard and delay responding until they are absolutely certain that the stimulus is present, whereas other variables may lead people to adapt a more liberal standard and initiate a response when they are "somewhat" certain that the stimulus is present. The race of the person in the prime stimulus might change the response threshold to the target. If it does, the strong effects in the weapon-identification task, relative to the other implicit tasks, may be due to both memory encoding (as in the other implicit tasks) and an additional psychological process such as a changing response threshold that is not present in the other implicit tasks.

As discussed above, there has been relatively little theorizing about the effects of stereotype dimension (e.g., gender, race, age) in the stereotype literature as stereotype activation and use are generally believed to be equivalent across different dimensions. The present meta-analysis coded three different stereotype dimensions—race, gender, and "other." The results suggest that stereotype dimension may be important—stereotype priming was significant for both dimensions of race and gender, but race was associated with significantly larger priming effects (d = 0.44, k = 43) than gender (d = 0.29, k = 39). No priming was observed for the other stereotype dimensions (d = 0.00, k = 4), perhaps due to high variability among the included effect sizes (SE = 0.14). Caution should be used when interpreting these results, however. We found that S-R tasks were more often used to examine racial stereotypes and S-S tasks were more often used to examine gender stereotypes. Given that effect sizes from S-R tasks were generally larger, greater priming for race may simply be an artifact of participant task. Future research that orthogonally manipulates stereotype dimension and participant task is needed to determine whether priming is truly stronger for race than gender.

As with task classification and dimension, SOA was a critical moderator of interest for this meta-analysis. That being said, SOA has received more consideration in the literature than the others (e.g., Clow & Esses, 2007; Neely, 1977). In this meta-analysis, we categorically coded SOA into three groups: SOAs that reflected the "typical" interval used in stereotype priming studies (SOAs 200 to 350 ms; k = 70), SOAs that were "short" in comparison (SOAs < 200 ms; k = 5), and SOAs that were "long" in comparison (SOAs > 350 ms; k = 11). The priming effect with short and typical SOAs was significantly larger than priming effects with long SOAs. This finding is interesting given that longer SOAs are associated with stronger, not weaker, priming effects in research on semantic priming (for review, see Neely, 1991). The presence of stronger priming with longer SOAs is generally credited to the operation of both automatic and controlled processes at long SOAs versus primarily automatic at short SOAs. While stronger semantic priming at long SOAs may be typical, the absence of stereotype priming at longer SOAs has been observed before (Kawakami et al., 1998). The proposed explanation in this article was that in a typical stereotype priming study, the nature of the primes and targets is likely to alert participants to the topic of study—stereotypes/prejudice. With this in mind, participants might be motivated to respond to targets in a socially desirable way, which would be more possible at longer SOAs than short. This pattern of findings is also found in evaluative priming paradigms with increased priming effects for smaller SOAs (Herring et al., 2013). A more systematic investigation of the impact of SOA on stereotype priming effects is warranted to understand the nature of these effects.

Results from several other analyses provide guidance with regard to best practices for conducting stereotype priming research and/or highlight areas where additional research is needed. When it comes to prime and target stimuli selection, targets appear to require more careful consideration. The type of prime (images, words, mixture) and prime content (e.g., traits, faces, names) do not appear to moderate the stereotype priming effect. By contrast, target type does moderate priming. Specifically, stereotype priming was significantly stronger when target stimuli were images than when target stimuli were words. However, these different image types were confounded with the type of judgment task (and thus the presence of S-R relations), that people made about the target so caution is warranted until this finding can be confirmed in controlled experiments. It is also valuable to note that prime (but not target) repetition appears to increase the amount of stereotype priming. This may speak to the efficiency (or inefficiency) of stereotype activation. In SPTs, people must process and make a judgment about the target but primes can typically be ignored. As primes are repeated, people may be able to identify them more readily and make the connections more easily between the prime and target that allows for stereotype priming. One explanation is that stereotype activation is not highly efficient and can only occur when sufficient cognitive resources are available. Familiarity with primes through repetition may reduce cognitive load and allow for stereotype activation. This explanation is very speculative but may warrant exploration in future research.

Although presenting stereotypical information as a prime or target did not affect the magnitude of stereotype priming, the location of the stereotypical information, particularly in S-R tasks, may affect the processes and assumptions about stereotype priming findings (D. Wentura, personal communication, March 6, 2017). For example, imagine an S-R gender stereotype task in which people indicate whether targets reflect males or females and the stereotypical information is presented either as a target (e.g., "FEMALE PICTURE"–NURSE) or as a prime (e.g., NURSE–"FEMALE PICTURE"). When stereotypical information appears as a target, people are explicitly using stereotypes to make the male/female judgment, and stereotype priming may reflect the fact that the prime picture is categorized as "female," which initiates a response that facilitates the response to the target. In this case, priming may mostly reflect how gender (category) priming facilitates making explicit stereotype judgments. However, when the stereotype information appears as a prime, the judgment about the target is not as explicitly stereotypical. For instance, people can categorize target pictures as male/female based on physical features that presumably have less overlap with the mental connections that form stereotype knowledge. If the stereotype content from primes activates a "male" or "female" response, this is more clearly a reflection of stereotype associations.

