January 2010

Response Mapping in Evaluative Priming

Laura Scherer
Washington University in St. Louis

WASHINGTON UNIVERSITY

Department of Psychology

Dissertation Committee:
Alan J. Lambert, Chairperson
David A. Balota
Pascal Boyer
Sandy Hale
Larry Jacoby
Stephen Ristvedt

RESPONSE MAPPING IN EVALUATIVE PRIMING

by

Laura Danielle Scherer

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August, 2010

Saint Louis, Missouri

ABSTRACT OF THE DISSERTATION

Response Mapping in Evaluative Priming

by

Laura D. Scherer

Doctor of Philosophy in Psychology

Washington University in St. Louis, 2010

Professor Alan J. Lambert, Chairperson

Recently, Scherer and Lambert (2009) proposed a new model of priming, which they called the Response Mapping (RM) Model. That model assumes that under some circumstances, priming effects are the result of an unintentional tendency for participants to impose the target categorization task onto the primes (which they are supposed to ignore). In the present dissertation, the RM model is reviewed, and the implications and boundary conditions of the model are explored. In Experiments 1 and 2, it was predicted and found that response mapping processes can result in evaluative conditioning effects. That is, priming tasks do not always simply measure attitudes, but rather these tasks can additionally create *new* attitudes towards the prime stimuli. In Experiments 3 and 4, two boundary conditions of the RM model were tested. In those experiments, it was found that evaluative priming effects depend on participants' ability to perceive the primes as belonging to distinct categories (boundary condition 1), and that those distinct categories must have different evaluative connotations (boundary condition 2). Importantly, results showed that priming effects are significantly stronger when primes are easily categorizable, relative to when they are not, even when the evaluative strength of the primes is held constant. Implications for theory and research involving priming measures and implicit attitudes more generally are discussed.

ACKNOWLEDGEMENTS

This dissertation was supported in part by an American Psychological Association Dissertation Research Award, as well as a Washington University Dissertation Fellowship. The author would like to thank Dr. Alan Lambert, Dr. David Balota, and Dr. Larry Jacoby for their help and guidance in the completion of this work. This dissertation was greatly improved as a result of their insights and suggestions. The author would also like to thank the other members of the dissertation committee, as well as the friends and family whose love and support made this dissertation possible.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

INTRODUCTION

Attitudes and attitude formation have long been a central focus of research in social psychology (Fazio & Petty, 2008). For the majority of social psychology's history, attitudes have been explored primarily by using direct inquiries into peoples' evaluative judgments. For example, participants are typically asked questions such as "How do you feel toward the Group X", and their answers are recorded on a Likert scale. However, this approach has quite a few obvious flaws, arguably the most important of which is that direct self-report measures are subject to social desirability pressures (Crowe & Marlowe, 1960; Fazio, Jackson, Dunton & Williams, 1995). That is, a participant might be less willing to provide an accurate report of his or her attitudes if the truthful answer is a socially unacceptable one. Hence, one longstanding priority in the social psychological field is to develop measures that sidestep this social desirability concern and tap into attitudes that participants might not ordinarily be willing or able to articulate explicitly (Devine, 1989; Fazio, Sanbonmatsu, Powell & Kardes, 1986; Fazio, Jackson, Dunton & Williams, 1995; Greenwald, McGee & Schwartz, 1998; Nosek & Banaji, 2001; Payne, 2001; Payne, Cheng, Govorun & Stewart, 2005; Scherer & Lambert, 2009).

This longstanding desire to tap into people's true feelings (as opposed to their socially desirable feelings) can explain, in part, why the past twenty years of attitude research has been flooded with a wave of interest in indirect tests of attitudes, also known as “implicit attitudes” (Bargh, 1999; De Houwer, 2006; Fazio et al. 1995; Fazio & Olsen, 2003; Gawronski, LeBel & Peters, 2007; Greenwald et al., 1998; Hassin, Uleman & Bargh, 2005; Olsen & Fazio, 2003; Payne, 2001). Although there is currently some lively debate as to the exact meaning of the term “implicit” (De Houwer, 2006), for the purposes of this article the main differences between explicit and implicit measures is that the latter (a) do not involve direct queries about the attitude object; (b) offer much less opportunity for people to exercise control over their responses and (c) are typically more sensitive in detecting the presence of, as well as changes in, evaluative associations with the attitude object.

To date, many psychological tests have been developed for the purpose of assessing attitudes indirectly. Perhaps the most popular and widely-used of these tests belong to a broad category of tasks, known collectively as priming tasks (e.g. Abrams & Greenwald, 2000; Fazio et al. 1986; Fazio et al. 1995; Payne 2001; Payne et al. 2005; for an excellent review and analysis of these tasks see Klauer & Musch, 2003). In a typical priming task, participants are asked to view images that appear in rapid succession on a computer screen. For example, in Payne’s (2001) priming paradigm, participants are presented with a picture of either a Black or White face (the prime), and after a few hundred milliseconds this face is replaced by a picture of either a gun or a tool (the target). Participants are instructed to ignore the face, and respond only to the target by indicating (with a key press) whether that target is a tool or a gun. In this task (as well as in other similar priming tasks), participants are not able to ignore the primes, and in fact, the primes actually end up systematically biasing participants' responses to the targets. For example, the presence of a Black prime makes it more likely that the target will be identified as a gun, even if it is actually a tool. The White prime produces exactly the opposite effect, insofar as participants are more likely to identify the target as a tool if it is preceded by a White face, even if the target is actually a gun. This biased pattern of responding is assumed to reveal participants' underlying stereotypic associations with Black and White faces. The upshot of these considerations is that priming tasks, such as the one just described, can potentially tap into participants' "real" evaluations of the primes, without having to directly ask participants to report their attitudes. For the reasons discussed above, these sorts of tasks represent a potentially huge advancement in attitude measurement and theory.

In spite of their potential utility, implicit measures have been the target of much controversy (e.g. Cunningham, Preacher & Banaji, 2001; De Houwer, 2006; Olsen & Fazio, 2003; Gawronski et al., 2007). Concerns have been raised about the reliability and predictive validity of these measures, as well as how these measures are related (or not) to explicit reports (Arkes & Tetlock, 2004; Karpinski & Steinman, 2006; Karpinski & Hilton, 2001; Lambert, Payne, Ramsey & Shaffer, 2005; Olsen & Fazio, 2003; Olsen & Fazio 2004; Payne, Burkley &

Stokes, 2008; Wilson, Lindsey & Schooler, 2000; Wittenbrink, Judd & Park, 1997; Wittenbrink & Schwarz; 2007). Moreover, there have been perhaps a dozen review articles in the past ten years devoted to trying to articulate exactly what an "implicit attitude" is, and what it means to have such an attitude (e.g. De Houwer, 2006; Gawronski et al. 2007). Implicit attitudes have been variably understood as unconscious, uncontrollable, unintentional, or some combination of these factors (Bargh, 1994). To date, there is still much controversy concerning what these tasks are measuring, and how these tasks produce their effects.

The Present Research

Although implicit attitude measurement represents a significant theoretical advance in the field of social psychology, clearly more work is needed to help clarify exactly what these tasks are measuring. In the present article, I will attempt to shed some light on this issue, by examining a new model of priming, called the *Response Mapping Model*. The basic features of this model, along with five studies offering support for its assumptions, were recently reported by Scherer and Lambert (2009). However, as will become clearer in the sections to follow, some fundamental questions remain about this model, including some ambiguities as to the exact nature of the processes involved. Hence, the overriding goal of my dissertation is to further explore the response mapping framework above and beyond any of the research that has been done thus far. In the present research, I will systematically examine both the boundary conditions and implications of the response mapping model.

Brief Introduction to the Response Mapping Model

As I will explain in greater detail in a subsequent section, response mapping presents a new way of thinking about and understanding both priming tasks and automatic behavior in general. The response mapping model begins with the assumption that people use salient goals and motives as a cue for categorizing the stimuli in their environment (for a detailed discussion of goal-related categorization, see Barsalou, 1991). Hence, automatic behaviors toward any given object are context-dependent; that is, they depend on the particular response tendency that is salient at the time of judgment, as well as the particular array of stimuli that are being judged.

To illustrate some of the basic elements of response mapping, it is useful to imagine a starving person who must decide what objects in his environment are a potential source of nourishment, and what are not. Since this person is starving, he will have a goal to eat. This goal will cause him to categorize things in his environment as either edible or inedible, and this categorization process is associated with a relevant response, namely, to eat or not to eat (Barsalou, 1991).

However, the starving person's categorization of an object as either edible or inedible, and his subsequent response to that object, will depend on the array of things that are available in his environment. For example, a person faced with a choice between an insect and a rock will probably attempt to eat the insect. However, that same person, when given a choice between the insect and an unfamiliar fruit, may view the insect as inedible. Note that the starving person may not actually think that an insect is a particularly good example of food, no matter what the circumstances, but his behavior toward that insect will nonetheless reflect the fact that he has categorized it as either "something to eat" or "something that I probably shouldn't eat". Also note that if this person was not starving, then there would probably be no attempt to classify *any* of these objects as food. In fact, from the perspective of a well-fed person, all of these objects (the rock, insect and strange fruit) might be viewed as inedible. The upshot of these considerations is that when faced with a strong motivation to categorize objects along a given dimension (i.e. food or non-food), people will attempt to fit their environment into those dimensions, even if this leads to unlikely or unusual categorization judgments (Barsalou, 1983; Barsalou, 1991). In sum, a person's response to any given object is guided by an interaction between a) the goal that is activated at the time of judgment, and b) the context in which that object is judged. These are the two basic tenets of response mapping.

Overview of the Present Dissertation

In the following pages, I will argue that both explicit and implicit judgments are guided by the aforementioned principles of response mapping. Although this paper will focus on models of priming, the overall goal is to demonstrate a novel account of automatic human behavior. In Part I of the following introduction, I will review two dominant models that have been used to explain priming effects, namely *spreading activation* and *response compatibility*. In Part II, I will describe the Response Mapping (RM) Model of priming, which was recently proposed by Scherer and Lambert (2009). In Part III, I will briefly describe a few experiments that support the RM model. In Part IV, I will discuss some important implications of the RM model, and how these implications provide the impetus for the present research. Finally, in Part V, I will present four experiments that constitute the present research. These experiments lend crucial insights into the RM model. Hence, the overall purpose of this dissertation is to explore the various implications of the response mapping framework, in order to gain a greater understanding of priming tasks and automatic processes that underlie them.

PART I:

Spreading Activation and Response Compatibility in Evaluative Priming

On the surface, priming effects seem to result from a fairly simple set of processes: The prime activates an attitude, and that attitude influences participants' subsequent response to the target. However, such a simplistic explanation leaves many questions unanswered, the most important of which is *how* primes exert their influence over participants' responses. In the following pages, two dominant models are considered: spreading activation and response compatibility. It is worth noting that both of these models have ardent supporters, and no research has been successful in developing a critical test of the viability of the two models (Klauer & Musch, 2003).

Spreading Activation Models of Priming

According to the spreading activation model of priming, memories are organized as a web of associated nodes. Each node represents a concept, such as “robin,” “sparrow” or “dog”, and the nodes are organized so that related concepts are located closer together on the web than unrelated concepts (Anderson & Pirolli, 1984; Balota & Lorch, 1986; Collins & Loftus, 1975; Neely, 1991; Neely, 1977; for a more recent review of related models, see Ratcliff et al. 2004). For example, the concepts “robin” and “sparrow” are closer together in this conceptual network than “robin” and “chair”. When a particular node is activated, it pre-activates closely related nodes. Therefore, uttering the word “robin” will pre-activate all sorts of related concepts, such as “sparrow,” “bird” and “fly”, and as a result, these secondary concepts will come to mind more easily than would other, unrelated concepts.

Spreading activation models were originally designed to explain findings from lexical decisions tasks (LDTs). A LDT is similar to the priming tasks described earlier, except that both the primes and targets are typically words, and participants are instructed to ignore the primes and identify targets as either real words or non-word letter strings (e.g. “kumph,” “lapgh”). Moreover, the primes and targets are paired so that they are either semantically related to each other (e.g. robin-bird) or unrelated (e.g. robin-chair). Typically, researchers find that responses are facilitated when the prime and target are related, relative to when they are not (Neely, 1991).

The spreading activation model explains LDT effects by assuming that the prime pre-activates related target words, making these related targets easier to identify than unrelated targets (c.f. Neely, 1991; for a more recent and sophisticated model, see Ratcliff et al. 2005). For example, when participants are primed with the word “bird”, this pre-activates all of the concepts that are closely associated with birds in memory. If the following target word is one of these related concepts, like “robin”, then the target will be easy to identify as a word because it is already partially activated. If the following target word is an unrelated concept (like “chair”), however, then participants will have to do a more detailed memory search for this item before it can be identified as a word. The result is that related target words are identified more quickly than unrelated target words.

A Spreading Activation View of Evaluative Priming

Spreading activation models were designed to account for semantic priming effects (that is, priming from “sparrow” to “robin”). According to some researchers, spreading activation can explain evaluative priming effects as well (Fazio, Sanbonmatsu, Powell & Kardes, 1986; Fazio, Jackson, Dunton &

Williams, 1995; Hill & Kemp-Wheeler, 1989; Bower & Cohen, 1982). According to Fazio's *attitude accessibility hypothesis*,

> "Presentation of an attitude object would automatically activate any strong association to that object. Such activation is assumed to spread along the paths of the memory network, including any evaluative associations. Consequently, the activation levels of associated evaluations are temporarily increased. If a target word that corresponds in valence to one of these previously activated evaluations is subsequently presented for judgment, then less additional activation is required…for a judgment to be made" (Fazio et al., 1986, p. 231).

On the surface, it might seem perfectly fitting to extend spreading activation from semantic priming to evaluative priming. Yet a spreading activation view of evaluative priming actually requires a significant extension of the original model. According to the original spreading activation model, primes activate a conceptual node, which pre-activates other related concepts. But according to a spreading activation model of evaluative priming, primes activate a global *evaluative* node, which pre-activates all other concepts that share a similar evaluative connotation (Bower, 1981; Bower 1987; Bower 1991; Fazio et al. 1986). For example, a cockroach is thought to activate a "negative node", which in turn pre-activates a broad array of other negative concepts. These pre-activated negative concepts need not be conceptually related to cockroaches. That is, a picture of a cockroach could pre-activate such broad-ranging concepts as "pain," "shame" and "death". According to the spreading activation account of evaluative priming, all of these words should receive equal activation as a result of viewing the cockroach picture, even though cockroaches usually don't cause shame or pain. This is an important extension of semantic priming, because it proposes that there is a generalized evaluative node that exists in addition to specific semantic nodes, and this evaluative node can produce priming in the absence of semantic relatedness. In fact, some researchers have proposed that the concept of a generalized evaluative node extends the spreading activation model too far (for a detailed discussion of this issue, see Klauer & Musch, 2003). As a result, an alternative *response compatibility model* has been proposed as another way of accounting for evaluative priming effects.

Response Compatibility Models of Priming

Response compatibility models were originally developed to explain effects from the Stroop task (Stroop, 1935; MacLeod, 1991). In the Stroop task, participants are presented with color words (e.g. "blue", "yellow") that are written in various colors of ink. Participants' task is to identify the color of the ink, while ignoring the written word. When the word and ink colors are the same, this task is fairly easy. However, the task becomes quite difficult when the word and ink colors are different. For example, participants are much slower to correctly respond to the word "blue" written in red ink than the word "blue" written in blue ink.

These effects are thought to occur as a result of interference between the response activated by the irrelevant stimulus (in this case, the word), and the response activated by the relevant stimulus (in this case, the ink color). When the two respective responses are compatible, then responding is facilitated. When they are incompatible, responding is inhibited. Importantly, the response compatibility model assumes that the conflict or compatibility between the responses is the driving force behind priming effects (Gawronski, Deutsch, LeBel & Peters, 2008; for a review, see Musch & Klauer, 2003; also see Flanker effects, Erikson & Erikson 1974). If the responses activated by the irrelevant and relevant stimuli do not produce any interference (for example, if the word "truck" is written in blue) then no priming effects should emerge.

A Response Compatibility View of Evaluative Priming

Extending the above logic to evaluative priming is relatively straightforward. In an evaluative priming task, the irrelevant stimulus is the prime, and the relevant stimulus is the target word. Participants' task is to identify the target words as either good or bad, and hence, these targets obviously activate either a good or bad response. However, even though participants are told to ignore the primes, they cannot, and the primes similarly end up activating either a good or bad response (Gawronski et al., 2008; see also Klauer & Musch, 2003). For example, a cockroach prime will activate the "bad" response, and if the following target word requires the "bad" response, then the correct response will be facilitated. If the following target word requires a “good” response, however, then the correct response will be inhibited. In this way, the “activation of the attitude associated with the [prime] suggests a response that either facilitates or inhibits the response to the following target word” (Ferguson & Bargh, 2003, p. 173).

One important prediction that is borne out of response compatibility models is that the nature of the priming effect should depend on the nature of the response that is required by the task. That is, response compatibility predicts that priming effects should be task dependent, and to some extent this has been found to be the case (De Houwer et al. 2000; Klauer & Musch, 2005; Klinger et al., 2000; Rothermund & Wentura, 1998). For example, priming effects that are evaluative in nature should occur only when the responses required are also evaluative (e.g. negative and positive word judgments). To illustrate, suppose that on a given trial, the prime is a cockroach and the word is “horrible”. If the task is to identify this word as negative or positive, then the negative response is activated by both the prime and target, causing facilitation. On the other hand, if the task is to identify the targets as either nouns or adjectives, then the cockroach should activate the “noun” response whereas the target word (horrible) should activate the “adjective” response. In this instance, therefore, the cockroach should not facilitate responses to the word “horrible”; in fact, the cockroach should inhibit responding to it, because the prime is a noun whereas the target is an adjective. These sorts of predictions have been verified empirically (Holender, 1992; Klauer & Musch, 2003). While spreading activation models can also account for such findings, it requires taking on a number of somewhat awkward additional assumptions.[1] Hence, these sorts of findings are considered good evidence for the response compatibility view.

PART II:

The Response Mapping Model

The RM model represents an entirely new model of priming that is distinct from both the spreading activation and response compatibility views. Although the RM model assumes that primes activate responses, the RM model goes considerably beyond response compatibility to argue that primes can sometimes activate responses for reasons apart from their semantic or evaluative meaning. That is, there might be some situations in which an ordinary household object could activate a positive response, or a negative response, even though that object has no positive or negative connotations whatsoever. Essentially, the RM model postulates a mechanism by which primes can activate responses that have no a priori relation to the primes themselves (I will explain this mechanism in full detail, below). For example, according to this model, a picture of a towel can, under the right circumstances, activate a “flower” response or a “food” response, even though towels have no characteristics of either flowers or food. To my knowledge, no other model of priming makes this claim.

