INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
DO ATTITUDES REALLY CHANGE?

DISSEPTION

Presented in Partial Fulfilment of the Requirements for
The Degree Doctor of Philosophy in the Graduate
School of Ohio State University

By
William Blair Gage Jarvis, C.V., B.A., M.A., M.A.

Ohio State University
1998

Dissertation Committee:  Approved by
Professor Richard E. Petty, Adviser
Professor John T. Cacioppo
Professor Philip E. Tetlock
Professor Edward M. Crenshaw

Adviser
Department of Psychology
ABSTRACT

One of the most common strategies to predict a person's response to a stimulus is to assess his or her current attitude toward the object. We argue that significant predictive power may be gained by also knowing of any prior attitudes the individual has held toward the object. We suggest that in many cases, these 'old' attitudes may remain at least partially intact and can influence current responses to stimuli under specifiable conditions. Through a series of four experiments, we developed a paradigm to create initial attitudes and then change them such that people appear to feel the same as others who never held the previously discrepant attitudes (as evidenced by measures of extremity, accessibility, ambivalence and behaviour). We then demonstrate how a person's prior attitude can still influence evaluative reactions even after it presumably has changed or been replaced with a new attitude. We discuss the potential theoretical dangers of assuming that attitudes necessarily 'change' and the problems associated with assuming that implicit measures are necessarily a direct route to a person's 'real' attitude.
Dedicated to Julie
ACKNOWLEDGEMENTS

I would like to extend the deepest gratitude imaginable to my advisor and collaborator Rich Petty. Thanks Rich. I would also like to thank Zak Tormala who also was instrumental in the conceptualization and execution of this research. Thanks Zak.

I also extend my most sincere appreciation to my family (Pat, Bill, Rob, Vesna, Erik, Aaron, Andrew, Patti, John, Barney & Ginger) and friends (most notably Dave, Shan, Stan & Steve) who all provided me with endless support and encouragement throughout my graduate years here at Ohio State. Thanks everyone!

Additionally, I would like to thank John Cacioppo, Phil Tetlock, Ed Crenshaw, Wendi Gardner, Brian Staub, Endel Tulving and all the members of the Group for Attitudes and Persuasion for taking the time to provide feedback on the conceptual and methodological issues raised in this paper. Special thanks also go to John Skelly who taught me everything I know about priming; Tiffany Ito and Peter Lange without whom our UCS collection could not have been the same; and to Nikolay Sergeevykh for his invaluable help with the tedious process of data collection throughout this program of research.

And, of course, an extra-special warm & fuzzy note to the OSU registrar who put me in the same class with Julie Barnes :) (smile because you made the process of writing this dissertation such a profoundly wonderful experience). Thanks so much.

Finally, I must also extend my deepest gratitude to myself for my patience, understanding and many valuable comments at all stages of this project. I couldn’t have done it without me.
VITA

March 1, 1966  Born Winnipeg, Manitoba, Canada (Pisces)

1990  B.A. Honours Psychology
      University of Western Ontario

1991  M.A. Psychology
      University of Western Ontario

1993  M.A. Social Psychology
      Ohio State University

1988 - 1990  Research Assistant, Cognitive Science
             University of Western Ontario

1990 - 1991  Course Assistant, Introductory Psychology
             University of Western Ontario

1991-1995  Lecturer, Social Psychology
           Ohio State University

1996-1998  Research Consultant
           The Strategy Team, Ltd.