The findings of this meta-analysis suggest that the size of priming effects in these two paradigms is equivalent, but that does not necessarily mean that the underlying processes are equivalent. The case where stereotype content is in the target position (and an S-R task is used) is more similar to an explicit stereotype measurement task and may correlate with and better predict explicit behaviors than when the stereotype content is in the prime position. Presenting stereotype content in the target position does, however, more closely approximate conditions that most interest researchers—encountering a member of a social group and then examining how this encounter influences later responses. If researchers wish to present stereotype content in the target position, they might consider using an S-S task instead of an S-R task. Results from the present meta-analysis would suggest that a semantic classification task is a better option than the LDT or pronunciation task. Alternatively, they could use an S-R task, but utilize procedures/techniques that examine the distribution of RTs and errors to tease apart the individual contributions of different processes (e.g., Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Voss, Nagler, & Lerche, 2013). These may help determine the extent to which observed priming effects are specifically due to distinct underlying processes.

The findings of this meta-analysis show that in sequential priming paradigms, stereotype priming effects are genuine, occurring across varied designs. We also showed that the primary influence on the magnitude of priming effects is due to the presence (or absence) of S-R relations, accounting for more than 40% of heterogeneity. SPTs with only S-S relations are generally thought to better capture "automatic" stereotype activation, but our results suggest that these stereotype priming effects are relatively weak. By contrast, stereotype priming effects containing irrelevant S-R relations are much stronger. The presence of S-R relations is determined largely by the judgment task participants perform, demonstrating that participant task is a theoretical as well as a procedural consideration when designing stereotype priming studies. The current meta-analysis also revealed that stereotype dimension, SOA, and target type affect the magnitude of the priming effect. The unusual findings for SOA provide a particularly interesting avenue for future research as they seem to suggest that stereotype priming may operate differently from semantic priming.

Appendix

Meta-Analysis Coding Manual

Any word(s) underlined are defined below in the "Glossary" section

Once we begin the actual coding process (i.e., after practice), you'll be given a "special instructions" sheet that will inform you what information to ignore or focus on for certain papers.

The code "–99" is used when information is missing, "–88" is used when the code is not applicable, and "–77" is used when you're not sure how to code something.

If you're coding multiple experiments in a study, assume that if the authors state only how the follow-up experiments changed, all else remained the same. For example, if in Experiment 1 the SOA was 150 but the authors only mentioned how Experiment 2 was different from Experiment 1, then assume the SOA was 150 in Experiment 2 also.

A. "ID"

- This info will be provided in the e-mails I send with your coding assignments

B. "Study #"

- We're defining the "paper" as the work as a whole (i.e., the article). A "study #" is more specific. For example, "Mal ran three experiments in his study." Because experiments are labeled in a number of ways, let's assign the first experiment "a," the second experiment "b," and so on. So if you're informed that you'll be coding Mal's experiment "b," you should be looking for the second experiment in Mal's study.
- Note: This variable is not asking the number of experiments in the study to then be converted to letters (e.g., there were two experiments so you put a "b"). Just use "a," "b," and so on sequentially for the experiments as you code.

C. "Total N"

- The number of participants in the study or the total sample size

D. "# EXCLUDED"

- Some studies report the number of participants that were excluded from the analysis
- If reported, enter that number
- If not reported, code as missing

Info Generally Found in the Stimuli & Procedure *Sections of Method of Articles for Coding*

E. "STEREOTYPE"

- What kind of stereotype is being examined in the study?
- **Gender**
- **Racial**
- **Other**

F. "STEREOTYPE OTHER"

- If other above, enter the kind of stereotype examined.

G. "Stereotype Comparison"

- Some stereotype studies make specific comparison. For example, in racial stereotyping, some studies make comparisons between African Americans and Caucasians.
- If a specific comparison is made, enter the comparison, for example, Black–White, Black–Hispanic, and so on.

H. "Prime Stereotype?"

- Does the prime set contain the stereotypic information?
- For instance, a prime could be gender or occupation. In gender priming, the occupation is the stereotype.
- **No, Yes.**

I. "PRIME TYPE"

- Were the primes evaluative symbols, words, or both?
- **Images (e.g., pictures, shapes, Chinese symbols, etc.)**
- **Words**
- **Both**

J. "PRIME CONTENT-WORDS"

- What was the content of the prime? For example, a prime could be NURSE or WARM, which are semantically different
- **Internal Trait, e.g., Kind**
- **External Trait, e.g., Beard**
- **Role, e.g., Mechanic**
- **Object, e.g., Truck**
- **Pronoun, e.g., She**
- **Category, e.g., Male**
- **Mixture**
- **Other**

K. "PRIME CONTENT-PICTURES"

- What is the content of the prime?
- **Face**
- **Full Body/Action, e.g., the picture shows a female doctor, or a female with a child**
- **Object**
- **Other**

L. "PRIME CONTENT-OTHER"

- If other above, state the kind of primes used in the text box.