Response Mapping in Explicit Judgments

Response mapping can occur in both explicit and implicit judgmental domains. For ease of exposition, I will first explain response mapping in explicit judgments. Then I will explain how this same concept can be applied to priming tasks.

Suppose for a moment that participants are asked to explicitly judge the size of a mouse using only one of two response options: "small" and "large". Further suppose that participants have already used the "small" response to refer to an extremely small target, such as a single celled organism. Under normal circumstances, participants would probably refer to mice as small, but use of that response option in this particular rating context violates participants' motivation to successfully convey that they recognize that mice are bigger than amoebas. Hence, the mouse is assigned the "large" rating by default, because the other response option—"small"—has already been reserved for the other stimuli being considered in that task. When pressed on the matter, we suspect that most respondents would admit that they don't *really* think that mice are "large" (but see Kosslyn, 1975). After all, this is not the type of language people usually use in connection with these animals. However, respondents might respond (with perhaps some indignation) that they were essentially doing the best they could with the response options that were provided for them by the experimenter.

This deliberately simple example illustrates what turns out to be fairly complex set of issues, which have been explored in a number of different judgmental models, most notably range-frequency theory (Parducci, 1965). One of the features of this model is that it assumes that overt ratings can reflect a compromise between a frequency principle (a bias towards using the available response alternatives equally often) and a range principle (a motivation to accurately match the range of responses to the range of underlying stimuli). (See Wedell et al. 2007 for a related discussion). Range-frequency theory represents a powerful, but rather complex, model that is capable of addressing judgmental settings far more complex than the simple example used here and full appreciation of it requires understanding of some rather complex mathematical modeling assumptions.

Nevertheless, for the present purposes, the critical assumption of this and other "response-based" models of judgmental contrast (Biernat, 2005) is that human judgment often involves a basic process of *response mapping*. In this case "response mapping" is defined as the process by which people select a particular response (e.g. "good", "2", "pretty") and use it to refer to a given stimulus. As part of this process, the mapping of response options onto extreme stimuli can displace ratings of less extreme stimuli onto other sorts of responses, and this displacement essentially describes a contrast effect. In other words, mice can be *large*, and elephants can be *small*, provided that the category labels that are normally used to refer to these stimuli have already been assigned to some other, even more extreme stimuli in that particular setting.

Response Mapping in Evaluative Priming Paradigms

The basic principles described above can, in principle, be applied to implicit judgment domains, including priming paradigms. Just as in the case of explicit judgments, priming paradigms can involve mapping of response options onto primes. Note, however, that there are a few important differences that characterize response mapping here as it might occur in priming paradigms.

In explicit rating paradigms, response mapping arises as part of participants' intentional goal, which is to respond to and rate the target stimuli. The situation in priming paradigms is somewhat more complex. Participants' responses to the *targets* are part of their primary, intentional goal (e.g. to classify words as positive and negative). However, participants are also responding to the *primes*, even though they are not supposed to be attending to these stimuli at all. This creates an interesting state of affairs in evaluative priming paradigms: While participants intentionally sort the targets into "good" and "bad" categories they are, at the same time, unintentionally sorting the primes into "good" and "bad" categories as well. That is, the response mapping framework proposes that participants *unintentionally impose the response categorization scheme onto the primes.* The primes become associated, or "mapped", with the responses as a result of this process (see Figure 1 for a graphic depiction of this process).

Under normal priming circumstances, the response mapping assumption is not particularly noteworthy, in the sense that researchers almost always present an exactly balanced ratio of positive and negative primes. In fact, we are aware of very few studies in the social psychological literature that do *not* follow this general rule. In other words, if participants end up mapping the positive primes with positive responses, and the negative primes with negative responses, this could be seen merely as a validation of the initial assumptions of the experimenter. For example, Figure 1a displays two types of primes that might be presented in a priming task. The picture of a shark activates a negative evaluation, and thereby elicits a negative response. Likewise, the picture of a butterfly activates a positive evaluation, and thereby elicits a positive response.

However, note what might happen when participants are presented with an array of priming stimuli consisting of extreme stimuli (e.g. unambiguously negative primes) along with an array of relatively neutral stimuli (Figure 1b). In Figure 1b, half of the primes are unambiguously negative, and half are neutral. According to the response mapping framework, participants will attempt to impose the good/bad classification scheme onto *all* of the primes, even though half of the primes are actually neutral. The neutral primes do not activate any particular evaluation, and therefore do not fit either the “good” or “bad” response labels perfectly. Nevertheless, there is a clear evaluative distinction between the neutral primes and the other unambiguously negative primes that are presented in the task. As shown in Figure 1b, the result is that the neutral primes will activate the “positive” response, even though those primes are not particularly positive. Just as a mouse may be mapped onto the “large” response in order to distinguish it from amoebas, a neutral prime may be mapped onto the “positive” response to distinguish it from extremely negative primes.

Hence, when the neutral primes are embedded among the negative primes, the unfavorable response option (“bad”) is mapped onto the negative primes, leaving the remaining (“good”) response option for the neutral stimuli. Conversely, when the neutral primes are presented along with the *positive* primes, the favorable response option (“good”) is mapped onto the positive primes, leaving the remaining (“bad”) response option for the neutral stimuli, by default (see Figure 1c). Consequently, the response mapping framework predicts that strong contrast effects will emerge for neutral primes when they are placed in a task along with either extremely positive or extremely negative primes.

These are precisely the sort of effects that Scherer and Lambert (2009) found in their research. In their experiments, Scherer and Lambert presented participants with evaluatively neutral prime pictures, and varied the context in which those neutral primes were presented. For half the participants, the evaluatively neutral primes were presented in a task along with extremely negative primes, so that half the primes were negative and half were neutral (see left side of Figure 2). For the other participants, the neutral primes were presented in a task along with extremely positive primes, again evenly divided between the two types of primes (see right side of Figure 2). (Figure 2 illustrates these two priming conditions in the form of an Attitude Misattribution Procedure (AMP) task, in which participants are instructed to identify Chinese characters as relatively pleasant or unpleasant, Payne et al. 2005).

It was expected that the extremely positive and negative stimuli would elicit strong priming effects, and this is exactly what the authors found. For example, the negative primes increased the probability of negative responses to the following target, and positive primes increased the probability of positive responses. However, of critical interest were participants' automatized reactions to the neutral prime stimuli. According to most priming theories, the neutral primes should have no effect at all, since these stimuli were purposefully selected for their lack of strong evaluative meaning. However, just as predicted by the response mapping framework, Scherer and Lambert found that the neutral stimuli influenced participants' responses in a way that was surprisingly similar to effects typically observed for unambiguous stimuli. When embedded among negative primes, neutral primes elicited priming effects analogous to those seen with positive primes. Conversely, when embedded among positive primes, neutral primes showed properties normally associated with negative primes. According to the response mapping framework, this contrast effect occurred because participants imposed the target categorization scheme—which included only positive and negative categories—onto the primes. In doing so, they mapped the neutral primes onto the positive or negative response by default, not because those responses reflected the actual evaluative connotations of the neutral primes themselves.

In order to understand the essential elements of the response mapping framework, it can be heuristically useful to regard these mappings as involving two different stages. (The term "stage" is used in a general sense here; the temporal ordering of these stages is not crucial for present purposes.) One stage of the response mapping process involves mapping the "good" and "bad" responses onto the *target* stimuli. This mapping is a direct consequence of the kinds of instructions that participants are typically given on priming tasks (e.g. "whenever you see a positive word, hit the good key"). The other stage of the response mapping process involves mapping responses onto the *primes*.

An obvious objection could be raised at this point. Why would participants do something so illogical as map stimuli onto a response that has no relation to the stimuli at all? There are two answers to that objection. First, keep in mind that the categorization processes occurring with the primes is happening *unintentionally* and, in all likelihood, outside of conscious awareness. Researchers have long recognized that automatic processing can involve the use of categories in ways that don't always follow strict rules of logic and rationality, and this could well be one of those examples. Second, even if participants were fully aware of how they are mapping responses onto the primes, this involves considerations not altogether different from those pertaining to explicit ratings. In the same way that people may not "really" think of elephants as particularly small, they may not "really" think of wooden stools as positive or negative.

Summary of the Response Mapping Framework

Although our framework may seem complex, the essential details are actually fairly straightforward, and can be summarized as follows.

1. Explicit and implicit attitude tasks involve a basic process of response mapping, which we simply define as the process of selecting any available response and using it to refer to a given stimulus.

2. In the case of explicit rating tasks, this process is part of an intentional effort to judge the target stimuli. In the case of implicit tasks, the response mapping process encompasses both the *intentional* goal to respond to the targets and well as the *unintentional* imposition of those response categories onto the primes.

3. The response mapping process is not necessarily "rational" and participants may not even be aware that such a process is even occurring, especially as it concerns the way that participants process information about the primes. Rather, response mapping can be better understood as an unintentional process of imposing the response categories onto the primes.

4. Relatively extreme and/or unambiguous stimuli will tend, by their very nature, to “dominate” the response mapping process. For example, if participants are presented with a randomized array of unambiguously negative and neutral stimuli, the former stimuli will tend to stand out and will be assigned the most negative response option available.

5. Contrast can be understood as a process of judgmental displacement as a result of the preceding step. In other words, after respondents have “used up” the available response options to refer to the unambiguous stimuli, less extreme or midrange stimuli will be assigned to whatever remaining response options are available. This explains both why recycling (a “mid-range” social issue) can be rated as either important or trivial depending on whether the available response options have been taken up by extremely important or extremely trivial issues (nuclear war, sidewalk spitting, respectively). This also explains why a towel could elicit automatic priming effects analogous to those associated with positive or negative stimuli, depending on the nature of the other primes presented in that context.

PART III:

Evidence for the Response Mapping Framework

Evidence for the response mapping framework was obtained by Scherer and Lambert (2009) using two different kinds of priming tasks: evaluative priming (Fazio et al, 1986) and the Attitude Misattribution Procedure (AMP; Payne et al. 2005). A detailed description of these tasks can be found in Appendix A. It is recommended that the unfamiliar reader review Appendix A before moving forward to the next section.

Evidence for Response Mapping Using Evaluative Priming and the AMP

Earlier I described an experiment in which Scherer and Lambert (2009) found that neutral primes elicit favorable priming effects when they are presented in a task along with negative primes, whereas they elicit unfavorable priming effects when the neutral primes are presented with positive primes. Although these contrasts effects provide initial evidence for the RM model, it is still possible that these effects were obtained because the neutral primes were actually perceived as being more or less favorable in the different task contexts. However, in two experiments Scherer and Lambert (2009) demonstrated that response mapping processes better accounted for these contrast effects. Those experiments are described below.

Testing the Response Mapping Framework: a 3-response paradigm

In their first test of the response mapping framework, Scherer and Lambert (2009) started with the following assumption: Response mapping processes should be contingent upon the kinds of response options that are available in the task (see point 5, above). According to the response mapping framework, contrast effects were observed for the neutral primes because those primes were displaced onto either a good or bad response. *Had a neutral response option been available, one would have expected the neutral primes to be mapped onto the neutral response, and not the negative or positive response.* Hence, if a neutral response option is included in the contrast paradigm (so that now there are three response options: good, bad and neutral), the neutral primes should facilitate the neutral response, and the contrast effect should disappear. On the other hand, the perceptual change view assumes that the neutral primes actually took on positive or negative connotations when presented alongside other, unambiguous primes. Hence, this latter view predicts that the neutral primes should not facilitate the neutral response, because in the task context they are no longer neutral.

In Scherer and Lambert's resultant experiment (2009, Experiment 3), participants performed one of two tasks, in which either positive and neutral, or negative and neutral primes were presented. This aspect of the experiment was exactly like the previous experiments that obtained evaluative contrast. However, in this experiment the targets included positive, negative *and neutral* words.

Participants were asked to identify each of the target words by indicating whether it was positive, negative or neutral. The results were exactly what were predicted by the response mapping framework. The neutral primes did not produce a contrast effect, but they did facilitate neutral responses in both of the task conditions. These results indicate that when participants are provided with response options that fit the evaluative connotations of the primes, this relieves the pressure to map the primes onto responses that don't share the same evaluative connotation. Hence, the contrast effects that were obtained in previous experiments were probably the result of providing participants with response options that could not account for the evaluative connotations of the primes in the task.

Testing the Response Mapping Framework: a semantic priming paradigm

In their second test of the response mapping framework, participants were presented with a priming task in which they were asked to identify targets as either food-related words (e.g. eat, hungry) or flower-related words (e.g. bloom, grow) (Scherer & Lambert 2009, Experiment 4). In one condition, the primes included pictures of flowers and household objects. In another condition the targets and judgments were identical, but the primes consisted of food and household objects. Hence, in each condition half the primes were related to one of the response categories (either flower or food pictures), and the other primes (household objects) were not related to either food or flowers.

In this task, the flower primes facilitated responses to flower words, and the food primes facilitated responses to food words. As for the object primes, all extant models of priming (i.e. both spreading activation and response-based models) predict that the object primes should not influence responses at all, because they are unrelated to both of the targets/responses. However, the response mapping framework uniquely predicted that the object primes would be mapped onto either the flower or food response, depending on the task context. When the object primes were presented along with flower primes, the objects should be mapped onto the food response, by default. In contrast, when objects were presented with food, the objects should be mapped onto the flower response, by default. That is, the RM model predicts that the object primes will be mapped onto whichever response is not already claimed by the other, unambiguous primes in the task. These are exactly the results that were obtained.

Hence, Scherer and Lambert found that the object primes acted like food primes when presented with flower primes, and acted like flower primes when presented with food primes. It is almost impossible to argue that the object primes somehow seemed more "food-like" or "flower-like" in the two conditions, and yet these primes produced effects that were very much like food and flower prime stimuli. As a result, this experiment demonstrates a situation in which systematic priming effects can occur in the absence of any logical relationship between the prime and the response. This experiment clearly shows that responses are mapped onto priming stimuli as a result of the prime context, and this process can occasionally result in priming effects that do no reflect the semantic or evaluative connotations of the primes.

PART IV:

Unresolved Issues in the RM Model

The aforementioned experiments validated the RM model. At this point, the question is not *whether* the RM model is valid, but rather *when* it is valid, and also what sorts of implications the model has for automatically activated evaluations. The RM model raises many novel and important questions for priming research and automatic evaluative processes. In the following section, I will identify several of these questions, and discuss how they can be addressed with further research. These questions and issues will lay the theoretical groundwork for the experiments to come.

Unresolved Issue #1: On the Fundamental Nature of Response Mapping Processes: Testing the Viability of the Central vs. Peripheral View

One longstanding issue surrounding response compatibility models is whether the locus of the affective priming effect is “central” or “peripheral” (for a relevant discussion, see Klauer & Musch, 2003). According to the peripheral view, primes simply activate response tendencies, such as “use your right hand” (Klinger et al., 2000; Musch, 2000). In contrast, the central view purports that affective priming involves a process of categorizing the primes as good or bad

(Abrams & Greenwald, 2000; Abrams, Klinger & Greenwald, 2003; Musch, 2000). As I will explain below, the same issue also arises for the RM model.

To illustrate the distinction between the central and peripheral explanations of affective priming, it is useful to consider a study reported by Abrams and Greenwald (2000). In this study, the researchers found that subliminal priming effects were stronger when participants had previously practiced classifying the primes as positive or negative. However, this previous practicing made it unclear whether participants simply learned to associate, say, a positive prime with a right-hand key press (assuming that the positive key is on the right-hand side), or if the act of identifying the primes actually reinforced participants' positive or negative associations with those stimuli. The former explanation illustrates the peripheral locus view, whereas the latter explanation is an example of the central locus view. In order to tease these alternate explanations apart, Abrams et al. (2003) performed another experiment in which participants were asked to classify the objects as either good or bad, and then perform a subliminal priming task. Importantly, just prior to performing the priming task, the experimenters switched the keys that were used to make positive and negative responses (for example, if the positive key was on the right for the classification task, it was on the left for the priming task). The authors found significant subliminal priming effects even when the keys were reversed, thus lending support for the central locus view.

In the RM model, a very similar issue arises. Recall that neutral primes show a contrast effect when placed in a task alongside extreme negative or positive primes. Importantly, it is not yet clear whether this contrast effect involves (a) a simple process of associating the neutral primes with a motor response (peripheral locus), or (b) a process of associating the neutral primes with an evaluative response (central locus). According to a peripheral locus view, response mapping merely reflects the fact that a motor response (rather than an evaluative association) has been mapped onto the primes. However, according to the central locus viewpoint, response mapping *actually causes these neutral objects to take on new evaluative associations*.

To further illustrate the central locus view, imagine a priming task in which half of the primes are negative, and half are neutral (see left side of Figure 2). Participants walk into the experiment with no preexisting evaluative associations with the neutral primes. Moreover, when they begin the task, the objects seem no more positive or negative than they would in any other context. However, note that in this example, the neutral primes will be mapped onto the positive response. At first, this positive response will merely be a default label with which to refer to the household objects. Yet as the trials proceed, participants will learn to respond to the objects favorably. That is, the label that was once applied by default actually comes to define the objects themselves. Hence, response mapping may not be simply a process of learning to associate primes with a particular motor response. Instead, response mapping may involve learning to associate primes with an *evaluative response.* Thus, priming tasks may have the capacity to create new evaluative associations.

The findings reported to date—including all of the experiments reported in Scherer and Lambert (2009)—are equally compatible with both a peripheral and central locus view. The purpose of Experiments 1 and 2 is to test the viability of these two respective views. To avoid a possible misunderstanding, the purpose of these studies is test whether the central (evaluative based) or peripheral (motor based) account provides the best explanation of the findings obtained in the present paradigm. However, failure to find support for the peripheral view does *not* necessarily imply that processes related to motor learning are unimportant and never occur.

In order to understand the logic of the studies to follow, it is worth emphasizing that, according to the central locus view, response mapping results in changes in liking for the neutral prime stimuli. If true, this view predicts that *the evaluative associations that are learned in the priming task should be reflected in future judgments of the primes*. For example, if participants learn negative associations with a picture of a towel, then this towel should elicit negative priming effects in a different priming task (Experiment 1), and it should also be explicitly rated relatively negatively as well (Experiment 2). On the other hand, if response mapping merely reflects a learned motor response, then the negative effects for that towel should only occur in the priming task, and should *not* carry over to future judgments of that stimulus.

Unresolved Issue #2: On the Mental Classification of Primes: On What Basis, Exactly, Are Participants Sorting the Primes into Different Categories?