1996-1998  Software Developer and President
           Empirisoft Research Software

PUBLICATIONS


FIELDS OF STUDY

Major Field: Psychology

Primary Specialization: Social psychology, attitudes, persuasion, research design

Additional Academic Training: Attitude measurement, influence and compliance, consumer psychology, cognitive psychology, advanced statistical analyses

Professional/Consulting Experience: Public opinion, market analysis, product testing and evaluation, assessing the effectiveness of persuasive appeals and strategies. Software development for psychological research.
## TABLE OF CONTENTS

Abstract................................................................................................................................. ii
Dedication........................................................................................................................... iii
Acknowledgments............................................................................................................... iv
Vita........................................................................................................................................ v
Table of Contents ............................................................................................................... vii
List of Tables ...................................................................................................................... xi
List of Figures .................................................................................................................... xii

Chapter 1-- Introduction ....................................................................................................... 1
  Clarification of Constructs and Terminology ............................................................... 2
  Reviewing Evidence Consistent with the PAST Model................................................ 4
  Classical Conditioning: Does Extinction Reflect Unlearning?...................................... 6
    Caveats.................................................................................................................. 10
  Stereotypes and prejudice............................................................................................ 11
    Devine................................................................................................................... 12
    More Recent Prejudice Models............................................................................. 15
    Caveats.................................................................................................................. 15
  Personality .................................................................................................................. 18
    Caveats................................................................................................................. 21
  Additional Conceptual Distinctions ........................................................................... 21
    Ambivalence and Affective-Cognitive Consistency ............................................... 22
    Context Dependent Attitudes ............................................................................. 23
    Implicit vs. Explicit Attitudes and Measures........................................................... 23
  Goals of the Current Research ................................................................................... 25
Targets ................................................................................................................................. 54
Similarity Condition ........................................................................................................ 56
Attitude Index .................................................................................................................. 57
Behavioural Choice ......................................................................................................... 57

Results .................................................................................................................................. 57
Attitudes ................................................................................................................................. 57
Choice of Partner ................................................................................................................ 58

Chapter 5 -- Study 4: Testing the PAST Model ....................................................................... 59

Method .................................................................................................................................. 60
Participants ............................................................................................................................ 60
Procedure ............................................................................................................................... 60
  Initial 'Prior' Attitude Induction ......................................................................................... 60
  'New Attitude' Manipulation ............................................................................................ 61
Design ..................................................................................................................................... 63
Implicit Attitude Measure .................................................................................................... 63
Choice Measure .................................................................................................................... 63

Results .................................................................................................................................. 65
Initial (Pre-test) Attitudes ..................................................................................................... 65
Post-Test: Attitude Change via Similarity Manipulation ..................................................... 68
Other Attitude Strength Indicators ...................................................................................... 69
Implicit Attitude Measure .................................................................................................... 71
Choice Measure .................................................................................................................... 72

Chapter 6 -- General Discussion .......................................................................................... 76

Methodological limitations ................................................................................................... 77
Other Directions for Future Research ................................................................................ 81
  Implications for ‘Implicit Attitude’ Research ..................................................................... 86
Why Wouldn’t Attitudes Just Change? .................................................................................. 88
Conclusion .............................................................................................................................. 89
LIST OF TABLES

1. Breakdown of Trials for Study 1 ................................................................. 31
2. Prime-Target Sequences for Six Faces Across Six trial Blocks ....................... 43
3. Design for Study 4 Including Counter-balancing ................................................. 63
4. Simplified Design for Study 4 ........................................................................ 64
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sample Screen from Rating Task</td>
<td>34</td>
</tr>
<tr>
<td>2</td>
<td>Raw Attitude Scores as a Function of Prior Conditioning</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>Transformed Standardized Attitude Scores as a Function of Prior Conditioning</td>
<td>39</td>
</tr>
<tr>
<td>4</td>
<td>Reaction Times for Judging Good and Bad Stimuli as a Function of Prime Prior Conditioning</td>
<td>46</td>
</tr>
<tr>
<td>5</td>
<td>Standardized ‘Positive Impact’ Scores Attributable to Prime as a Function of Prior Conditioning</td>
<td>49</td>
</tr>
<tr>
<td>6</td>
<td>Two Target Faces Used in Studies 3 and 4</td>
<td>55</td>
</tr>
<tr>
<td>7</td>
<td>Attitudes as a Function of Similarity Manipulation for Participants with Congruent Prior Attitudes</td>
<td>66</td>
</tr>
<tr>
<td>8</td>
<td>Attitudes as a Function of Similarity Manipulation for Participants with Incongruent Prior Attitudes</td>
<td>67</td>
</tr>
<tr>
<td>9</td>
<td>Implicit Attitudes as a Function of Similarity and Prior Attitude</td>
<td>73</td>
</tr>
<tr>
<td>10</td>
<td>Choice of Partner as a Function of Manipulated Similarity and Prior Attitude</td>
<td>75</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