M. "PRIME DURATION"

- Enter the time (ms) that the prime stimulus was presented.
- If multiple prime durations were used, enter what specific durations were used separated by commas

N. "PRIME SET SIZE"

- The number of prime stimuli used
- *Note*: Prime set size is independent from the trials in the study. This variable merely asks the total number of designated prime stimuli
- If researchers draws a certain number of primes randomly from an overall set, enter the number of the stimuli sampled (e.g., 20 primes were randomly drawn from a pool of 100 for each participant; we want to enter the 20 here so the repetition variables are computed properly)

O. "PRIME CATEGORY SIZE"

- Number of unique "categories" that primes might reflect. This number should be equal to or less than "PRIME SET SIZE." For example, in gender task, there might be two categories of primes—males and females

P. Supplemental priming? (Yes/No)

- Do the participants perform or complete any extra tasks or procedures that might boost the strength of priming in the study?
- For example, categorize a set of names or pictures as Black/White, male/female before completing the experimental trials; keep a mental tally of the number of Black/White, male/female pictures or names presented at the same time as performing the experimental task.

Q. "Target Stereotype?"

- Does the target set contain the stereotypic information?
- For instance, a target could be gender or occupation. In gender priming, the occupation is the stereotype.
- **No, Yes**.

R. "TARGET TYPE"

- Were the targets pictures, words, or both?
- **Pictures**
- **Words**
- **Both**

S. "TARGET CONTENT–WORDS"

- What was the content of the target? For example, a target could be NURSE or WARM, which are semantically different
- **Internal Trait, e.g., Kind**
- **External Trait, e.g., Beard**
- **Role, e.g., Hairdresser**
- **Object, e.g., Truck**
- **Pronoun, e.g., She**
- **Category, e.g., Male**
- **Mixture**
- **Other**

T. "TARGET CONTENT–PICTURES"

- What is the content of the target?
- **Face**
- **Full Body/Action, e.g., the picture shows a female doctor, or a female with a child**
- **Object**
- **Other**

U. "TARGET CONTENT–OTHER"

- If other above, state the kind of targets used in the text box.

V. "Target Offset Response"

- Does the target disappear when response is made?
- No, Yes

W. "TARGET DURATION"

- Enter the time (ms) that the target stimulus was presented.
- If multiple target durations, enter all separated by commas
- If response ends target, enter the maximum duration

X. "TARGET SET SIZE"

- The number of target stimuli used
- *Note*: Target set size is independent from the trials in the study. This variable merely asks the total number of designated target stimuli
- If a researcher draws a certain number of targets randomly from an overall set, enter the number of the stimuli sampled (e.g., 20 targets were randomly drawn from a pool of 100 for each participant; we want to enter the 20 here so the repetition variables are computed properly)

Y. "TARGET CATEGORY SIZE"

- Number of unique "categories" that targets might reflect. This number should equal two or less than "TARGET SET SIZE." For example, in weapons task, there might be two categories—gun and tool

Z. "STEREOTYPE SET SIZE"

- Number of stereotype items per stereotype condition—note that stereotype information might be in target or prime
- Example—if the study uses 5 male and 5 female items, enter "5"; most studies should be symmetrical like the example, but if a study uses an unequal number, compute the average (e.g., 4 male and 8 female items—average is 6)

AA. "SOA" (stimulus onset asynchrony) 1-3

- Enter the time (ms) of the SOA.
- If multiple SOA durations were used, enter separated by commas

BB. "TRIALS-EXPERIMENTAL"

- Number of total experimental trials in experiment
- Not including practice trials
- Not including "neutral" trials that are irrelevant to stereotype comparison/analyses
- Not including trials with nonwords as targets (for LDT-like tasks)
- Not including trials without either prime or target

CC. "TRIALS-TOTAL"

- Number of total trials that participants experienced are—practice, "nonword" trials in LDT, and so on.

DD. "BLOCKS"

- Total number of blocks used in the experiment
- Note: We assume there was at least one block when information pertaining to the blocks was missing.

EE. "TASK" (*Note*: It's important to read up on how these tasks are defined below)

- What sort of task was carried out?
- **Stereotype classification**
- **Semantic classification task**
- **Pronunciation (or naming) task**
- **Lexical decision task (LDT)**
- **Weapons Task**
- **Other**

FF. "Judgment"

- Enter specific judgment that people made, this is especially important for semantic and stereotype
- "Word-nonword"—e.g., LDT
- "Weapon-Tool—e.g., Weapons
- "Black-White," "Male-Female," etc.
- **Use "Link" if they have to make some judgment about target that requires prime—even if judgment is "yes-no" or "related-not"**
- Pronunciation tasks will be "−88"

GG. "TASK FAST RESPONSE"

- Does the task encourage/require a fast response by limiting RT (<1000) and/or providing feedback when slow responses occur
- Yes
- No

HH. Irrelevant S-R

- Is there something about the prime that might activate a response to the target? (see De Houwer, 2003—for description)

- Yes: When analyzed trials require a different **systematic** response for certain types of stimuli—e.g., in stereotype categorization tasks people may respond “male” or “female” to targets and all are included in the analyses; same is true for Weapon’s task—“weapon” vs. “tool”
- No: When people always make same response to analyzed trials (e.g., always say “word” or “yes” in LDT) or always have unique response to target such as pronouncing word in naming
- If Uncertain, use −77 so all can look.