The RM model may, at first blush, seem difficult to reconcile with some past research showing that priming effects depend on associative strength. For example, in an experiment by Fazio et al. (1986; Experiment 3), strongly valenced primes showed strong priming effects, whereas weakly-valenced primes showed weak effects. Response mapping might seem to have difficulty accounting for such effects because the model assumes that even weak primes will be mapped onto a response (in which case they will produce strong priming effects), or alternatively, the weak primes will not be mapped at all (resulting in zero priming effects). This all-or-none conceptualization of priming is not compatible with results such as those obtained by Fazio et al.

In my view, there is one particularly intriguing explanation that may resolve this apparent inconsistency. Remember that one of the central tenets of response mapping is that participants impose the target classification scheme onto the primes. In other words, response mapping relies on participants' ability to take a broad, response-based classification scheme and apply it to the prime stimuli. Hence, it is possible that response mapping processes *may be limited to situations in which the primes are easily sorted into a small number of* *homogenous categories.* For example, in Scherer & Lambert's (2009) experiments, the primes always fell into two very distinct and homogenous categories (e.g. flowers and objects, or animals and objects). By contrast, in the Fazio et al. (1986) experiments the primes included a broad array of unrelated words (e.g. "aquarium," "cake" and "recession") that did not form any coherent categories at all. As a result, it's possible that the participants in the Fazio experiments could not map the primes because they could not easily categorize them. Under these sorts of circumstances, it is likely that other processes (such as spreading activation) may play a more dominant role in producing priming effects (I will revisit this issue in the General Discussion).

One important implication of the above discussion is that many measures of implicit attitudes may actually be best understood as measures of *implicit categorization*. That is, these tasks may specifically measure participants' ability (or lack thereof) to categorically differentiate between various types of stimuli. If the primes are viewed as being categorically distinct, then they are likely to be mapped onto separate responses. In contrast, if the primes are categorically similar, then they are likely to be mapped onto the same response.

To illustrate, turn again to Figure 1b. This figure represents two primes that appeared in Scherer and Lambert's (2009) priming experiments. One prime is an extremely negative animal, and the other is a neutral household object. In this situation, participants can easily dissociate between the two primes, because they are evaluatively distinct (i.e. negative and neutral) and also categorically homogenous (animals and common household objects). As a result, participants map these primes onto different responses (negative and positive, respectively).

Now turn to Figure 3a and 3b. These figures also represent two primes that might appear in a priming task. However note that in Figure 3a the primes are categorically distinct (i.e. presented in greyscale or in color), but evaluatively identical. One important question is whether participants will use non-evaluative differences between the primes—such as the presence or absence of color—as a basis for mapping the primes onto different evaluative responses. If this is the case, then responses in a task like that represented in Figure 3a should be quite similar to the responses in a more typical priming task, like that represented in Figure 3b. This empirical question is addressed in Experiment 3.

Previously, I pointed out that in my past experiments all of the primes were a) evaluatively distinct, and b) categorically homogenous. For example, in one experiment the negative primes were always animals, whereas the neutral primes were household objects. Therefore, one question that remains unaddressed is whether response mapping occurs under circumstances in which the primes are *evaluatively* distinct, but are *categorically* diverse. For example, instead of presenting participants with a homogenous class of neutral household objects, participants might be presented with neutral pictures that include a wide array of things, such as landscapes, household objects, animals, etc. In Experiment 4 I will vary the homogeneity of the primes, in order to determine whether or not response mapping processes are dependent on the homogeneity of the prime categories.

Summary

At this point, I have identified two important unanswered questions concerning the processes underlying response mapping. The first question is whether response mapping reflects a learned motor response (the peripheral locus view), or if the priming task actually causes changes in liking for the neutral prime stimuli (the central locus view). This question will be addressed by Experiments 1 and 2. The second question asks on what basis, exactly, do participants classify and map the primes? This question will be addressed by Experiments 3 and 4.

EXPERIMENT 1

As demonstrated by previous research (Scherer & Lambert, 2009), a prime that is actually neutral may, in some circumstances, be mapped onto a positive or negative response by default. One interpretation of this effect is that it is a simple learned motor response that has absolutely no bearing on how participants feel toward the primes (henceforth the *peripheral locus hypothesis*). However, an alternative view is that response mapping actually causes participants to form *new* evaluative associations with those primes (henceforth the *central locus* *hypothesis*). That is, the simple act of pairing a prime stimulus with an evaluative response ends up changing the way participants feel toward that stimulus. The purpose of Experiment 1 is to determine whether response mapping is best understood as a learned motor association (the peripheral locus view) or learned evaluative association (central locus view) (c.f. Abrams & Greenwald, 2000; Abrams, Klinger & Greenwald, 2003).

The central locus hypothesis makes two related predictions. First, the associations that are formed in the initial priming task should carry over to future priming tasks, even if the nature of that subsequent task is different from the first. For example, suppose that participants are presented with two prime images; a towel and a threatening-looking dog. In this circumstance, the towel will be mapped onto the positive response as a way of dissociating it from the dog, which is clearly negative. The central locus hypothesis predicts that this act of mapping the towel onto a positive response will result in favorable attitudes toward the towel that did not exist prior to the priming task. Hence, if the towel is then presented again in a second priming task, this towel will continue to elicit positive responses, even if the threatening dog is no longer present. The central locus hypothesis additionally predicts that since the neutral primes are mapped onto an *evaluative* label, it should not matter which hand is used to make positive or negative responses. Continuing with my example, the towel prime will elicit positive responses in the second priming task, regardless of whether the hand that is used to make positive responses is the same as in the previous task, or different.

In contrast, the peripheral locus hypothesis suggests that participants simply learn to associate the towel with a particular motor response, for example, a right or left-hand key press. Continuing with the above example, if the positive label is on the right-hand side, then the towel will elicit positive right-hand responses. If the positive label is then switched just prior to performing the second task, then the towel will continue to elicit right-hand responses even though the right hand now represents the *negative* response. Hence, the peripheral locus hypothesis suggests that the towel becomes associated with a motor response that is not related to evaluative meaning.

Experiment 1 tested the viability of the central locus versus peripheral locus hypotheses. There were two experimental blocks in the present experiment. In Block 1, participants completed an AMP priming task in which negative and neutral, or positive and neutral, primes were presented. It was expected that Block 1 would replicate previous findings, insofar as the neutral primes will exhibit a contrast effect (i.e. elicit positive bias in the former context and negative bias in the latter). In a second block of the experiment, participants' automatic attitudes toward the neutral primes were assessed again, this time using a second AMP task. In the Block 2 priming task, half of the trials presented the same neutral primes that participants saw in phase 1, whereas the rest of the trials contained a grey square control prime.

An important factor in this experiment pertained to the position of the "pleasant" and "unpleasant" responses on the computer keyboard. For half of the participants, the hand that was used to make pleasant and unpleasant responses was switched between Block 1 and Block 2. For example, if the pleasant key was associated with the left hand in Block 1, it was later associated with the *right* hand in Block 2. The purpose of this key-switching manipulation was to completely rule out the possibility that participants associate the neutral primes with a particular hand (i.e. motor response), rather than an evaluative response. For the rest of participants, the key labels remained on the same side for both blocks.

In this experiment, the central and peripheral hypotheses made two opposing predictions with regard to the responses that the neutral primes will activate in Block 2. The central locus hypothesis predicts that in Block 2, the neutral primes will activate the same evaluative response that they activated in Block 1. This should be true even when a different hand must be used to make pleasant and unpleasant responses. In contrast, the peripheral hypothesis predicts that the neutral primes will activate the same *hand* across Block 1 and Block 2, irrespective of the evaluative label that is associated with that hand.

Participants and Design

Participants were 111 undergraduate students who participated in return for partial course credit. This experiment was a 2 X 2 X 2 X 2 mixed model design. The first factor pertained to the nature of the priming context in Block 1. Half the participants were assigned to a task in which positive and neutral primes were presented, and the rest of the participants were presented with negative and neutral primes. Hence, the neutral primes were presented in a "positive" or "negative" priming context. The second factor pertained to the primes in the two tasks, in which half were always neutral, and the rest were either evaluatively extreme primes (Block 1) or a grey square control prime (Block 2). The third factor pertained to the location of the pleasant and unpleasant keys at the beginning of the experiment. For half of the participants, pleasant responses were made with the left hand at the start of the experiment (and unpleasant responses were made with the right hand), and for the rest of the participants this configuration was reversed. The final factor pertained to whether the hand that was used to make pleasant and unpleasant responses was reversed between Block 1 and Block 2, or not.

Procedure

In Block 1, participants were randomly assigned to one of two conditions, pertaining to the nature of the primes that appeared in the AMP priming task. In the positive priming context, participants were presented with four positive and four neutral primes. In the negative priming context, participants were presented with four negative and four neutral primes. Participants were told that they would see images flash in the center of the computer screen, and that the first image would be a picture. They were told to ignore this picture (the prime), and to only respond to the following Chinese character target. They were told that their job was to indicate whether the target was aesthetically pleasant or unpleasant, as fast as possible.

In Block 2, participants completed another AMP priming task. In this block, the priming task was configured so that half of the trials presented the same neutral primes that participants viewed in Block 1. On the remainder of the trials, a grey square (control) prime was presented. As in Block 1, participants were told to ignore the primes and identify the subsequent targets as being either pleasant or unpleasant. In addition, the hand that was used to make pleasant and unpleasant responses was switched for half the participants between Block 1 and Block 2. For example, if pleasant responses were made with the left hand in Block 1, then they were made with the right hand in Block 2.

AMP Task Parameters

The priming task procedure closely followed Payne et al. (2005). On each trial, the rectangular prime appeared in the center of the screen for 75 milliseconds, and was immediately followed by the target Chinese character which remained on the screen for 100 milliseconds. The target was then replaced by a mask that was approximately the same size and shape as the prime and target pictures, and which remained on the screen until participants made their responses. After making their response, the next trial began immediately.

Participants were given 6 practice trials prior to beginning the key experimental blocks. As mentioned previously, each experimental block contained 2 types of primes, and each prime-type appeared for exactly half of the trials. The presentation of the primes was randomized, and each type of prime was followed by one of 80 randomly selected Chinese character targets. This design resulted in a total of 72 trials per block (8 primes X 9 repetitions).

Stimuli

Primes consisted of 4 neutral, 4 positive and 4 negative pictures, all of which were selected from the International Affective Picture System (IAPS, Lang, Bradley, & Cuthbert, 2008) on the basis of the normative data collected by Lang and his colleagues. Four of the pictures were rated as very positive (e.g. puppies, kittens, rabbits; average valence rating = 7.67, average SD = 1.43), four were rated as very negative (e.g. a snarling dog, cockroaches, spiders; average valence rating = 3.73, average SD = 1.88), and four pictures received ratings around the midpoint of the scale (e.g. a towel, mug, lamp; average valence rating = 4.87, average SD = 0.99). The IAPS scale ranged from 1 (negative) to 9 (positive). The targets were 80 Chinese character that were used in Payne et al. (2005).

Summary of Predictions

According to the central locus hypothesis, response mapping will cause participants to form new evaluative associations with the neutral primes. If this is true, then the evaluative associations that are learned in Block 1 should continue to influence participants' responses in Block 2. In contrast, the peripheral locus hypothesis assumes that response mapping produces a simple learned motor association. Hence, the peripheral locus hypothesis alternatively predicts that the neutral primes should continue to activate whichever hand they activated in Block 1, regardless of the evaluative label associated with that response.

Results

Analysis of Block 1: Test of Contrast Effect Replication

Data from Block 1 were coded so that pleasant responses were represented by the number 1, and unpleasant responses were represented by the number 0. For each prime type, the proportion of pleasant to unpleasant responses was calculated. The resultant index ranged from 0 (representing all unpleasant responses) to 1 (representing all pleasant responses). A score of 0.50 represents an equal number of pleasant and unpleasant responses.

Recall that in Block 1, participants were assigned to one of two conditions, in which the primes were either neutral and negative, or neutral and positive. Hence, it was expected that Block 1 would replicate the pattern of contrast effects that have been obtained in numerous past experiments (Scherer & Lambert, 2009). To test for that replication, a 2 (Prime: valenced vs. neutral) X 2 (Context: positive vs. negative) mixed model ANOVA was conducted on the data from Block 1. This ANOVA revealed significant a Prime X Context interaction ($F(1, 109) = 67.27$, $p < .001$). Table 1 shows that the contrast effects obtained in past research were indeed replicated. Simple effects tests revealed that the neutral primes elicited significantly more pleasant responses in the negative context as compared to the positive context, $F(1,109) = 28.74$, $p < .001$. Additional tests also revealed that, as expected, the positive primes elicited significantly more pleasant responses than the negative primes, $F(1,109) = 50.25$, $p < .001$.

Analysis of Block 2

The central and peripheral hypotheses make different predictions about the particular hand that participants will use to respond to the neutral primes in Block 2. Because of this fact, responses in Block 2 were coded to represent the hand that was used to make the response. Left-hand responses were represented by the number 1, and right-hand responses were represented by the number 0. For each prime type, the proportion of left-hand to right-hand responses was calculated. The resultant index ranged from 0 (representing all right-hand responses) to 1 (representing all left-hand responses). A score of 0.50 represents an equal number of right and left-hand responses.

Recall that in Block 2, participants performed a task that presented the same neutral primes that were presented Block 1, along with a control prime (a grey square). The peripheral locus hypothesis predicts that in this second task, participants will respond to the neutral primes using the same *hand* that was used to respond to those primes in Block 1. In contrast, the central locus hypothesis predicts that the neutral primes will continue to activate the same *evaluative* response as in Block 1, regardless of which hand is used to make that response. To analyze the primes in Block 2, a 2 (Prime: neutral vs. grey square) X 2 (Context in Block 1: positive vs. negative) X 2 (Response Location in Block 1: pleasant on left vs. right) X 2 (Switch condition: switch vs. no switch) mixed-model ANOVA was conducted. This analysis revealed a significant 4-way interaction ($F(1, 103) = 5.18$, $p < .05$), which indicated that the neutral and grey square primes elicited a different pattern of effects. Hence, the analyses below will conduct simple effects tests on the neutral and grey square control primes separately, in order to understand the nature of this complex interaction.

Analysis of the neutral primes: testing the Central locus vs. Peripheral locus hypotheses

The next set of analyses involves the critical predictions concerning the peripheral versus central hypotheses. To test the viability of these competing hypotheses, I conducted analyses on the neutral primes from Block 2. Three experimental factors were expected to influence participants' responses to the neutral primes in Block 2: 1) the evaluative context experienced in Block 1, 2) the location of the pleasant and unpleasant keys at the start of the experiment, and 3) whether the location of the pleasant/unpleasant keys were switched just prior to Block 2. If the peripheral locus hypothesis is correct, then this will emerge as a Context X Key Location interaction, because the neutral primes will activate whichever hand they activated in Block 1, regardless of whether the keys were switched, or not. However, if the central locus hypothesis is correct, then this will be expressed as a 3-way Context X Response Location X Switch Condition interaction, because when the key labels are switched just prior to Block 2, participants will respond to the neutral primes using the *opposite* hand that was used in Block 1.

A UNIANOVA analysis of the neutral primes revealed a 2-way Response Location X Switch condition interaction ($F(1,103) = 26.10$, $p < .001$), which simply indicated that participants made more pleasant than unpleasant responses overall (this is because the hand that was used to make pleasant responses depended on both the location of that response in block 1, as well as whether or not the response configuration was switched or not). However this interaction was qualified by the 3-way Context X Response location X Switch condition interaction that was predicted by the central locus hypothesis, $F(1, 103) = 9.93$, $p < .01$. To illustrate how these results support the central locus hypothesis, turn to Figure 4. The top half of Figure 4 displays the response patterns for participants for whom the location of the keys was *not* switched between blocks 1 and 2. Participants who were previously exposed to the negative priming context responded to the neutral primes as if they were positive, whereas the reverse was true for participants previously exposed to the positive priming context. This was true regardless of whether the "pleasant" key was on the right or left-hand side.

The bottom half of Figure 4 displays the data that are critical to the predictions made by the central locus hypothesis. These data display participants' responses when the locations of the pleasant and unpleasant keys are switched just prior to Block 2. Importantly, the contrast effect was still observed in this condition. That is, the neutral primes activated relatively more pleasant responses when the previous context was negative (black bars) than positive (white bars). These results suggest that participants associated the neutral primes with an *evaluative* response, rather than a right or left hand response.

Analyses Involving the Grey Square Prime

Next, I conducted another UNIANOVA, this time on the grey square control prime. This analysis revealed a Switch condition X Response Location interaction ($F(1, 103) = 33.58$, $p < .001$), which again simply indicated that participants tended to hit the pleasant key more often than the unpleasant key. Also, a Location X Context interaction emerged, $F(1, 103) = 3.98$, $p < .05$. Although this interaction was not qualified by the switch condition manipulation (3-way interaction $p = .12$), it is useful to display the means associated with the 3-way interaction so that a direct comparison can be made between the gray square and the neutral primes (recall that the precise nature of the earlier 4-way interaction has yet to be examined. The means displayed across Figures 4 and 5 represent that 4-way interaction). The top half of Figure 5 demonstrates that in the no-switch condition, the grey square elicited effects that were no different from that of the neutral primes (i.e. top halves of Figures 4 and 5 are not different, $F < 1.0$). In the switch condition (bottom half of Figure 5), however, the grey square elicited no effects involving the previous task context (all task context effects $F < 1.0$). In the switch condition, the effects for the grey square were significantly different from that of the neutral primes (i.e. bottom halves of Figures 4 and 5 are significantly different, $p < .01$). These results lend insight into the 4-way interaction that was reported earlier. Specifically, the difference between the neutral and grey square primes that was illustrated by that interaction can be explained, in part, by the lack of a contrast effect for the grey square prime in the switch condition.

Discussion

Experiment 1 provides strong evidence for the central locus hypothesis. That is, evaluative priming tasks can generate new evaluative associations that did not exist prior to the task. This is important, because the vast majority of research assumes that these sorts of tasks measure *preexisting* attitudes, whereas the present research suggests that priming tasks both measure *and create new* attitudes.

Experiment 1 showed that priming tasks can cause evaluative conditioning effects. The RM model suggests a process by which such evaluative conditioning can take place:

1) The priming task instructions ask participants to sort the target stimuli into good and bad categories. As a result of these task instructions, a good-bad classification scheme becomes salient.
2) The salience of the good-bad classification scheme causes participants to apply good and bad responses to all the stimuli in the task, including the primes (which they were instructed to ignore).
3) The unambiguous primes are classified in accordance with their evaluative implications, as might be expected. However, the neutral primes are *also* classified as either positive or negative (as a result of the RM assumptions described earlier), even though they possess no evaluative associations.
4) As a result of 1-3, each time a neutral prime appears, it activates an evaluative response (either positive or negative, depending on the nature of the other primes in the task).
5) Participants develop a new evaluative association with the neutral prime stimuli, as a result of step 4. An evaluative response is now tied to the neutral primes, and as a result, future responses to these stimuli will be more positive or negative than would otherwise be the case.