The primary purpose of the current research is to formally challenge and directly test the implicit assumption that when attitudes change, the old attitude is literally replaced with the newer attitude. We suggest that in many cases, prior attitudes can remain at least partially intact and can influence current evaluative responding under specifiable conditions.

We support this position by demonstrating that two people holding an equally strong attitude (e.g., in terms of extremity, accessibility, ambivalence and so forth; see Petty & Krosnick, 1995) toward an object respond to that object differently if one of them previously held a discrepant attitude whereas the other did not. This finding is inconsistent with a general model which presumes a single global evaluative association with any given object and that prior attitudes must therefore be replaced or updated when attitude change occurs (we refer to this as the Attitude-Change or AC model; e.g., see Anderson, 1971; Fazio, 1995). If such a model were valid, then how an individual used to feel should have no significant influence on current evaluative responding above and beyond one's current attitude. Since under specifiable circumstances, previous attitudes can be shown to have an impact on current evaluative responding even when current attitudes and attitude strength are controlled, our data appear to be more consistent with
what we refer to as the PAST model (Prior Attitudes are Still There) which allows for past attitudes to remain independently associated with a given object.

We will argue that the alternative Attitude-Change model is in fact commonly assumed in attitude research and we will outline a number of the theoretical and practical implications associated with neglecting to consider people’s prior attitudes.

Although a number of researchers have alluded to something like the PAST model for global attitudes, (e.g., Cacioppo, Marshall-Goodell, Tassinary & Petty, 1992; Petty, Baker & Gleicher, 1991; Petty, Gleicher & Jarvis, 1994; Wilson, Lindsey, Schooler, 1998), only a few studies have provided data that are relevant to a direct test of this position. Accordingly, to demonstrate the plausibility of the PAST model, we briefly review existing data from the research domains of classical conditioning, stereotyping and prejudice, and personality. We point out how data from these domains are consistent with the PAST model yet are insufficient due to both conceptual and methodological issues. We then discuss how our theoretical approach to this issue differs from related ideas in the field. Finally, we will discuss how we have attempted to help clarify this issue with new research.

Clarification of Constructs and Terminology

First, by an attitude we are referring to a relatively enduring positive or negative global evaluation associated in memory with a particular object (e.g., see Fazio, 1995; Petty & Cacioppo, 1981; Eagly & Chaiken 1994). The meanings of current and prior with respect to attitudes depends on the model one assumes. Under the traditional AC model, prior attitudes simply refer to any attitudes that temporally precede an individual’s current attitude. Since only a single attitude is expected to be associated with an object in
memory, the use of the term prior in the AC model appears to be a literal referent to an association that used to exist but is now presumed to be gone or dissociated.

The PAST model, in contrast, assigns the term prior additional meaning beyond its literal definition. If, as the PAST model proposes, any attitude associated with an object in memory has the potential to remain associated with the object, how then would an individual know how he or she actually feels about that object? With multiple attitudes associated with a single object, then some mechanism would be required to inhibit acting on inappropriate associations that come to mind. We suggest that Gilbert's (e.g., Gilbert, Krull & Malone, 1990; Gilbert, 1991; Gilbert, Tafarodi & Malone, 1993; see also Smith, 1998) line of research on the subjective belief mechanism is relevant in this regard. This research has addressed the question of how the validity or truth or information is stored in memory. Gilbert has argued that information by default is not associated with its subjective truth value. Rather, he argues that comprehension is a sufficient condition for belief. Disbelief, in contrast, requires an active 'false' association to be encoded with the information when it is processed. If, and only if, this active association occurs will the individual later recognize the information as intrinsically false.