II. Relevant S-R

- Is there something about the prime or target that is directly relevant to the response to the target (see De Houwer, 2003, for description)
- Yes: If there is something about the prime or target that “points” to a response hand/button—e.g., and arrow or symbol like “<” that points to left button/hand
- No: This will most likely be your code

Information From the Results section

JJ. “Within Conditions”

- How many unique within-subject conditions were examined? This variable might be used to see how many trials per condition

KK. Dependent Variable (RT or Error rates)

- Did the researchers measure and analyze response times or error rates?

LL. Analysis

- How were the priming effects analyzed?
- **T-test**
- **Main Effect** (use this if ANOVA has congruency as one factor—others might be target or prime)
- **Prime-Target** (use this if ANOVA has prime and target factors—e.g., prime gender by target gender
- Other

MM. F-value

- Enter the reported *F* value
- If the analysis is only reported as nonsignificant, enter 0 for the *F* value

NN. T-value

- Enter the reported *t* value
- If the analysis is only reported as nonsignificant, enter 0 for the *t* value

OO. DF

- Enter the reported degrees of freedom

PP. Reported effect size

- If an effect size (Cohen’s d) is reported, enter the reported effect size
- If an effect size other than Cohen’s d is reported, enter the effect size and the kind of effect size

QQ. Mean incongruent

- If reported, enter the mean of the incongruent trials
- If means for separate types of incongruent trials are reported, enter each mean into subsequent boxes

RR. SD incongruent

- If reported, enter the standard deviation of the incongruent trials
- If means for separate types of incongruent trials are reported, enter each standard deviation into subsequent boxes

SS. Mean congruent

- If reported, enter the mean of the congruent trials
- If means for separate types of congruent trials are reported, enter each mean into subsequent boxes

TT. SD congruent

- If reported, enter the standard deviation of the congruent trials
- If standard deviations for separate types of congruent trials are reported, enter each mean into subsequent boxes

UU. P-value

- Enter the *p* value for the analysis
- If the analysis is only reported as nonsignificant, enter 1 for the *p* value.

Glossary

a. Block: a segment of trials intended primarily to spread out the experiment to allow rest but also used for experimental manipulations

b. Lexical decision task (LDT): task in which participants determine whether a word is a word or nonword. If they make this determination aloud, it is still considered an LDT, not a pronunciation task. For the purposes of this meta-analysis, this classification will include primed tasks.

c. Pronunciation (or naming) task: a task in which the participant says what the target is without making any sort of assessment. For example, if the participant were shown the word “delightful” they would just say the word aloud. *NOTE*: If participants are saying whether the target is male or female, this is a stereotype classification task, not a pronunciation task.

d. **Prime:** the first stimulus presented in the stimulus pair. For example, in the pair "Spock"–"delightful," Spock is considered the prime stimulus. See also Figure 1.

e. **Reaction time (RT) or response latencies**: the time in milliseconds it takes a participant to respond after seeing the target stimulus

f. **Semantic categorization task**: a task in which the participant makes a semantic classification of the target stimulus. For example, person vs. animal, alive vs. inanimate, etc. *Note*: This is different than classifications made based on the stereotype of interest, e.g., male–female.

g. **SOA (stimulus onset asynchrony)**: the time from the onset of the prime stimulus to the onset of the target stimulus. See Figure 1.

h. **Stereotype classification task:** a task in which the participant makes a stereotype classification of the target stimulus. For example, Black–White, male–female. These are usually based on the stereotype of interest.

i. **Stimulus-Response compatibility:** a task where the congruity of the trials is defined as a congruity between the prime and the response to the target; e.g., stereotype classification, semantic classification, and weapons' tasks fit within this category.

j. **Stimulus-Stimulus compatibility:** a task where the congruity of the trials is defined as a congruity solely between the prime and the target; naming/pronunciation tasks and LDTs fit within this category.

k. **Target**: the second stimulus presented in the stimulus pair. For example, in the pair "Spock"–"delightful," delightful is considered the target stimulus. See also Figure 1.

l. **Task**: the type of response the participant made to the target stimulus

m. **Trial**: the presentation of a fixation cross ("+") all the way to an intertrial interval (ITI) represents one trial. See Figure 1 (this is the sequence of one trial)

n. **Weapons Task:** a priming task examining racial stereotypes where the targets are tools and weapons.

o. **Within- vs. between-subjects designs**: in a between-subjects design, the participants are randomly assigned to a condition (e.g., 12 participants assigned to 300 SOA & 12 to the 1,000 SOA). In a within-subjects design, all 24 participants would be assigned to both the 300 and the 1,000 SOA conditions.

p. **Word attribute task (an "other" task)**: a task in which participants assess a characteristic of the target word.

a. For example, was the target word a "noun" vs. "adjective"

b. For example, was the target word "capitalized" vs. "lower case"

Authors' Note

Ciara K. Kidder is now at Marian University, Fond du Lac, Wisconsin. Katherine R. White is now at Kennesaw State University, Kennesaw, Georgia.