The process described above can be referred to as *task initiated categorization and evaluative conditioning*. That is, the task instructions initiate categorization of the prime stimuli, which leads to response mapping and then evaluative conditioning. These effects have important implications for those who wish to study social attitudes, because priming tasks are generally employed as a way of measuring participants' attitudes towards other people, especially outgroup members. The results of a given priming task are assumed to reflect participants' preexisting attitudes, but the present research suggests that the task may in fact create attitudes that participants did not possess when they walked into the experiment. For example, if a given task shows negative bias toward the elderly and positive bias toward the young, it could be that the task has, in fact, created a negative association with the elderly that did not exist until the onset of the task.

The effects elicited by the grey square prime were not relevant to the critical hypotheses that were tested in this experiment, but the effects were nonetheless interesting. When the keys were not switched prior to block 2, the grey square activated response patterns that were virtually indistinguishable from those elicited by the old neutral primes. Another way of stating this is that the new primes showed an *assimilation* effect (e.g. Strack & Schwarz, 2007). However, when the key location *was* switched, this assimilation effect disappeared. It is unclear why the grey square elicited assimilation-like effects in the no-switch condition, but not in the switch condition. One possible explanation is that the grey square was similar enough to the neutral household objects (i.e. both were neutral and dull) that they could show assimilation under certain conditions, but that they were not sufficiently similar to show assimilation in the face of disruptive task changes, such as the switch in key label. Another possibility is that the grey square revealed some sort of underlying response bias, but that this response bias cannot entirely account for the effects for the neutral primes. These questions are important and interesting, and are certainly deserving of future exploration. However, these questions do not qualify the central findings of this experiment, which demonstrated that the task caused the (previously) neutral stimuli to activate evaluatively meaningful responses.

Perhaps the most compelling implication of the central view of response mapping is that response mapping processes result in a new attitude toward previously neutral stimuli. Importantly, this means that priming tasks may sometimes create, as well as measure, attitudes. In the next experiment, I will further test the boundary conditions of this effect. Specifically, I will determine whether *overt ratings* of the stimuli can also be influenced by prior response mapping.

EXPERIMENT 2

The previous experiment provided some initial evidence that response mapping leads to the formation of new evaluative associations. This is important, because it suggests that priming tasks can both measure attitudes on one hand, and create attitudes on the other. However, one limitation of Experiment 1 was that both the manipulation and the dependent measure were priming paradigms. Hence, it is possible that the "new attitudes" toward the neutral primes are only activated within the limited circumstances of a priming task. For example, Experiment 1 showed that the neutral primes can activate "pleasant" or "unpleasant" priming task responses, but this does not necessarily mean that these primes will activate the kinds of positive or negative feelings that might be reflected in other, more overt or deliberate types of judgments. Moreover, if one were interested in using the AMP as an evaluative conditioning paradigm, it would be theoretically important to know that the evaluations that are formed in the task will influence a broad array of behaviors and judgments, rather than simply influencing priming task judgments. Hence, the purpose of the present experiment is to determine whether the evaluative response that is learned in the priming task can influence later, explicitly expressed attitudes.

As in Experiment 1, there were two blocks in Experiment 2. In Block 1, participants completed an AMP task that was exactly like Block 1 of Experiment 1. However, Block 2 involved viewing the neutral primes again and *explicitly* *rating* those images along three different dimensions: pleasantness/unplesantness, positivity/negativity, and valuable/worthless. According to the central view of response mapping, participants will form new evaluations in the Block 1 priming task. This new evaluation should, in turn, influence later explicit judgments of pleasantness, positivity and value.

Participants and Design

Participants were 32 undergraduate students who participated in return for partial course credit. There was one between-subjects factor in this experiment, pertaining to the nature of the primes that were presented in the AMP task (positive versus negative).

Procedure and Stimuli

In Block 1, participants were randomly assigned to complete an AMP task that presented either positive and neutral, or negative and neutral, primes. This AMP task was identical to Block 1 of Experiment 1. In Block 2, participants were presented with the neutral prime images again, only this time they were asked to rate the images with respect to three dimensions: *pleasantness-unpleasantness, positive-negative,* and *valuable-worthless*. Each rating was made on a -4 to +4 Likert scale. The order of the pictures was randomized, as was the order in which participants were asked each of the three distinct types of questions.

Summary of Predictions

The predictions for Experiment 2 were similar to those made for Experiment 1. If the central locus view is correct—meaning that participants form new evaluative association with the primes—then these new associations should influence participants' later, explicit evaluations of those primes. For example, when the priming context is negative, then the neutral primes should be mapped onto the positive response. If this response mapping process results in the formation of a new, positive attitude, then participants should later rate the neutral primes more favorably than would otherwise be the case.

Results

For each of the three types of primes (positive, negative and neutral), an index was created that represented the proportion of pleasant to unpleasant responses. This index ranged from 0 (all unpleasant responses) to 1 (all pleasant responses). A score of 0.50 represents an equal number of unpleasant and pleasant responses. For each type of explicit rating (pleasantness, positivity, value), I averaged across participants' responses to the four neutral primes. This resulted in three explicit rating indices: 1) pleasantness/unpleasantness, 2) positive/negative, and 3) valuable/worthlessness.

Priming Task Results

Before a test of the critical hypotheses can be conducted, it is first important to establish that the contrast effect was replicated within the Block 1 priming task. A 2 (Prime type: unambiguous vs. neutral) X 2 (Context: positive vs. negative) mixed model ANOVA revealed the expected contrast effect in the form of a 2-way Prime X Context interaction, $F(1,30) = 17.08$, $p < .001$ (see Table 2). Simple effects tests revealed that the neutral primes elicited significantly more pleasant responses in the negative context as compared to the positive context, $F(1,30) = 6.23$, $p < .05$. Additional tests also revealed that, as expected, the positive primes elicited significantly more pleasant responses than the negative primes, $F(1,30) = 9.28$, $p < .01$.

Explicit Ratings Results

The central locus hypothesis predicts that the priming task (Block 1) should influence participants' later explicit ratings of the neutral stimuli (Block 2). In order to test this hypothesis, a 3 (Type of rating: pleasant vs. positive vs. value) X 2 (Prior Priming Task Context: positive vs. negative) mixed-model ANOVA revealed a main effect of context ($F(1,30) = 4.62$, $p < .05$), indicating that each of the three types of ratings were influenced by the prior priming task context. All other effects were not significant, $p > .20$. Figure 6 displays the means associated with this test. This figure shows that participants who viewed the neutral primes in a negative priming context later rated those primes as more pleasant, more positive and more valuable than participants who were exposed to a positive priming context.

Discussion

These results call into question an important and commonly-held assumption about priming tasks. Specifically, priming tasks are widely assumed to measure participants' automatic associations. Yet the present research shows that these tasks can both measure attitudes (in the case of the extreme primes), and also *create attitudes where none existed previously* (in the case of the neutral primes). This is the first time (to my knowledge) that priming tasks have been proposed as vehicles for attitude creation in addition to attitude measurement.

In Experiment 2, the initial priming task influenced participants' subsequent explicit judgments of the neutral object stimuli. When the initial priming context was negative, the neutral primes activated more pleasant responses in the priming task, and elicited more favorable overt ratings as well. The reverse was true when the initial task context was positive. Moreover, every type of rating—pleasantness, positivity, and value—was influenced by the prior context. For example, if participants responded to a picture of a towel with a "pleasant" key press, they also later tended to think that the towel was more valuable than would otherwise be the case. Although one might have expected this effect to occur at least for the explicit pleasantness ratings (after all, the priming task also asked for pleasant and unpleasant responses), it turned out that the biggest effect was actually for the value rating (see Figure 6, far right bars).

These results lend support for the central locus view of response mapping. That is, the simple act of responding to the neutral primes with a positive or negative key press eventually results in a positive or negative attitude toward those images. In this experiment, participants do not begin the task with strong feelings toward the neutral primes. However, participants finish the experiment with evaluative associations that were not present before the task began. As such, response mapping can cause evaluative conditioning effects. If participants make a “pleasant” key press every time a towel appears, they will come to view that towel in a positive light. If they make an “unpleasant” response following a towel, then the towel will later be perceived as negative.

EXPERIMENT 3

Experiments 1 and 2 addressed some important implications of the response mapping model. In the next two experiments, I will shift the focus slightly and instead explore some of the potential boundary conditions of the response mapping model. Hence, the purpose of Experiments 3 and 4 is to identify conditions in which response mapping is, or is not, likely to occur.

Experiment 3 addresses two different proposed forms of the RM model. One form, which I will refer to as the “strong” form of response mapping, suggests that participants will *always* map the primes onto separate responses, so long as there are two or more classes of primes that are categorically distinct, and regardless of whether the primes differ on an evaluative dimensions. However, another view of the model is that response mapping has at least one clear boundary condition, which is that the primes must be evaluatively distinct in order for participants to map the primes onto the positive and negative responses. I will refer to this latter view as the "evaluative distinctiveness" hypothesis. The evaluative distinctiveness hypothesis fits well with current conceptualizations of priming, which assume that priming effects are the result of the evaluative implications of the primes. The strong form of RM, however, suggests that priming effects can and do occur, even in the absence of any strong evaluations toward any of the primes.

To illustrate the strong form of response mapping, turn to Figures 3a and 3b. In this figure, there are two prime images (in this case, two faces). In Figure 3a, both of the primes are the same male face, wearing a neutral expression, but one of the faces is presented in color, and the other is presented in black-and-white. In contrast, in Figure 3b the two faces have different evaluative connotations (happy and angry). The strong form of the RM model predicts that priming effects will occur in both Figure 3a and 3b: In Figure 3b, the primes will be mapped onto different responses because the primes are evaluatively distinct. Importantly, in Figure 3a, the primes will *also* be mapped onto different responses, because the color (and lack thereof) of the primes will act as an evaluative dimension. That is, participants will identify the fact that there are two distinct types of faces (color and grayscale), and then will choose a positive or negative response to refer to each type of prime, perhaps because they have a mild preference for color images over grayscale (or vice versa). Hence, the strong form of the RM model predicts participants who are shown primes like those illustrated in Figure 3a will show significant priming effects, even though the primes in that condition have no obvious evaluative connotations.

In the case of Figure 3a, the arrows suggest that the color prime is mapped onto the positive response, and the grayscale prime is mapped onto the negative response. That is, the figure assumes that participants have a preference for color over grayscale. However, participants could easily map the primes in the opposite fashion, perhaps as a result of a preference for grayscale over color. Yet regardless of how participants map the primes, the strong form of the RM model suggests that participants will always end up strongly associating the primes with an evaluative response, because of slight preferences for one group of primes over another group.

Now suppose that the strong form of RM is true, and about half the participants prefer color images and half prefer grayscale. How would these preferences be reflected in the group data as a whole? The answer is that if one were to average across all of the participants, then it would appear as though the sample as a whole possesses no particular preference at all. That is, a null effect (i.e. approximately equal effects for both primes) would be obtained. Yet that “null effect” would actually conceal strong priming effects for each individual participant. In sum, if half of the sample has one set of preferences, and the rest of the sample has an opposite set of preferences, then the sample as a whole will appear to have no preferences at all. This can explain how the strong form of the RM model could potentially be true, even when researchers sometimes obtain null findings.

Of course, it remains to be determined whether this strong form of the RM model has any merit. According to the alternative evaluative distinctiveness hypothesis, the evaluative implications of the primes play a crucial role in producing priming effects. To illustrate, turn again to Figure 3a, in which the primes are categorically distinct but evaluatively identical. The evaluative distinctiveness hypothesis predicts that participants will simply fail to map any of the primes in this case, because they all share the same evaluative connotation. However, the evaluative distinctiveness hypothesis *does* predict that priming will occur in circumstances where the primes have different evaluative connotations, such as the task displayed in Figure 3b.

The purpose of Experiment 3 is to test the strong form of the response mapping framework, and compare it to the alternative evaluative distinctiveness hypothesis. In this experiment, participants completed a single priming task in which the primes were just like those displayed in either Figure 3a or Figure 3b. The strong RM model predicts that equal priming effects will be obtained in both conditions. The evaluative distinctiveness view predicts that priming effects will only obtain when the primes have clear evaluative connotations, such as in Figure 3b.

Participants and Design

A total of 39 students participated in this experiment in return for partial course credit. There was one between-subjects factor in this experiment, pertaining to the nature of the primes presented in the priming task. In one condition, the primes were faces that were distinguishable only on the basis of whether they were presented in color or not. I will henceforth refer to this as the *non-evaluative priming* condition. In the other task, all of the primes were presented in color, but they were evaluatively distinct; that is, half of the primes were clearly positive (happy), and the rest were clearly negative (angry). I will refer to this as the *evaluative priming* condition.

Procedure

All participants were randomly assigned to complete one of the two AMP tasks. The trial configuration and timing of the AMP tasks was identical to Experiments 1 and 2. However, in the present experiment, all of the primes were White male faces.

In the non-evaluative priming condition, the primes were pictures of four men, each of whom were wearing a neutral expression. For half the trials, these four faces were presented in color. For the rest of the trials, the faces were presented in grayscale. In the evaluative priming condition, the primes were pictures of the same four men. However, in this second task, the men were wearing positive expressions in half of the trials, and negative expressions for the other half of trials. After participants completed the AMP task, they were asked to rate the pleasantness of each of the pictures that they saw in the priming task on a -4 to +4 Likert scale.

Stimuli

Stimuli were created using FaceGen software. There were four White male faces used as primes in this experiment. Each face was altered so that it expresses positive, neutral, and negative emotions. Hence, there were 12 different face stimuli in total (4 faces X 3 expressions).

FaceGen allows the user to manipulate the emotional expression of each face by moving tabs that correspond to the eyes, mouth, eyebrows, etc. The negative faces in this experiment conveyed anger (rather than sadness or anxiety). The extremely negative faces were constructed by moving the “anger” tab in FaceGen to the 100%. This resulted in a face that possessed the maximum amount of expressed anger, allowable by FaceGen, and includes furrowed eyebrows, squinted eyes, and a snarling mouth (see Figure 3b).

The positive faces in this experiment conveyed happiness. Since there is no “happiness” tab in FaceGen, these faces were created by moving individual components of the face. These extremely positive faces were constructed by using a combination of 100% smile closed, 100% smile open, 100% eyebrows up, and 50% eye squint. These features correspond to what has been defined as a Duchenne smile. The neutral faces were constructed by setting all of FaceGen's emotional expression tabs to zero.

Summary of Predictions

The strong form of the response mapping model predicts that, so long as there is some kind of categorical difference between the primes, participants will map these primes onto separate "pleasant" and "unpleasant" responses. Participants are expected to map the primes onto separate responses even when the primes are evaluatively neutral and distinguishable only on the basis of the presence or absence of color. Alternatively, response mapping may rely on the presence of an *evaluative* distinctiveness between the primes. More generally, this view predicts that response mapping will only occur if the primes can be sorted according to the categories that are specified by the response labels. If this latter hypothesis is correct, then we should find much stronger response mapping when the primes are evaluatively distinct, but little or no response mapping when the primes are evaluatively similar.

Results

For each of the four types of primes (happy, angry, neutral- grayscale, neutral-color), an index was created that represented the proportion of pleasant versus unpleasant responses. This index ranged from 0 (all unpleasant responses) to 1 (all pleasant responses). A score of 0.50 represents an equal number of unpleasant and pleasant responses.

Data Analysis Strategy

In this section I will describe in detail the procedure that was used to analyze the data and test the two opposing hypotheses (the strong RM model vs. the evaluative distinctiveness hypothesis). Recall that in one condition, primes included grayscale and color pictures of men wearing neutral expressions (see Figure 3a). In this *non-evaluative priming condition*, it could easily be the case that half of the participants prefer the color primes whereas the rest prefer the grayscale primes. If this turns out to be the case, then averaging across all the participants will result in a null effect, because any given prime will tend to elicit pleasant responses from half the sample and unpleasant responses from the rest of the sample. As a result of these considerations, it is necessary to create an index that represents the difference in response rates to the two types of primes, but that is indifferent to the particular (pleasant or unpleasant) response that is activated by any given prime.

To arrive at such an index, I first calculated the proportion of pleasant responses for each prime-type. This resulted in four indices, one each for the happy, angry, neutral-color and neutral-grayscale primes. Next, I subtracted the effect for one prime category from the effect for the other prime category, to arrive at a difference score. For example, in the evaluative priming condition, this difference score was the difference between the angry and happy primes. In the non-evaluative condition, this score was the difference between the grayscale and color primes. Finally, I took the absolute value of that difference score. The resulting index ranged from 0 (no difference between the effects for the two types of primes) to 1 (one prime category elicited 100% pleasant responses, while the other prime category elicited 100% unpleasant responses). Hence, this index reflects the difference between participants' responses to each type of prime, irrespective of which particular response each prime happened to facilitate. This index can be thought of as a direct measure of response mapping, with high numbers indicating that the primes activated completely separate responses (i.e. strong response mapping), and lower numbers indicating that the primes don't activate any particular response (i.e. little or no response mapping). I will henceforth refer to this index as the *response mapping score.*

Traditional Analysis of the Priming Data

The purpose of this section is to perform priming task analyses of the sort that are traditionally used in this area of research. Such analyses typically involve testing for differences in response rates for the primes in the task. Since this was not a crossed design, a 2 (evaluative vs. non-evaluative condition) X 2 (prime type) analysis is not meaningful. Instead, I separately analyzed the response rates for the primes in each between-subjects condition. In the evaluative priming condition, the angry primes elicited significantly fewer pleasant responses than the happy primes, $F(1,18) = 15.01$, $p < .001$. In the non-evaluative condition there was no difference between the color and grayscale primes, $F(1,19) = 1.73$, $p > .20$.

Testing the Critical Hypotheses

The strong form of the response mapping model predicts that the primes in both the evaluative and non-evaluative conditions will activate strong priming effects. The previous analyses showed that the non-evaluative condition did not produce significant effects when averaging across the entire sample. However, as I have previously explained, it could be the case that strong preferences are in fact hidden by collapsing across the sample as a whole. Hence, to test the viability of the two hypotheses, a response mapping score was calculated for each type of prime using the methods described earlier (i.e. the absolute value of the difference between the primes, for each individual participant). Hence, each participant received a score that reflected the degree to which he or she tended to associate the primes with a particular positive or negative response.