We suggest that Gilbert's model can be applied to understanding how we know which of our evaluative associations actually represents how we feel about an object. Following Gilbert's model, we suggest that when new attitudes are formed, or when attitudes are no longer considered appropriate, the individual actively encodes an association with the rejected evaluation that, in essence, tags it "false." Consequently, we define prior attitudes in the same manner that Gilbert defines "false" information in memory. That is,
if an evaluative association is tagged as false with sufficient strength such that the "false" association comes to mind with the attitude, then it will be perceived as a prior attitude.

Following this logic, there would be at least three ways an attitude could be reasserted following what we would normally consider attitude change. First, as suggested by Gilbert's research, any failure to explicitly reject an existing attitude when a new one is formed should facilitate its potential to impact future responding. If, for example, the prior association were ever to come to mind before the newly formed attitude, there would be no immediate reason not to act upon it since it would initially be perceived as true. Second, even if the existing attitude is explicitly rejected when the new attitude is formed, this tag would also have to be recalled in order for subsequent rejection to occur (much like evaluative associations must come to mind in order to have any impact). Third, even if the existing attitude is rejected, and when it later comes to mind the individual recalls that it is false, then it would still be required that the individual be able to inhibit the impact of the false attitude.

If attitudes literally change then none of this will be of any consequence. However, if the PAST model is correct, then interfering with an individual's ability to encode it as false, retrieve it as false, or to inhibit its impact should consequently produce evidence of the persisting evaluative association—rendering the AC model unable to account for the resulting response.

**Reviewing Evidence Consistent with the PAST Model**

In reviewing the past literature relevant to our question, we first looked for experimental research in which attitudes had been measured, subsequently changed, and where some attempt had then been made to determine if evidence of the prior attitude
could still be found. Although we found none that directly matched these criteria, we did
find a variable array of research and theory consistent with the idea that associations in
general can often give the outward appearance of having been changed despite that
specifiable conditions can produce evidence of their persistence. The data we review
seem to suggest that when change seems to have occurred, it can often be accounted for
either by an inhibitory mechanism, or by the activation and application of an independent
and relatively more accessible association. For example, as Gilbert has demonstrated,
the changing of a belief reflects the tagging of that association as false—not a literal
dissociation of the rejected association itself.

Research in three areas will be reviewed. In the case of classical conditioning,
associations shown to have been fully extinguished, can best be accounted for by an
inhibitory mechanism which, when disrupted, also allows for evidence of the association
presumed to have been unlearned. In the case of stereotypic traits associated with
minority groups, the changing of beliefs appears also to best be accounted for by an
inhibitory mechanism which, if disrupted, allows for evidence of the rejected
associations. Finally, research from the personality domain offers provocative evidence
suggesting that changes in one's personality though the life span do not necessarily
reflect the dissociation of prior predispositions, but rather the application of new
associations in the self-concept.

We will now briefly review the relevant data from each of these areas as it applies to
our question regarding attitudes, or global evaluative associations. At the conclusion of
each review we also will briefly discuss what we saw as the important conceptual and
methodological caveats to keep in mind when applying these ideas to the domain of attitudes and attitude change.

**Classical Conditioning: Does Extinction Reflect Unlearning?**

> Extinction must be regarded as a special form of inhibition... It cannot be regarded as an irreparable destruction of the conditioned reflex.

Pavlov, 1927 (p. 60)

Pavlov (1927) offered an account of what he termed the internal inhibition of conditioned reflexes that seems to closely parallel more contemporary conceptions of inhibiting rejected associations. Pavlov found that his participants, despite their apparent lack of sophistication, could come to reject associations that were no longer appropriate and respond as if those associations had never been established in the first place. In a typical study, one of Pavlov's dogs would be exposed to a neutral stimulus (e.g., a ticking metronome, or a quantity of meat powder) immediately prior to feeding. It is a well known finding that after repeating this process a number of times, the presentation of the neutral stimulus alone became sufficient to produce a reaction similar to that of feeding alone (e.g., as measured by drops of saliva). Through the repeated pairing of the conditioned and unconditioned stimuli, the agreeable dog would come to associate the neutral stimulus with the tasty event and salivate accordingly.