Acknowledgment

The authors would like to acknowledge Dirk Wentura, Adriaan Spruyt, and an anoymous reviewer for their invaluable feedback during the review process. The authors would also like to acknowledge David R. Herring for his guidance during this research.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Wentura and Degner (2010) assert that participant task is an important consideration for both evaluative and stereotype priming, but direct evidence of its importance for stereotype priming has not been directly demonstrated.
2. Although we did not specifically exclude studies with negative stimulus onset asynchronies (SOAs), backward priming, the authors were unable to find studies of sequential stereotype priming that used them.
3. A "boosting" method is one where the salience of the stereotype dimension or the strength of the association between stimuli and the stereotype dimension of interest is increased. For example, Wittenbrink, Judd, and Park (1997) asked participants to categorize names as Black or White prior to completing a race sequential priming paradigm with an lexical decision task (LDT). A "reduction" method is any number of interventions or techniques introduced to reduce the stereotype priming effect. For example, Sassenberg and Moskowitz (2005) activate a "creative" mind-set in some participants to examine whether this mind-set reduces or eliminates the stereotype priming effect.
4. During the submission and revision process, we did additional searches and/or received unpublished data to include studies that were published after our initial search (k = 28). The numbers for these searches are not included here because the process was somewhat different.
5. If exact F or t values for main effects or interactions of interest were not reported in the manuscript, the authors were contacted in an attempt to get those values.
6. Although d_z can also be directly calculated from means and standard deviations if the exact F is not reported, the correlation between the congruent and incongruent trials must also be known. One study was excluded because the necessary data were not reported and could not be obtained.
7. Many of these interactions are not included in this manuscript, but can be found in the supplemental materials published on the Open Science Framework website.
8. These analyses can be found in the supplemental materials.

9. Analysis can be found in supplemental materials.
10. One effect size coded as "long" SOA came from a study that employed masking after the prime. Analyses were conducted without this effect size, and the pattern of results was the same, so we reported the results with the effect size included.

References

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

*Amodio, D. M., Harmon-Jones, E., Devine, P. G., Curtin, J. J., Hartley, S. L., & Covert, A. E. (2004). Neural signals for the detection of unintentional race bias. *Psychological Science*, *15*, 88-93. doi:10.1111/j.0963-7214.2004.01502003.x

*Anderson, S. P. (2011). *The effects of global and local processing on gender stereotype activation and inhibition* (Doctoral dissertation, University of Minnesota). Available from ProQuest Dissertation and Thesis database. (UMI No. 3474716)

*Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, *7*, 136-141. doi:10.1111/j.1467-9280.1996.tb00346.x

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. J. Wyer & T. K. Srull (Eds.), *Handbook of social cognition: Basic processes; Applications* (Vol. 1-2, 2nd ed., pp. 1-40). Hillsdale, NJ: Lawrence Erlbaum.

*Bessenoff, G. R., & Sherman, J. W. (2000). Automatic and controlled components of prejudice toward fat people: Evaluation versus stereotype activation. *Social Cognition*, *18*, 329-353. doi:10.1521/soco.2000.18.4.329

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, *6*, 242-261. doi:10.1207/S15327957PSPR0603_8

*Blair, I. V., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology*, *70*, 1142-1163. doi:10.1037/0022-3514.70.6.1142

*Cacciari, C., & Padovani, R. (2007). Further evidence of gender stereotype priming in language: Semantic facilitation and inhibition in Italian role nouns. *Applied Psycholinguistics*, *28*, 277-293. doi:10.1017/S0142716407070142

Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review*, *16*, 330-350. doi:10.1177/1088868312440047

*Castelli, L., Macrae, C. N., Zogmaister, C., & Arcuri, L. (2004). A tale of two primes: Contextual limits on stereotype activation. *Social Cognition*, *22*, 233-247. doi:10.1521/soco.22.2.233.35462

Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. New York, NY: The Guilford Press.

*Chen, M. (1999). *Racial category activation as a precursor to nonconscious interpersonal motivation: A reconceptualization of stereotype priming effects* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9945264)

Clow, K. A., & Esses, V. M. (2007). Expectancy effects in social stereotyping: Automatic and controlled processing in the Neely paradigm. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, *39*, 161-173. doi:10.1037/cjbs20070013

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407. doi:10.1037/0033-295X.82.6.407

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology*, *89*, 469-487. doi:10.1037/0022-3514.89.4.469

Correl, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, *83*, 1314-1329. doi:10.1037/0022-3514.83.6.1314

*Cothran, D. L. A. (2005). *The role of facial affect in the perception of threat posed by black faces* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3191280)

De Houwer, J. (2003). A structural analysis of indirect measures of attitudes. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 219-244). Mahwah, NJ: Lawrence Erlbaum.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, *135*, 347-368.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5-18. doi:10.1037/0022-3514.56.1.5

*Dickter, C. L. (2006). *The effects of contextual information in processing race: Exploring the nature of the black as fear stereotype* (Doctoral dissertation). Available from ProQuest Dissertation and Thesis database.

*Dickter, C. L., & Kittel, J. A. (2012). The effect of stereotypical primes on the neural processing of racially ambiguous faces. *Social Neuroscience*, *7*, 622-631. doi:10.1080/17470919.2012.690345

Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology*, *22*, 22-37. doi:10.1016/0022-1031(86)90039-9

Duval, S. (2005). The trim and fill method. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 127-144). Chichester, UK: John Wiley.

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

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*(1), 297-327. doi:10.1146/annurev.psych.54.101601.145225

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*, 229-318. doi:10.1037/0022-3514.50.2.229

Fox, J., & Weisberg, S. (2011). *An {R} companion to applied regression* (2nd ed.). Thousand Oaks, CA: SAGE.