As predicted by the evaluative distinctiveness hypothesis, the non-evaluative condition produced significantly less response mapping ($M = .11$, $SD = .09$) than the evaluative priming condition ($M = .35$, $SD = .33$), $F(1, 37) = 10.10$, $p < .01$. That is, participants tended to show a much stronger preference for one prime over the other when the primes had clear evaluative implications, relative to when they did not. However, it should be noted that the index was indeed significantly different from zero in both the evaluative and non-evaluative conditions ($t = 5.38$ and 4.62, respectively, both $p < .001$). Hence, these data show that response mapping processes are significantly weakened (but not absent altogether) when the primes lack strong evaluative meaning. These data do not support the strong form of response mapping, which predicted that equal amounts of response mapping should occur across the evaluative and non-evaluative conditions. Instead, these data are more consistent with the evaluative distinctiveness view.

Analysis of Explicit Ratings

Immediately after completing the priming task, participants were asked to explicitly rate the images that they had seen in the task. Participants in the evaluative priming condition rated the angry faces much more negatively than the happy faces ($F(1,18) = 321.55$, $p < .001$). Participants in the non-evaluative priming condition did not rate the color and grayscale faces differently, $F(1,19) = 1.00$, $p > .30$. These results parallel the findings from the priming task. Hence, both explicit and implicit evaluations show strong differences between the happy and angry primes, and no difference between the neutral color and grayscale primes.

However, like the priming task, it is possible that participants did indeed perceive an evaluative difference between the color and grayscale faces, but that half preferred the color images whereas the rest preferred grayscale (note that this logic is identical to that used in the preceding section). To explore this possibility, I created an explicit response mapping index that was analogous to the response mapping scores used in the preceding analyses (i.e. the absolute value of the difference between participants' ratings of the two types of images). That index ranged from 0 (faces were liked equally) to 9 (faces were rated using opposite ends of the Likert scale). (The scale ranged from 0 to 9 because responses were made on a -4 to +4 Likert scale.) A UNIANOVA involving the resultant index revealed that, just like the priming data, the explicit response mapping score for the angry and happy faces ($M = 4.92$, $SD = 1.19$) was significantly greater than the response mapping score for the color and black-and-white faces ($M = .98$, $SD = .76$), $F(1,37) = 152.77$, $p < .001$. Also mirroring the priming data, both of these indices were significantly different from zero ($t = 17.93, 5.73$ for the evaluative and non-evaluative conditions, respectively, both $p < .001$).

Discussion

The strong form of the RM model states that participants will map any categorically distinct primes onto the available responses, even if the primes have no evaluative connotations whatsoever. Experiment 3 did not support that view. When the primes belonged to two neutral categories, little or no RM was observed. However, when the primes belonged to two evaluative categories, one extremely negative and the other extremely positive, the primes were clearly mapped onto the negative and positive responses, respectively. Instead of showing evidence for the strong form of the response mapping model, these data illustrate that one necessary condition for response mapping is that at least one of the primes must have clear evaluative connotations. While this may seem obvious in retrospect, this represents an important boundary condition of the RM model that was not evident prior to the present experiment.

At this point it is worth revisiting the contrast effects that were obtained in Experiments 1 and 2. In those experiments, only half of the primes were extreme, and the rest were neutral. Hence, the two types of primes were always evaluatively distinct, even though half of the primes had no particular evaluative connotations. The present experiment hypothesized that perhaps evaluative distinctiveness is a critical factor that is necessary for response mapping to take place. The data showed that this was indeed the case. Hence, one way to summarize what has been learned about the RM model thus far is that while the prime categories must be evaluatively distinct to activate RM, it is *not* critical that all primes have strong evaluative implications (instead, only *half* of the primes must have strong evaluations, and the rest can be neutral).

EXPERIMENT 4

Experiment 4 examines another boundary condition of the RM model. Specifically, Experiment 4 addresses whether RM depends on the presence of obvious, homogenous prime categories, or not. In all but one of my past RM experiments, the primes not only had clearly different evaluative connotations, they were also members of discrete classes of objects as well (e.g. animals vs. objects). One possibility is that participants' ability to map the primes onto one or the other response depends, in part, on their ability to easily detect *categorical* differences between the primes (in addition to detecting evaluative differences). If this is the case, then participants should show stronger priming effects when the primes consist of negative animals and neutral objects (for example), than if the negative and neutral images do not form any coherent internal category of objects. However, an alternative possibility is that response mapping may occur under any conditions in which the primes are evaluatively distinct, irrespective of whether each evaluative category has any internal coherence. In this case, participants' performance should be unaffected by the homogeneity of the prime categories, but instead should simply be contingent on the evaluative connotations of the primes.

In the present experiment, participants completed a priming task in which the primes were evenly divided between two evaluative groups (positive and neutral, or negative and neutral). In this respect, the present experiment was the same as block one in Experiments 1 and 2. However, the categorical homogeneity *within* each of those evaluative groups was experimentally manipulated (see Figure 7). Figure 7 displays the prime images that were presented in each of the four experimental conditions. The top half of Figure 7 displays the *homogenous category condition*, in which the extreme primes constituted an internally consistent semantic category (note that "homogeneous" refers to the *semantic* quality of the primes, not their evaluative connotations. All of the primes belonged to an evaluatively homogenous category.) In this condition, all of the extreme primes belonged to the category "animals". The neutral primes, however, did not constitute any particular category of objects, and included a diverse array of images, such as a hanging light bulb and a checkerboard pattern.

The bottom half of Figure 7 displays the *heterogeneous category condition*. In this condition, none of the primes created an internally consistent semantic category (although note that the primes still constituted two different *evaluative* categories). For example, the negative primes consisted of pictures of a tornado, skulls, a man with a gun, and a crying boy. None of these primes can be easily identified as belonging, a priori, to a single category of objects.

The critical question at hand is whether response mapping depends on participants' ability to categorize the primes, or not. If response mapping is dependent on ease of categorization, then priming effects should be bigger in the homogenous category condition (top half of Figure 7) than in the heterogeneous category condition (bottom half of Figure 7). On the other hand, if response mapping is not affected by the coherence of the prime categories, then response mapping should occur in both conditions, because both conditions present evaluatively distinct groups of objects.

Participants and Design

There were 75 undergraduate students who participated in this experiment in exchange for partial course credit. This experiment consisted of one within-subjects factor, pertaining to the nature of the two types of primes in the AMP task (neutral vs. unambiguous). There were also two between-subjects factors (see Figure 7), the first of which pertained to whether the unambiguous primes were positive or negative. The second between-subjects factor pertained to the homogeneity within the prime categories. Approximately half of the participants were assigned to the homogenous category condition, and the rest were assigned to the homogeneous category condition.

Procedure

Participants were randomly assigned to complete one of the four AMP tasks (see Figure 7). The AMP timing and configuration was exactly the same as the previous experiments. After completing the AMP task, participants were debriefed and dismissed.

Stimuli

The stimuli were selected from the IAPS collection (Lang, Bradley, & Cuthbert, 2008) on the basis of the normative data collected by Lang and his colleagues. In the homogenous category condition, the average affect rating of the unambiguous animal primes was 3.73 and 7.67, for the negative and positive animals, respectively. In the heterogeneous category condition, the average ratings of the negative and positive primes were 2.94 and 7.75, respectively. The neutral primes were the same across the homogenous and heterogeneous conditions, and these primes received an average rating of 5.09. An observant reader will note that the ratings for the unambiguous primes in the heterogeneous condition were slightly more extreme that those in the homogenous condition. Rather than being a fatal flaw, this actually provides a strong test of my hypothesis. If anything, one would expect response mapping to become stronger as the primes get more extreme, and so according to this view RM should be strongest in the heterogeneous condition. However, I predict that exactly the opposite will occur; namely, that little or no response mapping will occur in the heterogeneous condition, because the primes cannot be as easily categorized.

Summary of Predictions

Earlier, I considered the possibility that priming tasks reflect a categorization process. If this is true, then RM may depend on the ease with which participants can categorize the primes. When the primes belong to a homogenous semantic category, participants may have an easier time categorizing the primes compared to when the primes do not belong to a homogenous category. Hence, the typical contrast effects that we find for neutral primes may be blunted when the primes are heterogeneous, compared to when the primes are clear members of a homogenous category.

RESULTS

For each of the five prime categories (neutral, homogenous-negative, homogenous-positive, heterogeneous-negative, heterogeneous-positive), an index was created that represented the proportion of pleasant versus unpleasant responses. This index ranged from 0 (all unpleasant responses) to 1 (all pleasant responses). A score of 0.50 represents an equal number of unpleasant and pleasant responses.

A 2 (Prime: unambiguous vs. neutral) X 2 (Unambiguous prime: negative vs. positive) X 2 (Category condition: heterogeneous vs. homogenous) mixed-model ANOVA revealed a Prime X Context interaction ($F(1,71) = 11.10$, $p < .01$), however this effect was qualified by a 3-way Prime X Context X Cateogry condition interaction, $F(1,71) = 3.92$, $p = .05$ (see Figure 8). This 3-way interaction signified that, as predicted, response mapping occurred in the homogenous condition, but not in the heterogeneous condition. Simple effects tests confirmed that in the homogenous prime condition there was a significant Prime X Context interaction, $F(1,35) = 8.69$, $p < .01$. This interaction signified that the positive primes elicited relatively more pleasant responses as compared to the negative primes ($p < .05$). However, in the heterogeneous condition there were no significant effects at all (all $p > .10$). In fact, in the heterogeneous condition even the difference between the unambiguously negative and positive primes was not significant ($p > .3$).

Discussion

Experiment 4 demonstrated that evaluative priming effects are significantly influenced by the categorical coherence of the primes. According to most common models of evaluative priming (e.g. spreading activation, response competition), the effect for any given prime should be the result of its evaluative implications, not its categorical membership. However the present experiment shows that even an extreme prime—such as skulls and guns—will not produce any discernible priming effect when its category membership is unclear.

In my past experiments (both in this dissertation and those described in the introduction), the primes were easily sorted into two categories that differed in both their semantic meaning (animal/object) and evaluative implications (good/bad). In this respect, those past experiments were quite similar to the majority of past research in this area, which typically contrasts just two or three distinct types of primes (see Fazio et al. 1986; Fazio et al. 1995; Payne et al. 2005). However, in the present experiment, the primes always belonged to different evaluative categories, but they did not necessarily belong to clear-cut semantic categories. In the heterogeneous condition, the lack of categorical clarity seemed to obscure the fact that there were, in fact, two classes of *evaluative* stimuli. As a result, no evaluative priming effects occurred at all in this condition. Not only was there no contrast for the neutral prime, but there was also no discernible effect for the unambiguous primes, either. By contrast, in the homogenous condition the primes were easier to categorize, and as a result these primes produced significant priming effects.

The RM model provides a good framework with which to understand the present effects. According to that model, participants sort the primes and map them onto the response options. When participants have a difficult time sorting the primes, they will be unable to map those primes onto a response. Hence, participants do not respond to each individual prime, but instead respond to the prime categories as a whole. When the primes aren't easily categorized, even the strongest of primes—such as human skulls—do not exert a strong influence on participants' judgments.

The present findings may seem inconsistent with past research that has found strong priming in the absence of homogeneous prime categories. For example, Payne et al. (2005) used a multitude of heterogeneous prime stimuli, and yet still found significant priming effects for both negative and positive stimuli using the AMP. However, there is one key difference between our respective experiments that could potentially explain this discrepancy. In Payne et al.'s experiments, the primes were selected so that half of the primes were very positive, and half were very negative. In contrast, in the present research, extreme primes were always contrasted with *neutral* primes. Hence, one possible explanation is that in Payne et al.'s experiments, the extremity of the primes' evaluative connotations made the categorical boundaries between the primes very obvious, even in the absence of semantic homogeneity. However, in the present research, it is likely that participants needed additional "help" categorizing the primes (in the form of semantic homogeneity), because the prime connotations did not represent opposite ends of an evaluative spectrum. These postulates are clearly in need of formal testing. Nonetheless, the present experiment raises a number of important issues with regard to priming that would not have been apparent in the absence of the response mapping framework.

GENERAL DISCUSSION

The RM model proposes a new way of understanding priming tasks and automatic behaviors more generally. The model states that salient goals influence how people classify stimuli in their environment. Importantly, it does not particularly matter whether participants have the expressed intent to categorize those stimuli or not. Instead, the classification process can be an unintentional consequence of the salient goal at hand. For example, a starving person might classify everything in his environment as edible or inedible, even if he is not explicitly thinking about classification per se. Another important point is that the particular category label that ends up being associated with an object depends largely on the array of things that are in the person's environment. When presented with a poisonous-looking berry, for example, the starving man might end up classifying it as edible if there is nothing else available that is remotely edible, whereas it may be classified as potentially poisonous—and therefore inedible—if better food options can be found. Hence, our perception of objects is the result of both salient goals and perceptual context.

The RM model is a model of automatic judgment. Hence, the processes involved in deciding whether to eat a berry (or not) are very similar to the kinds of automatic judgments that are involved in priming tasks. In priming tasks, a goal is made salient by the experiment instructions. These instructions typically ask participants to sort the targets into two distinct categories (e.g. good-bad, gun-tool, etc.). This salient categorization goal is then applied to all of the stimuli in the task, including the primes (which participants are supposed to ignore). The net effect of these processes is that participants learn to associate the primes with the task responses. That is, participants "map" the primes onto the available responses.

Sometimes the connotations of the primes don't exactly fit the connotations of the responses in the priming task. Neutral primes, for example, don't fit the typical evaluative priming response labels, which are usually “positive” and “negative”. Hence, just as a potentially hazardous berry can be categorized as either edible or inedible depending on the availability of other foods, a neutral stimulus can be categorized as either positive or negative, depending on the particular task context. If a neutral prime is compared with something that is unambiguously bad, then the neutral stimuli will be categorized as good, and vice versa. The upshot of these considerations is that a neutral prime can elicit effects that appear to reflect strong prepotent evaluations, but that in fact reflect the way in which participants automatically categorize and map the primes. This conclusion has important implications for implicit attitude research, because it is generally assumed that priming tasks tap directly into participants’ automatized evaluations.

The main purpose of the present research was to resolve two important and as-yet unaddressed issues regarding the RM model. The first question, which was addressed by Experiments 1 and 2, concerned the nature of the association that forms between primes and responses. The second question, which was addressed by Experiments 3 and 4, concerned the boundary conditions of response mapping processes. Since these two questions are rather distinct, I will discuss the answers to, and implications of, these questions in separate sections, below.

Experiments 1 and 2: Implications, Limitations and Future Directions

Experiments 1 and 2 showed that simply performing a brief AMP priming task can cause lasting changes in the way participants' perceive the prime stimuli. When neutral primes had previously appeared along with extremely negative primes, they continued to elicit relatively favorable responses, both on a second priming task and on explicit scale ratings. In contrast, the neutral primes elicited relatively negative implicit and explicit scores when they had previously appeared alongside extremely positive stimuli.

Participants almost certainly did not begin the task with strong feelings toward the neutral primes. However, participants clearly finished each experiment with relatively positive or negative evaluative associations toward those prime stimuli. These results suggest that sequential priming tasks have the power to create new attitudes where none existed previously. This finding rests in contrast to the majority of the evaluative priming literature, which largely assumes that priming tasks measure attitudes, and do not create them.

The present data demonstrated that the AMP task can measure attitudes when the primes are extreme, but cause evaluative conditioning when the primes are not extreme. Hence, these data suggest that in cases where all of the primes are extreme, then evaluative conditioning will not occur. In light of this fact, one might make the faulty assumption that evaluative conditioning is not a concern for the vast majority of research on implicit attitudes. After all, researchers usually assume that the primes that they select have strong evaluative connotations. However, it is important to recognize that researchers often don't know, a priori, how participants feel toward the prime stimuli. For example, a researcher might assume that a particular prime is negative (perhaps because it elicits negative priming effects), when in fact it is neutral. Or, to make matters more complex, a particular prime might be negative for one person but neutral for another. This latter case may be especially true for complex social stimuli. The present research suggests that the outcome of a priming measure cannot be viewed as support for researchers' assumptions about the evaluative implications of their primes. If a particular prime stimulus happens to exhibit negative priming effects, this could be due to a) a preexisting negative association, or b) a *new* negative association that was learned in the task. For example, much research has shown that in a priming task involving Black and White primes, the Black primes typically evoke negative responses whereas the White primes evoke positive responses. The present research suggests that we cannot necessarily say whether the White primes are indeed positive (or the Black primes negative), because either of those effects could be due to pre-existing associations, or new associations that were learned as a result of the task itself.

Attitude Creation via Response Mapping

The effects observed in the present research can be understood as a consequence of response mapping. According to the response mapping model, participants mapped the neutral primes onto a pleasant or unpleasant response, not because they literally perceived those primes as pleasant or unpleasant, but rather because participants assigned the neutral primes to whichever response is not already “taken up” by the extreme primes (Scherer & Lambert, 2009). Hence, one important concept that runs throughout the present research is that the initial assignment of the neutral primes to the “pleasant” or “unpleasant” keys is not necessarily the result of an evaluative process. This was one of the major points made by Scherer & Lambert (2009).

The present research extends that earlier research by suggesting that one *consequence* of response mapping is that participants ended up perceiving the primes differently. Specifically, participants end up viewing the primes in accordance with the evaluative implications of the mapped response. Hence, response mapping can be likened to an evaluative conditioning process. For example, if participants make a “pleasant” key press every time a picture of a towel appears, they will come to view that towel in a positive light.

On the Automaticity of Evaluative Conditioning

Research on implicit attitudes has long made a distinction as to the various ways in which an evaluation can be “automatic” (Bargh, 1994). On one hand, something can be said to be “automatic” if it is unintentional. It appears that the evaluative conditioning described here is automatic in this sense. In spite of the fact that participants were supposed to ignore the primes, participants unintentionally categorized the primes and learned to associate those primes with a response. Therefore, in the present research it is reasonable to assume that an unintentional process led to evaluative conditioning.

However, automaticity can also refer to the extent to which participants are able to control their responses. On this point, the automaticity of evaluative conditioning is less clear. It is possible that participants could control their tendency to map the neutral primes, if one were to ask them to do so, but the present data do not speak to this issue. Moreover, automaticity can also refer to participants' awareness of the processes involved in the task. It is possible, for example, that participants were aware that they used the positive or negative response to refer to the neutral primes. However, it is equally likely that participants were mostly unaware of the mapping processes that occurred in the task. More research is clearly needed to clarify these issues.