Gawronski, B. (2009). Ten frequently asked questions about implicit measures and their frequently supposed, but not entirely correct answers. *Canadian Psychology/Psychologie Canadienne*, *50*, 141-150. doi:10.1037/a0013848

Gawronski, B., & Bodenhausen, G. V. (2005). Accessibility effects on implicit social cognition: The role of knowledge activation and retrieval experiences. *Journal of Personality and Social Psychology*, *89*, 672-685. doi:10.1037/0022-3514.89.5.672

*Gawronski, B., Deutsch, R., Mbirkou, S., Seibt, B., & Strack, F. (2008). When "just say no" is not enough: Affirmation versus negation training and the reduction of automatic stereotype activation. *Journal of Experimental Social Psychology*, *44*, 370-377. doi:10.1016/j.jesp.2006.12.004

Glaser, W. R. (1992). Picture naming. *Cognition*, *42*, 61-105. doi:10.1016/0010-0277(92)90040-O

*Govorun, O., & Payne, B. K. (2006). Ego—Depletion and prejudice: Separating automatic and controlled components. *Social Cognition*, *24*, 111-136. doi:10.1521/soco.2006.24.2.111

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4-27. doi:10.1037/0033-295X.102.1.4

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464-1480. doi:10.1037/0022-3514.74.6.1464

*Greenwald, A. G., & Nosek, B. A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift Für Experimentelle Psychologie*, *48*(2), 85-93. doi:10.1026//0949-3946.48.2.85

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

*Harrison, L. A. (2001). *The effects of state self-esteem and individual prejudice level on explicit and implicit stereotyping* (Doctoral Dissertation). Retrieved from http://digitalcommons.unl.edu/dissertations/AAI3009722/

Herring, D. R., White, K. R., Jabeen, L. N., Hinojos, M., Terrazas, G., Reyes, S. M., . . . Crites, S. L. (2013). On the automatic activation of attitudes: A quarter century of evaluative priming research. *Psychological Bulletin*, *139*, 1062-1089. doi:10.1037/a0031309

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*, 1369-1385. doi:10.1177/0146167205275613

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, *85*, 616-626. doi:10.1037/0022-3514.85.4.616

Ito, T. A., & Urland, G. R. (2005). The influence of processing objectives on the perception of faces: An ERP study of race and gender perception. *Cognitive, Affective, & Behavioral Neuroscience*, *5*, 21-36. doi:10.3758/CABN.5.1.21

*Jones, C. R., & Fazio, R. H. (2010). Person categorization and automatic racial stereotyping effects on weapon identification. *Personality and Social Psychology Bulletin*, *36*, 1073-1085. doi:10.1177/0146167210375817

*Karylowski, J. J., Motes, M. A., Wallace, H. M., Harckom, H. A., Hewlett, E. M., Maclean, S. L., . . . Vaswani, C. L. (2001). Spontaneous gender-stereotypical categorization of trait labels and job labels. *Current Research in Social Psychology*, *6*, 77-90. Retrieved from http://www.uiowa.edu/crisp/volume-6-issue-6-march-19-2001

Kawakami, K., Dion, K. L., & Dovidio, J. F. (1998). Racial prejudice and stereotype activation. *Personality and Social Psychology Bulletin*, *24*, 407-416. doi:10.1177/0146167298244007

*Kawakami, K., Dion, K. L., & Dovidio, J. F. (1999). Implicit stereotyping and prejudice and the primed Stroop task. *Swiss Journal of Psychology/Schweizerische Zeitschrift Für Psychologie/Revue Suisse de Psychologie*, *58*, 241. doi:10.1024//1421-0185.58.4.241

*Kawakami, K., & Dovidio, J. F. (2001). The reliability of implicit stereotyping. *Personality and Social Psychology Bulletin*, *27*, 212-225. doi:10.1177/0146167201272007

*Kimura, A., Wada, Y., Goto, S., Tsuzuki, D., Cai, D., Oka, T., . . . Dan, I. (2009). Implicit gender-based food stereotypes. Semantic priming experiments on young Japanese. *Appetite*, *52*, 521-524. doi:10.1016/j.appet.2008.11.002

Klauer, K. C., Dittrich, K., Scholtes, C., & Voss, A. (2015). The invariance assumption in process-dissociation models: An evaluation across three domains. *Journal of Experimental Psychology: General*, *144*, 198-221. doi:10.1037/xge0000044

Klauer, K. C., Musch, J., & Eder, A. B. (2005). Priming of semantic categorizations: Late and response related, or earlier and more central? *Psychonomic Bulletin & Review*, *12*, 897-903. doi:10.3758/BF03196783

Klauer, K. C., Teige-Mocigemba, S., & Spruyt, A. (2009). Contrast effects in spontaneous evaluations: A psychophysical account. *Journal of Personality and Social Psychology*, *96*, 265-287. doi:10.1037/a0013248

*Kubota, J. T., & Ito, T. A. (2014). The role of expression and race in weapons identification. *Emotion*, *14*, 1115-1124. doi:10.1037/a0038214

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta analysis* (Vol. 49). Thousand Oaks, CA: SAGE.