Implications for the Literature on Contrast Effects

Contrast phenomena are familiar to most psychologists, in part because they are so ubiquitous. Contrast effects have been demonstrated in judgments that are quite diverse, from rating the importance of recycling, to judging the heaviness of lifted weights (Helson, 1947; Sherman, Ahlm, Berman & Lynn, 1978). Despite the ubiquity of contrast effects, an important and difficult issue has riddled the contrast literature for years. That issue pertains to whether contrast effects result from a change in people's perception of the stimuli, or instead are the result of a change in the way people use the rating scales (for a discussion of this issue, see Scherer & Lambert, 2009). To illustrate the latter case, imagine a situation in which participants are asked to rate the size of a mouse. When compared to an elephant, a mouse might be described as “very small”. But when compared to an amoeba, that same mouse might be described as “very large”. Importantly, using the different labels reflects a change in how “large” and “small” are defined, not in a change in how participants perceive the size of the mouse.

A definitive resolution of this issue was never reached. It was Sherman et al. (1978) who first suggested that both views might be correct. Those authors proposed that people may initially select a particular response because of the way that the context causes them to use the scale. However, this act of response selection may, in fact, form a stable association between the stimulus and response. The response association becomes, in effect, a bonafide change in perception. Using the above example, a mouse might be assigned the label “very large” when compared to an amoeba, even though participants know that the mouse is not any bigger than when it is compared to elephants. However, this act of assigning the “very large” label to the mouse will then cause the mouse to be perceived as being somewhat larger than would otherwise be the case.

The present research suggests that Sherman was probably correct. The contrast effects observed in the present experiments were almost certainly caused, at least initially, by a response mapping process that was not a reflection of participants' evaluations per se. However, the act of mapping the prime stimuli onto an evaluative response resulted in a lasting change in the way that participants responded to the stimuli. Hence, the act of choosing a response to refer to a given stimulus resulted in a new attitude.

Directions for Future Research

The present research suggests numerous avenues for future research. I have already suggested that it will be important to explore the various ways in which the present effects could be labeled "automatic". Evaluative conditioning in the AMP is probably unintentional, but it remains to be determined whether it is also uncontrollable and inaccessible to conscious thought. Another important avenue for future research is to determine the longevity of the evaluative conditioning effects. It could be that the association between prime and response lasts only a few minutes, or alternatively, it could potentially last for hours or even days. Additional research is needed to determine which of these possibilities is correct.

Additionally, the present research focused on primes that were not social in nature. Future research should apply these methods to a wide range of social stimuli, in various combinations, to determine what kinds of social stimuli are most likely to be subject to evaluative conditioning effects in the AMP. For example, research has shown that elderly primes generally elicit negative responses when they are placed in a task alongside extremely positive primes (e.g. Scherer & Lambert, 2009, Experiment 6). As I have already discussed, this sort of finding could potentially reflect preexisting negative attitudes, or alternatively it could reflect new negative attitudes toward the elderly that were generated by the task itself, and which last beyond the end of the experiment. It will be important for future research to determine whether priming tasks have the power to cause long-lasting changes in attitudes toward social stimuli.

Experiments 3 and 4: Implications, Limitations and Future Directions

Experiments 3 and 4 clarified two important boundary conditions of the RM model. Experiment 3 demonstrated that the response mapping process is dependent on the presence of clear, evaluative distinctiveness between the primes in the task. That is, response mapping will be strong when the two classes of primes are evaluatively distinct (e.g. negative and positive, or negative and neutral), but will be weak or not occur at all when the two classes of primes are categorically distinct but evaluatively neutral (i.e. neutral color and neutral grayscale). This represents a boundary condition to the RM model. Specifically, one criterion for response mapping is that the primes must be *evaluatively* distinct.

Experiment 4 demonstrated that even when the primes meet the evaluative distinctiveness criterion, the primes must also be categorically homogenous in order for response mapping to occur. Even though the primes in that experiment always represented two distinct *evaluative* categories (e.g. negative and neutral, or positive and neutral), participants nonetheless failed to map the primes when the *semantic* category boundaries were unclear. This finding demonstrates that priming effects depend, in part, on a categorization process. When participants cannot easily sort the primes into two distinct categories, then they cannot map the primes onto two distinct responses.

Addressing some alternative explanations of Experiment 4

One of the most surprising findings of the present research occurred in Experiment 4, which showed that even the most negative (e.g. skulls) and positive (e.g. ice cream) primes do not produce priming effects when the categorical boundaries between the primes are unclear. This finding is surprising because it is somewhat rare to find circumstances under which these sorts of extreme stimuli do *not* produce priming effects. Hence, before discussing the implications of these findings for the AMP, it is critical to examine some alternative explanations for this counterintuitive result.

First, it is important to point out that the lack of significant priming in the case of the heterogeneous primes could not have been due to lack of strong attitudes toward these stimuli. In fact, the heterogeneous prime stimuli were even more evaluatively extreme than the primes in the homogenous condition. For example, the normed ratings of the heterogeneous negative primes (i.e. skulls, tornado, gun and crying boy) were more extremely negative than the ratings of the homogeneous negative primes (i.e. snarling dog, snake, cockroach and shark). Moreover, the heterogeneous positive primes were more extremely positive than the homogeneous positive primes.

Second, while it is technically possible that the lack of effects for the extreme primes could have been due to random sampling error, this possibility is also highly unlikely. For one thing, AMP effects tend to be quite reliable. This suggests that null effects in the AMP are probably quite reliable as well.

Third, there could have been something about the heterogeneous primes themselves that resulted in null effects. However, neither the heterogeneous positive nor heterogeneous negative primes produced any significant effects. From this perspective, it seems highly unlikely that some unknown variable, apart from prime heterogeneity, could have nullified the effects of both the positive and negative heterogeneous primes. Hence, these three factors (lack of attitudes, lack of reliability, or third variable issues) cannot easily account for the surprising fact that extreme primes produced no priming when they belong to heterogeneous categories.

Implications for the processes underlying the AMP

The AMP is thought to have its effect through a process of attitude misattribution (Payne et al., 2005). That is, on each trial the participant views the prime, has an evaluative reaction, and then misattributes that evaluative reaction to the Chinese character target. According to this view, AMP effects should be determined by participants' feelings toward each individual prime, and not by their ability to perceive the categorical relationships *among* the primes. Hence, a misattribution conceptualization of the AMP is difficult to reconcile with the present findings. From the perspective of the current article, the AMP task might be better understood as a *categorical* misattribution procedure, insofar as participants misattribute the evaluative implications of the prime *category* to the targets.

Although the RM model can account for the observed effects more easily than a misattribution model, more research is needed to substantiate these claims. For example, the RM model predicts that the heterogeneous primes will produce strong priming effects if participants are somehow able to categorize them (for example, if participants learn the categories prior to the priming task). In future experiments, it might be instructive to encourage participants to think of the extreme stimuli as members of ad hoc categories (e.g. "terrible events", or "things that make you feel good"). Under these circumstances, one would expect the extreme stimuli to produce strong priming effects in spite of the fact that they do not belong, a priori, to a clear semantic category.

Applications and Limitations of the Response Mapping Model

There are a number of priming paradigms that can currently be used to assess implicit attitudes and associations. These include lexical decision tasks (Neely, 1991), the AMP (Payne et al., 2005), the IAT (Greenwald et al. 1998), evaluative priming (Fazio et al. 1986), the Go/No-go task (GNAT; Nosek & Banaji, 2001), the affective Simon (De Houwer, 2003), weapon identification tasks (Correll, Park, Judd & Wittenbrink, 2002; Payne, 2001), and many others. One important question is how far reaching the implications of the RM model actually are; that is, whether the model can explain a large portion, or only a small slice, of the existing implicit attitude and priming literature. As it is currently understood, the RM model applies to any priming paradigm that fits the following criteria:

1) Participants are asked to make categorical responses (e.g. “good” and “bad”, or “fruit” and “vegetable”).

2) The prime stimuli consist of two or more categories that can be construed to correspond to the response labels being used in the relevant task. For example, negative and neutral primes can be construed to correspond to negative and positive responses, by treating the neutral primes as “positive” in that particular task context.

Importantly, the above criteria describe a large portion of priming research in social psychology, including most research involving the AMP, the IAT, evaluative priming, and other similar tasks, such as weapon paradigms (Fazio et al. 1986; Greenwald et al. 1998; Payne et al. 2005; Payne 2001). In those tasks, participants are asked to categorize target stimuli using two response categories (criteria one). As for the primes, researchers usually select stimuli that are extreme or highly categorizable (criteria two). Under these conditions, the assumptions of the RM model should apply.

A case in point is the classic experiment by Fazio, Jackson, Dunton & Williams (1995), on which much of the subsequent literature in this area is based. In those experiments, automatic reactions to black and white faces were measured. On the critical trials, the primes were evenly divided between Black and White faces, and participants were asked to sort target words into “good” and “bad” categories. Given these task parameters, it is easy to imagine how participants might have very easily applied the target categorization task to the primes, since both the primes and targets were evenly divided into two obvious categories.

The RM model could also conceivably apply to the IAT, although no research has examined this issue. In the IAT, the response mapping is, in fact, an explicit part of the experimental paradigm. In fact, “response mapping” is often the exact phrase that is used to describe IAT effects. For example, in Experiment 3 of Greenwald et al.’s classic article (1998), participants were asked to sort words according to two categorical dimensions: race (i.e. Black vs. White), and valence (i.e. pleasant vs. unpleasant). Participants were faster when “Black” and “unpleasant” responses were made with the same key press, relative to when “Black” and “pleasant” responses were made with the same key press. In this case, the responses to Black targets were *literally* mapped onto the “pleasant” or “unpleasant” response, as a result of explicit task instructions. Hence, there is reason to presume that the sorts of effects predicted by the RM model should hold true for the IAT.

There are a number of instances in which the RM model will almost certainly *not* apply, however. In those cases, other processes (such as spreading activation or response competition) can better explain observed priming effects. For example, the RM model is not expected to apply in cases where the task judgment is independent from the effect of the primes, such as when the judgment is “word” vs. “non-word”, and the priming effect occurs only within the responses to real words (e.g. Gaertner & McLaughlin, 1983; Wittenbrink, Judd & Park, 1997). In other priming paradigms, the dependent variable is an impression, rather than a target identification, and the RM model does not apply in these cases either (e.g. Devine, 1989). Furthermore, the RM model is somewhat difficult to apply to paradigms such as the GNAT (Nosek & Banaji, 2001), and the affective Simon (De Houwer, 2003). In sum, the processes described in the present article can be applied to many evaluative priming, AMP and IAT tasks, but are in no way capable of accounting for all implicit attitude effects.

On the RM Model and the moderating effect of evaluative strength

There is one other potential limitation of the RM model that warrants additional comment. The RM model may, at first blush, seem difficult to reconcile with some past research showing that priming effects are moderated by evaluative strength. For example, in Fazio et al.'s (1986) seminal experiments, the authors claimed to have found strong effects for the subset of primes that were known to have strong attitude associations, but weak effects for primes known to have weak attitude associations. The RM model cannot explain these effects, because the model assumes that primes will be responded to categorically. That is, the positive primes should be treated as being equal members of a "positive" category, regardless of how favorable any particular prime happens to be.

Yet upon a closer look at Fazio et al.'s experiments, it turns out that the majority of the data in fact support the RM model. In two out of three experiments, the weak primes did not show weak priming effects; instead, the weak primes showed *no effects at all* (moreover, when the weak primes finally did show priming, in Experiment 3, they did so only in one of the two SOA conditions). That is, weak positive and weak negative primes generally showed priming effects that were not different from each other, and those effects were intermediate between the extreme primes. In this sense, the weak primes exhibited effects that were exactly what one would expect from a set of control stimuli.

The RM model can explain these effects by assuming that the strong primes were associated with a response, and the weak primes were not. That is, participants categorized the primes as being positive, negative and neutral. The negative primes were mapped onto the negative response, the positive primes were mapped onto the positive response, and the rest of the primes were not mapped at all, because both of the available responses had already been "taken up" by the extreme primes. One implication of this observation about Fazio et al.'s classic research is that it is not currently clear whether priming tasks can, in fact, show graded effects of evaluative strength beyond more than three levels (positive, negative, neutral) of primes.

Directions for future research

The present experiments suggest numerous avenues for future research. I have already mentioned that an additional experiment is needed to lend additional support to the conclusions of Experiment 4. Furthermore, it might be instructive to employ the data analysis strategy from Experiment 3 to experiments that find null effects, especially those that involve social stimuli. For example, social stimuli can be perceived in many different ways, depending on the context and who is doing the perceiving. Some people may find elderly faces to be quite comforting, whereas others may find such faces ugly or negative. If a researcher is interested in implicit attitudes towards the elderly, and finds null effects using a priming task, it could be that the participants are evenly distributed between liking for the elderly and dislike. The analysis strategy employed here could potentially lend insight into such situations.

Finally, while the present experiments have demonstrated two boundary conditions to the RM model, there are almost certainly more. Further research is needed to discover what these boundary conditions are, and what the implications are for the interpretation of priming task results. Such research will almost certainly lend further insight into the complex nature of priming tasks and automatic behaviors more generally.

REFERENCES

Balota, D.A., Paul, S., Spieler, D. (1999). Attentional control of lexical processing pathways during word recognition and reading. In S. Garrod & M. Pickering (Eds). Language Processing (pp. 15-57). Psychology Press Ltd.

Banse, R. (2001). Affective priming with liked and disliked persons: Prime visibility determines congruency and incongruency effects. *Cognition and Emotion, 15,* 501-520.

Bargh, J. A. & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54,* 462-479.

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

Barsalou, L. W. (1983). Ad Hoc Categories. *Memory & Cognition, 11,* 211-227.

Barsalou, L. W. (1991). Deriving Categories to Achieve Goals. In Bower, G. (Ed.) *The Psychology of Learning and Motivation.* Academic Press, NY.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C. & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5,* 323-370.

Biernat, M. (2005). *Standards and expectancies: contrast and assimilation in judgments of self and others.* New York: Routledge.

Biernat, M. & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology, 66,* 5-20.

Blair, I. V., Ma, J. E. & Lenton, A. P. (2001). Imagining stereotype away: the moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology, 81*, 828-841.

Brickman, P., Coates, D. & Janoff-Bulman, R. (1978). Lottery winners and accident victims: *Is happiness relative? Journal of Personality and Social Psychology, 36,* 917-927.

Correll, 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.

Dasgupta, N. & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admires and disliked individuals. *Journal of Personality and Social Psychology, 81,* 800-814.

Devine, P. (2001). Implicit prejudice and stereotyping: How automatic are they? Introduction to the special section. *Journal of Personality and Social Psychology, 81*(5), 757-759.

Deutsch, R. & Gawronski, B. (in press). When the method makes a difference: Antagonistic effects on "automatic evaluations" as a function of task characteristics of the measure. *Journal of Experimental Social Psychology*.

De Houwer, J. (2006). What are implicit measures and why are we using them. In R. W. Wiers & A. W. Stacy (Eds.), *The handbook of implicit cognition and addiction* (pp. 11-28). Thousand Oaks, CA: Sage Publishers.

Eder, A., & Klauer, K. C. (2007). Common valence coding in action and evaluation: Affective blindness toward response-compatible stimuli. *Cognition & Emotion*, *21*, 1297-1322.

Eimer, M., Schlaghecken, F. (1998). Effects of masked stimuli on motor activation: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1737-1747

Eiser, J. R. (1986). *Social psychology: Attitudes, cognition, and social behavior*. Cambridge,UK: Cambridge University Press.

Eiser, J. R. (1990). *Social Judgment.* Pacific Grove, CA: Brooks/ColePublishing Co.

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion,* 15, 115-141.

Fazio, R. H., Jackson, J. R., Dunton, B. C. & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69,* 1013-1027.

Ferguson, M. J. & Bargh, J. A. (2003). The constructive nature of automatic evaluation. In J. Musch & K. C. Klauer (Eds.), The psychology of evaluation: Affective processes in cognition and emotion (pp.169-188). NJ: Erlbaum.

Gawronski, B, Deutsch, R., & Seidel, O. (2005). Contextual influences on implicit evaluative: A test of additive versus contrastive effects of evaluative context stimuli in affective priming. *Personality and Social Psychology Bulletin, 31*, 1226-1236.

Glaser, J. & Banaji, M. R. (1999). When fair is foul and foul is fair: Reverse priming in automatic evaluation. *Journal of Personality and Social Psychology, 77,* 669-687.

Greenwald, A. G., McGhee, D. E. & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology, 74*, 1464-1480.

Judd, C., Blair, I., & Chapleau, K. (2004). Automatic stereotypes vs. automatic prejudice: Sorting out the possibilities in the Payne (2001) weapon paradigm. *Journal of Experimental Social Psychology*, *40*(1), 75-81.

Karpinski, A. & Steinman, R. B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91*,16-32.

Klauer, K. C. & Musch, J. (2003). Affective priming: findings and theories. In J.

Musch, & K.C. Klauer (Eds.), *The Psychology of Evaluation,* (pp. 7-36). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Klauer, K. C., Robnagel, C. & Musch, J. (1997). List-context effects in evaluative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 246-255.

Klauer, K. C., Teige-Mocigembad, S. & Spruyt, S. (2009). Contrast effects in spontaneous evaluations: a psychophysical account. *Journal of Personality and Social Psychology, 96*, 265-287.

Klinger, M. R., Burton, P. C., & Pitts, G. S. (2000). Mechanisms of unconscious priming: I. Response competition not spreading activation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26,* 441-455.

Kosslyn, S. M. (1975). Information Representation in visual images. *Cognitive Psychology*, *7,* 341-370.

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

Larsen, R. J., Mercer, K. A. & Balota, D. A. (2006). Lexical characteristics of words used in emotional Stroop experiments. *Emotion*, *6,* 62-72.

Livingston, R. W., & Brewer, M. B. (2002). What are we really priming? Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology, 82,* 5–18.

Maier, M. A., Berner, M. P. & Pekrun, R. (2003). Directionality of affective priming: Effects of trait anxiety and activation level. *Experimental Psychology, 50,* 116-123.

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

Ostrom, T. M. (1970). Perspective as determinant of attitude change. *Journal of Experimental Social Psychology, 6,* 280-292.

Payne, B. (2005). Conceptualizing control in social cognition: How executive functioning modulates the expression of automatic stereotyping. *Journal of Personality and Social Psychology, 89*(4), 488-503.

Payne, B. K., Cheng, C. M., Govorum, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89*, 277-293.

Payne, B. K., Jacoby, L. L. & Lambert, A. L. (2005). Attitudes as accessibility bias: Dissociating automatic and controlled components. In R. Hassin, Bargh, J. & Uleman, (Eds.), *The New Unconscious.* Oxford.