Lupker, S. J., & Pexman, P. M. (2010). Making things difficult in lexical decision: The impact of pseudohomophones and transposed-letter nonwords on frequency and semantic priming effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1267-1289.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*, 93-120. doi:10.1146/annurev.psych.51.1.93

*Macrae, C. N., & Cloutier, J. (2009). A matter of design: Priming context and person perception. *Journal of Experimental Social Psychology*, *45*, 1012-1015. doi:10.1016/j.jesp.2009.04.021

*Macrae, C. N., & Martin, D. (2007). A boy primed Sue: Feature-based processing and person construal. *European Journal of Social Psychology*, *37*, 793-805. doi:10.1002/ejsp.406

*Macrae, C. N., Mitchell, J. P., & Pendry, L. F. (2002). What's in a forename? Cue familiarity and stereotypical thinking. *Journal of Experimental Social Psychology*, *38*, 186-193. doi:10.1006/jesp.2001.1496

Magee, M. W. (2011). *The interpersonal foundations of anti-atheist prejudice* (Doctoral dissertation, The City University of New York). Retrieved from https://dedicated.pstnet.com/downloads/eprime/publications/Magee_Michael_Doctoral_Dissertation.pdf

Marsh, K. L., Johnson, B. T., & Scott-Sheldon, L. A. (2001). Heart versus reason in condom use: Implicit versus explicit attitudinal predictors of sexual behavior. *Zeitschrift Für Experimentelle Psychologie*, *48*, 161-175.

Martin, D., & Macrae, C. N. (2007). A face with a cue: Exploring the inevitability of person categorization. *European Journal of Social Psychology*, *37*, 806-816. doi:10.1002/ejsp.445

*Mellott, D. S. (2003). *Measuring implicit attitudes and stereotypes: Increasing internal consistency reveals the convergent validity of IAT and priming measures* (Doctoral dissertation, University of Washington). Available from ProQuest Dissertation and Thesis database. (UMI No. 3091043)

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227-234. doi:10.1037/h0031564

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*, 297-326. doi:10.1037/0033-2909.132.2.297

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-group designs. *Psychological Methods*, *7*, 105-125. doi:10.1037/1082-989X.7.1.105

*Müller, F., & Rothermund, K. (2014). What does it take to activate stereotypes? Simple primes don't seem enough: A replication of stereotype activation (Banaji & Hardin, 1996; Blair and Banaji, 1996). *Social Psychology*, *45*, 187-193. doi:10.1027/1864-9335/a000183

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*, 226-254.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264-336). Hillsdale, NJ: Lawrence Erlbaum.

Neely, J. H., & Keefe, D. E. (1989). Semantic context effects on visual word processing: A hybrid prospective-retrospective processing theory. In J. H. Bower (Ed.), *The Psychology of learning and motivation* (Vol. 24, pp. 207-248). San Diego, CA: Academic Press.

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition*, *19*, 625-666. doi:10.1521/soco.19.6.625.20886

Olson, M. a., & Fazio, R. H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science*, *14*, 636-639. doi:10.1046/j.0956-7976.2003.psci_1477.x

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*, 776-783. doi:10.1037/h0043424

*Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving weapons. *Journal of Personality and Social Psychology*, *81*, 181-192. doi:10.1037//0022-3514.81.2.181

*Payne, B. K. (2002). *Cognitive control in automatic attitude measurement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3083583)

*Payne, B. K. (2005). Conceptualizing control in social cognition: How executive functioning modulates the expression of automatic stereotyping. *Journal of Personality and Social Psychology*, *89*, 488-503. doi:10.1037/0022-3514.89.4.488

*Payne, B. K., Lambert, A. J., & Jacoby, L. L. (2002). Best laid plans: Effects of goals on accessibility bias and cognitive control in race-based misperceptions of weapons. *Journal of Experimental Social Psychology*, *38*, 384-396. doi:10.1016/S0022-1031(02)00006-9

*Plaza, M., Boiché, J., Brunel, L., & Ruchaud, F. (2016). Sport = male . . . but not all sports: Investigating the gender stereotypes of sport activities at the explicit and implicit levels. *Sex Roles*, *76*, 202-217. doi:10.1007/s11199-016-0650-x

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, *95*, 385-408.

Rice, S. A. (1926). "Stereotypes": A source of error in judging human character. *Journal of Personnel Research*, *5*, 267-276.

Rogers, T. T. & McClelland, J. L. (2014). Parallel distributed processing at 25: Further explorations in the microstructure of cognition. *Cognitive Science*, *6*, 1024-1077. doi: 10.1111/cogs.12148

Rothermund, K. (2017). [Sequential Stereotype Priming Studies]. Unpublished raw data.

Rumelhart, D. E., & McClelland, J. L., & CORPORATE PDP Research Group. (Eds.). (1986a). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Vol. 1. Foundations*. Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L., & CORPORATE PDP Research Group. (Eds.). (1986b). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Vol. 2. Foundations*. Cambridge, MA, USA: MIT Press.

Sassenberg, K. & Moskowitz, G.B. (2005). Do not stereotype, think different! Overcoming automatic stereotype activation by mindset priming. *Journal of Experimental Social Psychology*, 41 (5), 317-413. doi:10.1016/j.jesp.2004.10.002

Sherman, S. J., Presson, C., Chassin, L., Rose, J., & Koch, K. (2003). Implicit and explicit attitudes toward cigarette smoking: The effects of context and motivation. *Journal of Social and Clinical Psychology*, *22*, 13-39.

Shulman, H. G., & Davidson, T. C. (1977). Control properties of semantic coding in a lexical decision task. *Journal of Verbal Learning and Verbal Behavior*, *16*, 91-98.

Smith, E. R., & DeCoster, J. (2000). Dual process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*, 108-131.

Smith, E. R., & Zarate, M. A. (1990). Exemplar and prototype use in social categorization. *Social Cognition*, *8*, 243-262. doi:10.1521/soco.1990.8.3.243

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

Spruyt, A., De Houwer, J., Hermans, D., & Eelen, P. (2007). Affective priming of nonaffective semantic categorization responses. *Experimental Psychology*, *54*, 44-53. doi:10.1027/1618-3169.54.1.44

Spruyt, A., Hermans, D., Houwer, J. D., & Eelen, P. (2002). On the nature of the affective priming effect: Affective priming of naming responses. *Social Cognition*, *20*, 227-256. doi:10.1521/soco.20.3.227.21106

Sterne, J. A. C., Becker, B. J., & Egger, M. (2005). The funnel plot. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 73-98). Chichester, UK: John Wiley.

*Stewart, B. D., & Payne, B. K. (2008). Bringing automatic stereotyping under control: Implementation intentions as efficient means of thought control. *Personality and Social Psychology Bulletin*, *34*, 1332-1345. doi:10.1177/0146167208321269

Stewart, T. L., Latu, I. M., Kawakami, K., & Myers, A. C. (2010). Consider the situation: Reducing automatic stereotyping through Situational Attribution Training. *Journal of Experimental Social Psychology*, *46*, 221-225. doi:10.1016/j.jesp.2009.09.004

Stone, G. O., & Van Orden, G. C. (1992). Resolving empirical inconsistencies concerning priming, frequency, and nonword foils in lexical decision. *Language and Speech*, *35*, 295-324.

*Todd, A. R., Thiem, K. C., & Neel, R. (2016). Does seeing faces of young Black boys facilitate the identification of threatening stimuli? *Psychological Science*, *27*, 384-393. doi:10.1177/0956797615624492

Verhaeghen, P., Aikman, S. N., & Van Gulick, A. E. (2011). Prime and prejudice: Co-occurrence in the culture as a source of automatic stereotype priming: Prime and prejudice. *British Journal of Social Psychology*, *50*, 501-518. doi:10.1348/014466610X524254

Viechtbauer, W. (2010). Conducting meta-analysis in R with metafor package. *Journal of Statistical Software*, *36*, 1-48.

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: A practical introduction. *Experimental Psychology*, *60*, 385-402. doi:10.1027/1618.3169/a000218

*Weber, D. R. (2008). *A study of the automatic stereotypic activation of driving behavior prediction, based on vehicle type categories* (Master's thesis). Available from ProQuest Dissertations and Theses database. (UMI No. 1450615)

Wentura, D. (1999). Activation and inhibition of affective information: For negative priming in the evaluation tasks. *Cognition and Emotion*, *13*, 65-91.

Wentura, D. (2000). Dissociative affective and associative priming effects in the lexical decision task: Yes versus no responses to word targets reveal evaluative judgment tendencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 456-469. doi:10.1037/0278-7393.26.2.456

Wentura, D., & Brandtstädter, J. (2003). Age stereotypes in younger and older women: Analyses of accommodative shifts with a sentence-priming task. *Zeitschrift Für Experimentelle Psychologie*, *50*, 16-26. doi:10.1027//1618-3169.50.1.16

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 95-116). New York, NY: The Guilford Press.

Wentura, D., & Rothermund, K. (2014). Priming is not priming is not priming. *Social Cognition*, *32*, 47-67. doi:10.1521.soco.2014.32.supp.47

*White, K. R. (2009). *Attention allocation and the variability of the stereotype priming effect* (Master's thesis, University of Texas at El Paso). Available from DigitalCommons@UTEP. (AAI1473903)

*White, K. R., Danek, R., Herring, D. R., Taylor, J. H., & Crites, S. L. (2017). *Stereotype priming effects in sequential paradigms: A matter of design and dimension*. Manuscript Submitted for Publication.

Wilson, J. P., Hugenberg, K., & Rule, N. O. (2017). Racial bias in judgments of physical size and formidability: From size to threat. *Journal of Personality and Social Psychology*. Advance online publication. doi:10.1037.spsi0000092

*Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, *72*, 262-274. doi:10.1037/0022-3514.72.2.262

*Wittenbrink, B., Judd, C. M., & Park, B. (2001). Evaluative versus conceptual judgments in automatic stereotyping and prejudice. *Journal of Experimental Social Psychology*, *37*, 244-252. doi:10.1006/jesp.2000.1456

Yu, L., Yang, X., Lu, Z., & Yan, Z. (2014). Effects of subliminal affective priming on occupational gender stereotypes. *Social Behavior and Personality*, *42*, 145-154. doi:10.2224/sbp.2014.42.1.145