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review, 72,*407-418.

Petty, R. E. & Cacioppo, J. T. (1981). Attitudes and persuasion: Classic and contemporary approaches. Dubuque, IA: William Brown Publishers.

Scherer, L.D. & Lambert, A. J. (2009). Contrast effects in priming paradigms: Implications for theory and research on implicit attitudes. *Journal of Personality and Social Psychology, 97,* 383-403.

Schwarz, N., Groves, R. M. & Schuman, H. (1998). Survey methods. In Gilbert, D. T., Fiske, S. T. & Lindzey, G. (Eds.) *The handbook of social psychology* (pp. 143-179). New York, NY, McGraw-Hill.

Schwarz, N., & Bless, H. (1992). Constructing reality and its alternatives: Assimilation and contrast effects in social judgment. In L.L. Martin & A. Tesser (Eds.), The construction of social judgment (pp. 217-245). Hillsdale, NJ: Erlbaum.

Sherif, M. & Hovland, C. I. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change.* New Haven, CT: Yale University Press.

Sherman, S.J., Ahlm, K., Berman, L., & Lynn, S. (1978). Contrast effects and the relationship to subsequent behavior. Journal of Experimental Social Psychology, 14, 340-350.

Simpson, D. D. & Ostrom, T. M. (1976). Contrast effects in impression formation. *Journal of Personality and Social Psychology*, *34,* 625-629.

Skowronski, J. J. & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin, 105*, 131-142.

Spruyt, A., Hermans, D., De Houwer, J., Vandromme, J. & Eelen, P. (2007). On the nature of the affective priming effect: Effects of stimulus onset asynchrony and congruency proportion in naming and evaluative categorization. *Memory and Cognition, 35,* 95-106

Steele, C. M. & Ostrom, T. M. (1974). Perspective-mediated attitude change: When is indirect persuasion more effective than direct persuasion? *Journal of Personality and Social Psychology, 29,* 737-741.

Upshaw, H. S. (1969). Stimulus range and the judgmental unit. *Journal of Experimental Social Psychology, 5*, 1-11.

Volkmann, J. (1951). Scales of judgment and their implications for social psychology. In J. H.Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads* (pp. 273-294). New York: Harper.

Wedell, D. H., Hicklin, S. K. & Smarandescu, L. O. (2007). Contrasting models of assimilation and contrast. In Stapel, D. A. & Suls, J. (Eds.), *Assimilation and contrast in social psychology* (pp. 45-74). New York: Psychology Press.

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

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.

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.

Footnotes.

1. For example, a spreading activation account of these effects would have to assume that the accessibility of a cockroach's "badness" is heightened when participants are making evaluative responses, whereas accessibility of a cockroach's "noun-ness" is heightened when participants are making adjective/noun responses (Klauer & Musch, 2003).

Table 1.

	Negative Context	Positive Context	Difference
Valenced primes	.36	.63	-.27***
Neutral primes	.70	.53	.17***

Replication of contrast effect, Experiment 1. Higher numbers indicate more pleasant responses. Also note that in the Negative Context, valenced primes are threatening animals. In the Positive Context, valenced primes are unthreatening (baby) animals.
**** $p < .001$*

Table 2.

	Negative Context	Positive Context	Difference
Valenced primes	.42	.62	-.20**
Neutral primes	.68	.52	.16*

Replication of contrast effect, Experiment 2. Higher numbers indicate more pleasant responses. Also note that in the Negative Context, valenced primes are threatening animals. In the Positive Context, valenced primes are unthreatening (baby) animals.

* $p < .05$

**$p < .01$

Figure 1a.

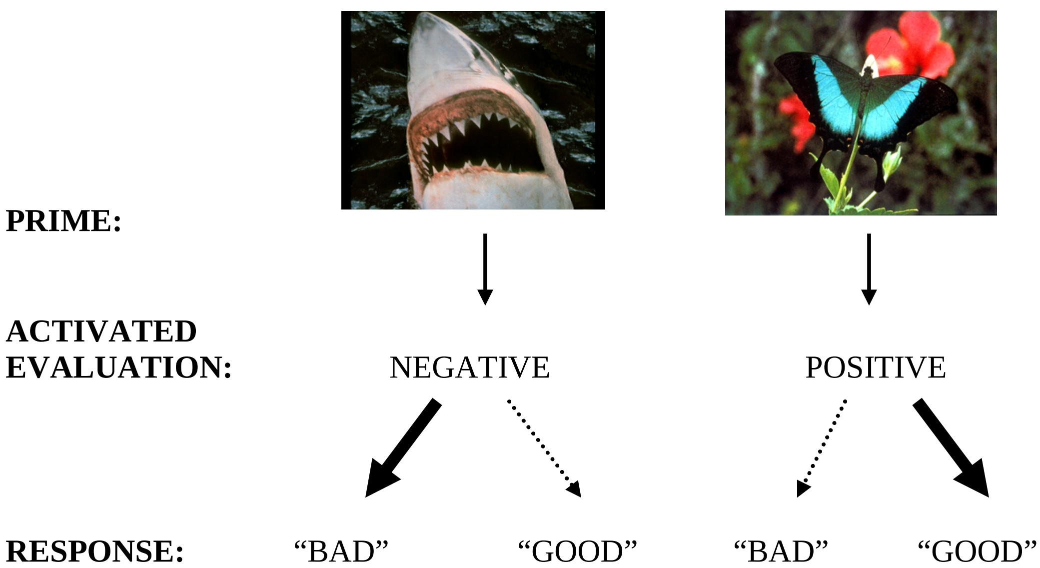


Figure 1b.

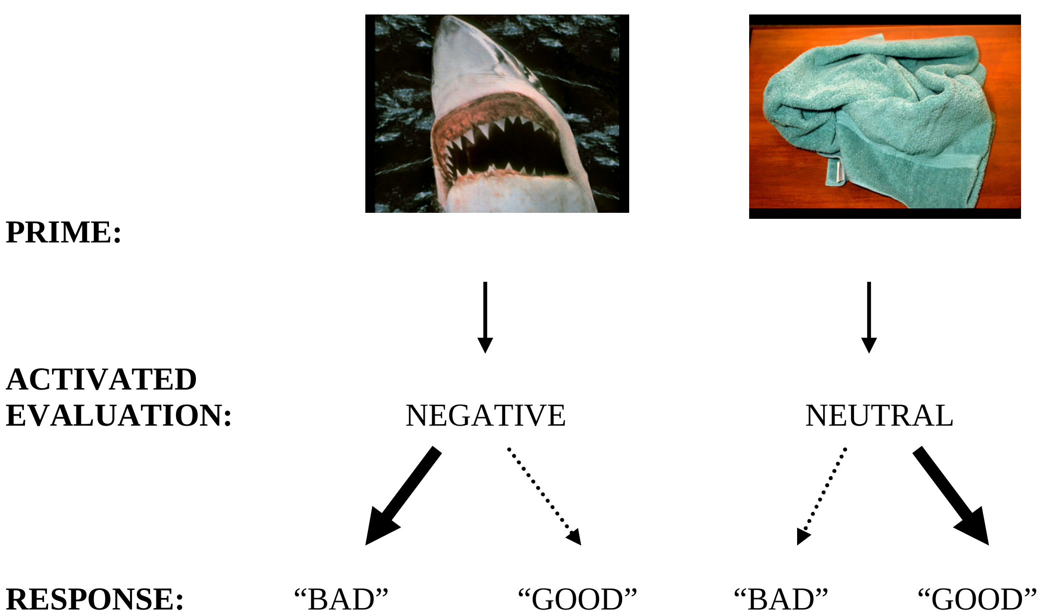


Figure 1c.

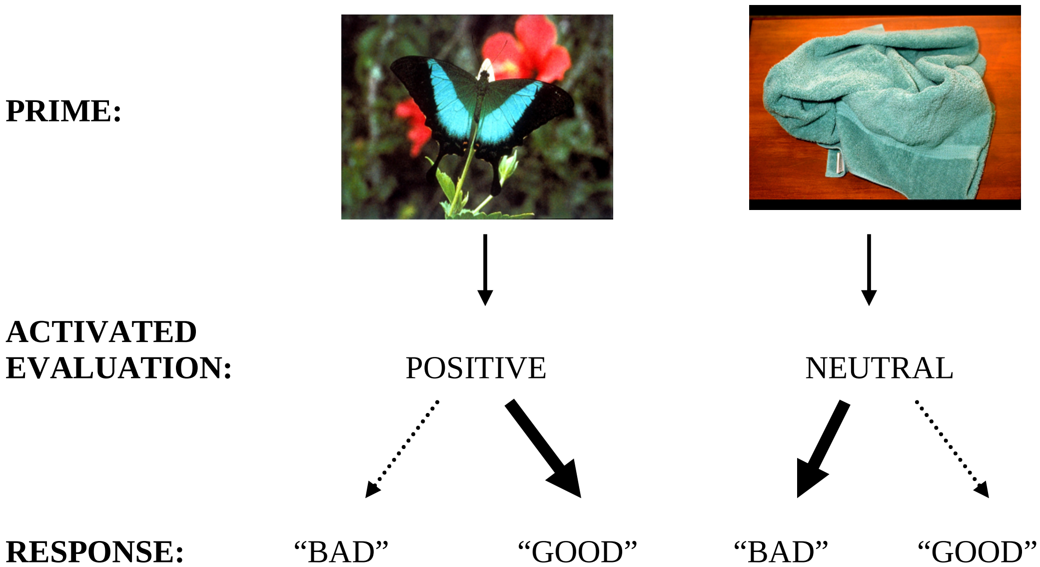

PRIME:
ACTIVATED EVALUATION:
POSITIVE
NEUTRAL
RESPONSE:
“BAD”
“GOOD”
“BAD”
“GOOD”

Figure 2.

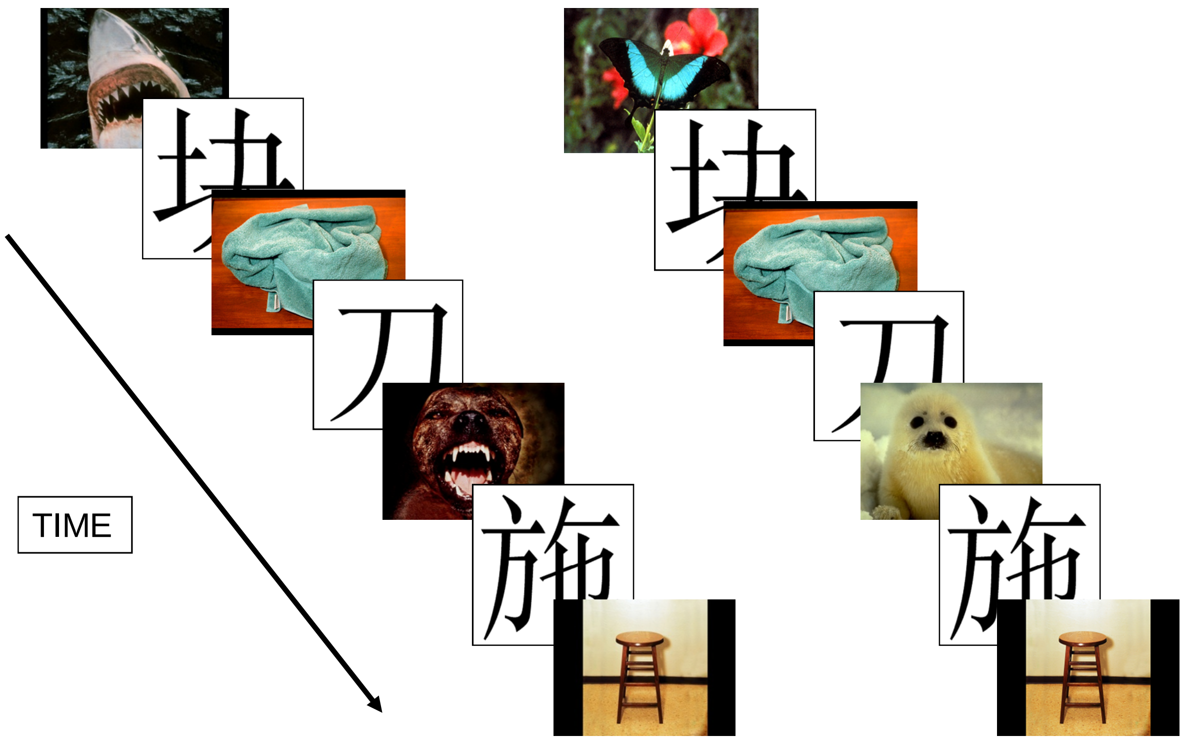


Figure 3a

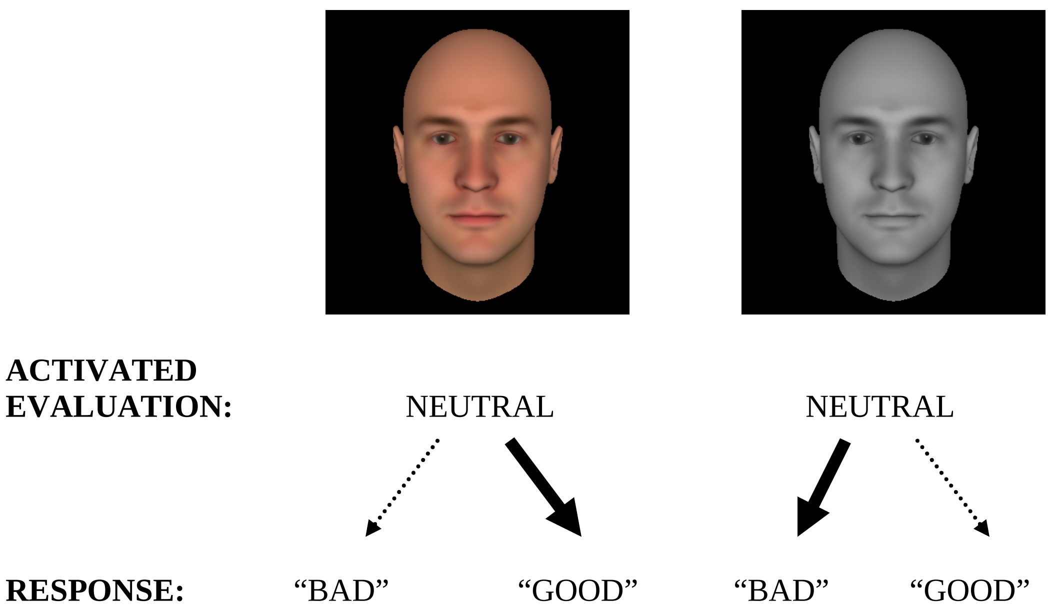


Figure 3b

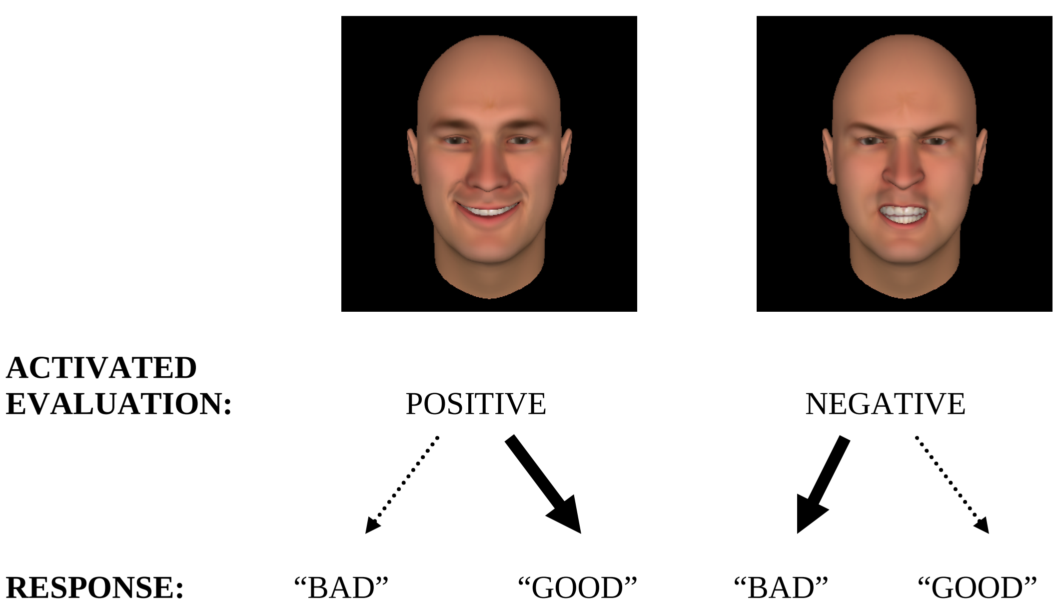
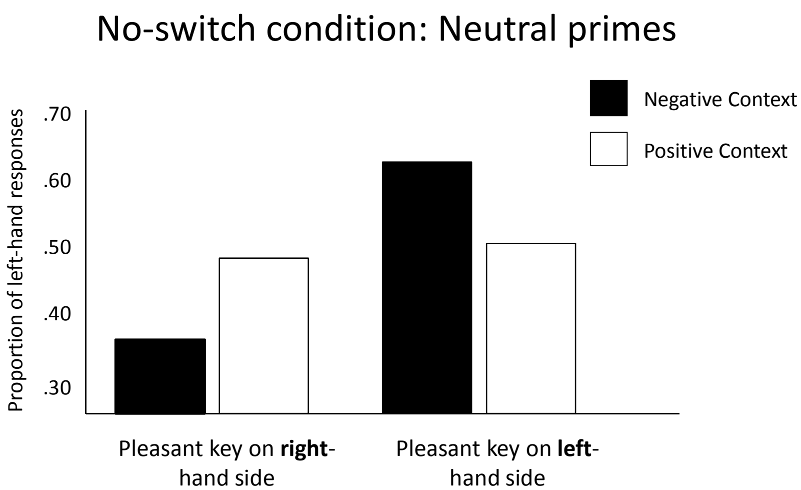


Figure 4.


No-switch condition: Neutral primes
Negative Context
Positive Context
Proportion of left-hand responses
.70
.60
.50
.40
.30
Pleasant key on **right**-hand side
Pleasant key on **left**-hand side

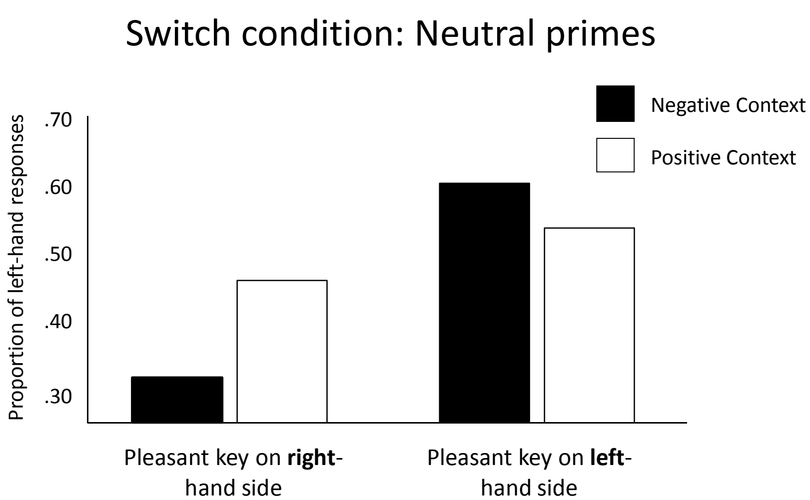
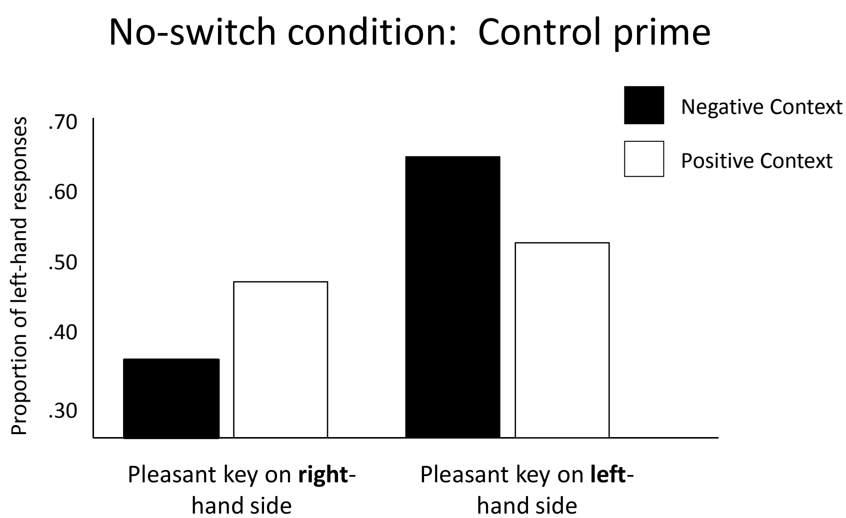

Switch condition: Neutral primes
Negative Context
Positive Context
Proportion of left-hand responses
.70
.60
.50
.40
.30
Pleasant key on **right**-hand side
Pleasant key on **left**-hand side

Figure 5.


No-switch condition: Control prime
Negative Context
Positive Context
Proportion of left-hand responses
.70
.60
.50
.40
.30
Pleasant key on right-hand side
Pleasant key on left-hand side

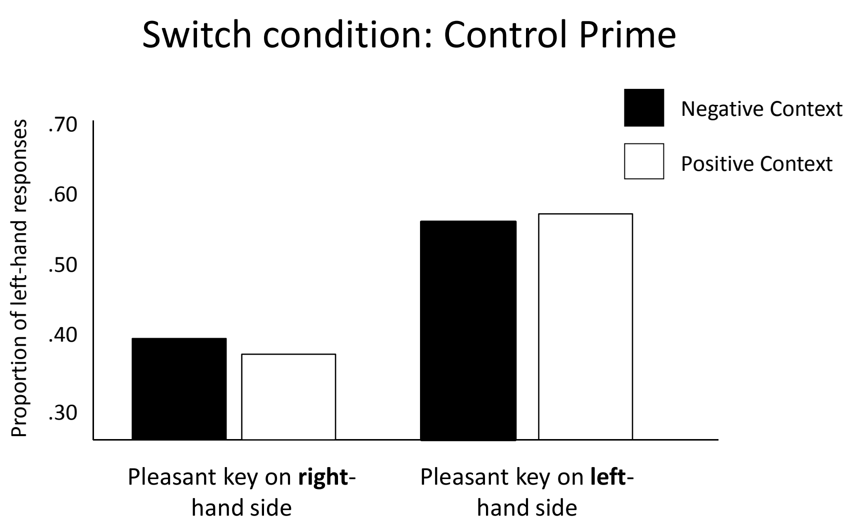

Switch condition: Control Prime
Negative Context
Positive Context
Proportion of left-hand responses
.70
.60
.50
.40
.30
Pleasant key on right-hand side
Pleasant key on left-hand side

Figure 6.

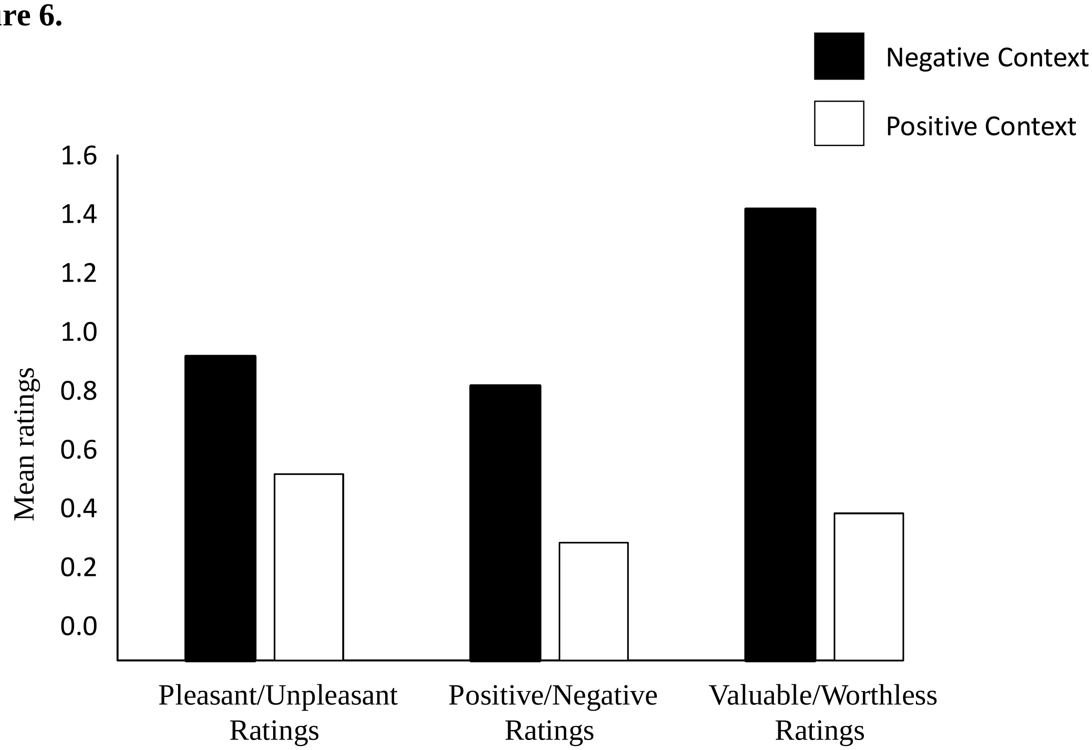
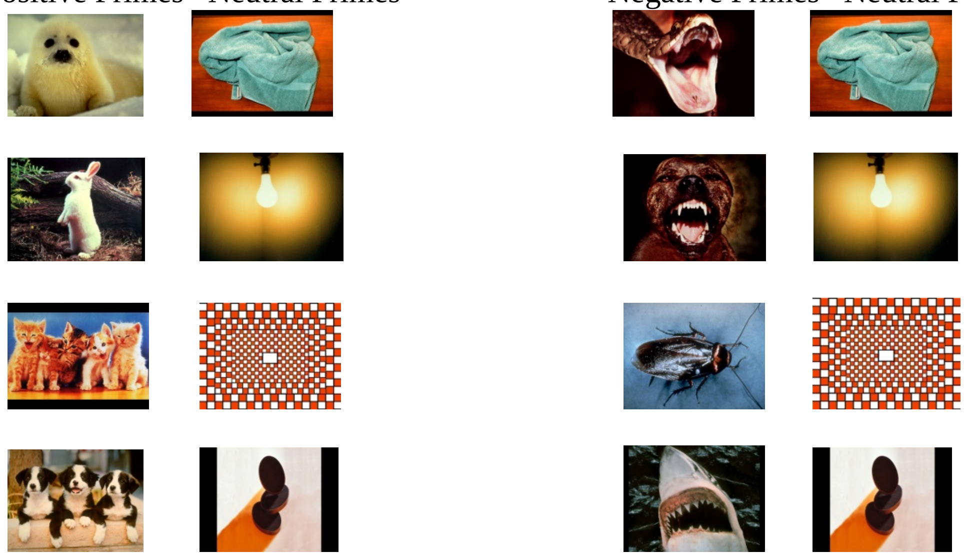

Negative Context
Positive Context
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Mean ratings
Pleasant/Unpleasant Ratings
Positive/Negative Ratings
Valuable/Worthless Ratings

Figure 7. HOMOGENEOUS CATEGORY CONDITION

Positive Context — Positive Primes | Neutral Primes

Negative Context — Negative Primes | Neutral Primes



HETEROGENEOUS CATEGORY CONDITION

Positive Context — Positive Primes | Neutral Primes

Negative Context — Negative Primes | Neutral Primes

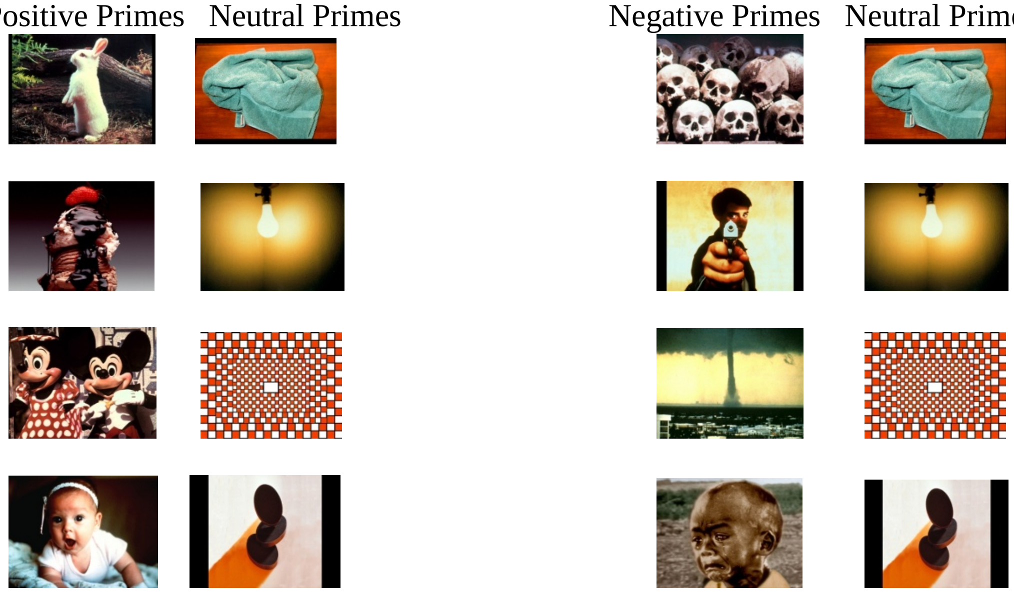
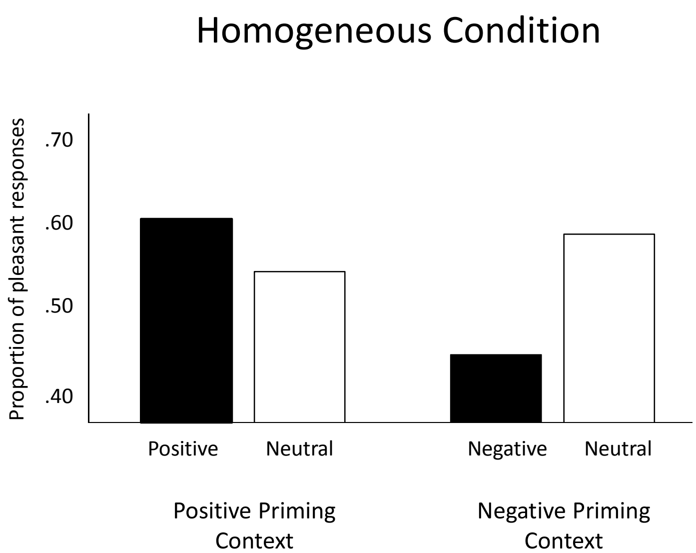
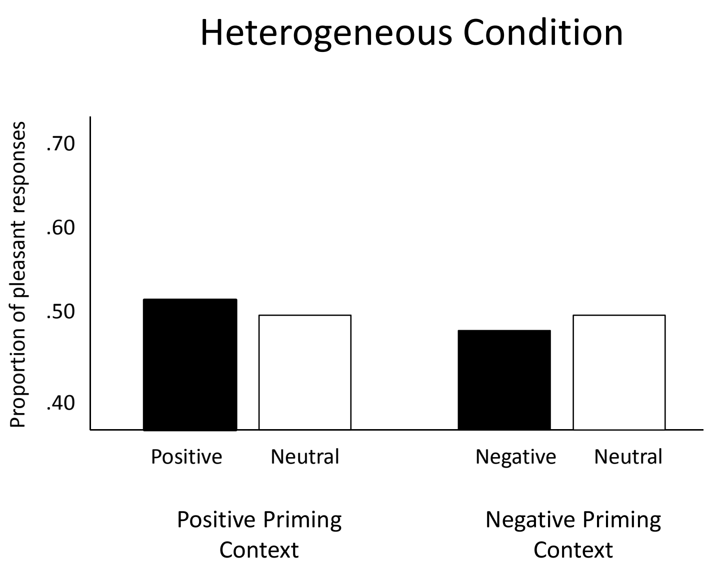


Figure 8.


Homogeneous Condition
Proportion of pleasant responses
.70
.60
.50
.40
Positive
Neutral
Negative
Neutral
Positive Priming Context
Negative Priming Context


Heterogeneous Condition
Proportion of pleasant responses
.70
.60
.50
.40
Positive
Neutral
Negative
Neutral
Positive Priming Context
Negative Priming Context

APPENDIX A

The Evaluative Priming Task

Researchers often vary the exact configuration of evaluative priming tasks, and so the following description is meant to be somewhat general (e.g. I will not specify exact presentation times, ISIs, masking procedures, etc. that were used in Scherer and Lambert's research). In the evaluative priming task, participants are presented with a series of rapid trials, each of which consists of a prime followed by a target. The prime is often a picture, although it can also be a word (in the experiments that will be described, the prime is always a picture). The target is always a positive or negative word, such as "wonderful" or "horrible". On each trial, the prime picture appears very briefly (approximately 100 milliseconds), and is immediately followed by the target word. Participants are instructed to ignore the prime stimulus and respond only to the target word by indicating whether the target is positive or negative, using one of two appropriately labeled keyboard keys.

Even though participants are told to ignore the primes, these stimuli systematically bias participants' responses to the targets. For example, an evaluatively negative prime will make participants faster and more accurate in responding to negative targets. In contrast, the negative prime will make participants slower and less accurate when responding to a subsequent positive target. As a result, the researcher can infer participants' evaluations of the prime by observing how that prime influences participants' responses to the targets. This task is considered indirect, or "implicit", because attitudes towards the primes are assessed without ever asking participants how they feel about the prime stimuli directly. Also, since participants are specifically instructed to ignore the primes, any influence of the prime on responses is usually assumed to be unintentional (assuming that participants are following instructions) and automatic (Fazio et al., 1995).

In Scherer and Lambert's experiments (2009), the primary dependent variable was participant accuracy rates, rather than reaction times. Typically, accuracy data in evaluative priming are interpreted as follows: Greater errors on positive than negative words following a particular prime indicate negative bias, whereas the reverse indicates positive bias. The logic behind this interpretation is simple: A negative prime will activate the negative response, making participants more prone to use this response when the word is actually positive. Likewise, a positive prime will activate the positive response, with the end result being that participants will often use the positive response when the word is actually negative.

One advantage of using errors as a dependent variable is that errors represent a lapse in cognitive control (Jacoby, 1991; Payne, 2001). That is, participants are instructed to identify words accurately, and an error represents an instance in which they fail to accomplish this task. Hence, errors are thought to be independent—at least, more so than reaction times—of the sorts of controlled behaviors that participants might ordinarily exert when answering explicit questionnaires, such as self-monitoring and presentational biases (Fazio et al., 1995). Of course, it is important to keep in mind that errors are not a pure measure of participants' automatized reactions to the prime; that is, tasks are not process pure (Jacoby, 1991). Errors can also be caused by factors that are independent of the prime's influence, such as arbitrary distractions, or task set expectations (for example, a participant may expect that the next trial will contain a negative word because the previous five trials were positive words).

The Attitude Misattribution Paradigm Task

The AMP task (Payne et al., 2005) is similar in many ways to the evaluative priming task. It consists of rapid trials in which a prime is followed by a target. Participants are asked to ignore the prime and make binary positive/negative judgments to the targets. However, unlike the evaluative priming task, in the AMP the target is always a Chinese character. The participant's task is to identify the characters as either more or less pleasant than average by pressing one of two appropriately labeled keys on a keyboard.

Participants typically cannot read these characters (and those who can read the characters are removed from analyses), and the characters are presented extremely briefly (typically less than 100ms). In addition, participants are told to respond very quickly, based on their "gut" response to the target. As a result, participants' responses are not based on the objective attributes of the character itself. Instead, participants tend to misattribute their evaluations of the primes as evaluations of the targets, because they have nothing else on which to base their judgments. For example, a negative prime tends to increase the likelihood that the following target will be rated as "unpleasant", whereas the reverse is true for positive primes. One important assumption of the AMP is that this process occurs outside of participants' conscious awareness. That is, participants use the primes as a basis for judging the targets, but they are unaware that they are doing so.

One potential weakness of the AMP is that it is more susceptible than evaluative priming tasks to participants' attempts to control their responses. Participants could easily choose to respond in a way that reflects the attitudes that they want to portray. For example, if the primes included pictures Black and White faces, participants could attempt to diminish the appearance of racial bias by purposefully pressing the "pleasant" key following the Black faces. However, experiments by Payne and colleagues (2005) suggest that participants are unable to control the influence of the primes on their responses. For example, in one experiment, participants were warned that their responses might be influenced by the primes. Moreover, they were told that under no circumstances should they let the primes influence their responses. In spite of these directions, participants were still significantly influenced by the primes, in the expected direction.

Therefore, even when participants are given a strong incentive to follow directions and ignore the primes, they cannot. These findings suggest that the results obtained in the AMP are usually the result of automatic evaluations of the primes, and not presentational biases.