Another well known finding is that when Pavlov continued to repeatedly expose his dogs to these conditioned stimuli in the absence of subsequent feeding, the conditioned response could be effectively extinguished—that is, the mere presentation of the conditioned stimulus would no longer be sufficient to produce the conditioned response. Essentially, the dogs would come to dissociate the initially neutral object from its
conditioned response (e.g., thinking “Ah, it’s just a metronome.”). Had the dogs unlearned the previous association?

In a lesser known continuation of this paradigm, Pavlov repeatedly demonstrated that the prior conditioned response had not literally been extinguished but rather it only appeared so, he argued, due to a process of internal inhibition. Essentially, these dogs had realized that the neutral stimulus was no longer associated with feeding. It is as if they had tagged this association as false and subsequently rejected it when it came to mind. Consequently, no evidence of the conditioned response could be found. Although this is an obvious anthropomorphism, the way Pavlov describes it, it actually does resonate well with the way we talk about how low-prejudice individuals inhibit the highly accessible, but rejected, associations that come to mind about minority groups. This parallel can be drawn even further with Pavlov’s demonstrations in which the extinguished conditioned association would reveal itself if the dog was distracted, or otherwise cognitively loaded, upon presentation of the now neutral object (c.f. Gilbert and Hixon, 1991). In the simplest cases, knocking on a table or having someone walk into the room during presentation was sufficient to make the dog think “Hey, what’s that!?” and produce evidence of the old response. Pavlov referred to this as disinhibition, and described it as the disruption of the inhibitory mechanism required to prevent the old association from impacting on the dog’s response. This brief survey is of course a simplification of Pavlov’s procedures and represents only a small portion of the data upon which he based his conclusions. However, in consideration of all the data, Pavlov concluded,
Any hypothesis of an irreparable destruction of the conditioned reflex in the process of experimental extinction cannot possibly stand for a moment.

Pavlov, 1927 (p. 67)

Interestingly, contemporary accounts of extinction have apparently neglected Pavlov’s absolute insistence that extinction is a process involving the gradual inhibition of an intact association—and not a process of elimination or change. In a recent review of theory and research relevant to extinction, Baeyens, Eelen, & Crombez (1995) concluded that numerous contemporary accounts of extinction do not follow from Pavlov’s original conceptualization (e.g., Rescorla & Wagner, 1972; Wagner, 1976; 1979; Mazur & Wagner, 1982; Wagner, 1981; Wagner & Brandon, 1989; see Donegan, Gluck & Thompson, 1989 for a review). Baeyens et al. argue that contemporary accounts do actually seem to assume that extinguished associations are, at least functionally, unlearned:

...they share the basic assumption that reinforced CS presentations (acquisition) and unreinforced CS presentations (extinction procedure) act on the strength of the associative link in a way opposing each other’s effect. Hence, a sufficient number of unreinforced CS presentations will finally reduce the (net) associative strength between CS and US to an asymptote of zero, which means that the CS re-enters a state which is functionally identical to the state of a neutral stimulus which was never involved in a CS-US contingency. (p.128)

And further, that,

[Pavlov’s] point of view has not received the attention it deserves in the majority of traditional accounts and formal models of Pavlovian learning. Hopefully then, the evidence which is available now will demonstrate more convincingly the basic truth of Pavlov’s conception of Pavlovian extinction.
Baeyens et al. reasserted Pavlov's position and resurrected Pavlov's intuitive notion that if the initial CS-UCS association necessarily changes through the process of extinction, then it should not be possible to show evidence of the pre-extinction association after it is presumed extinct. Baeyens et al. then continued to review a fascinating and wide array of data that strongly support Pavlov's original conceptualization (see especially Bouton, 1988; Bouton & Bolles, 1985; Bouton & Swartzentruber, 1991).

For example, a number of studies have supported Pavlov's finding that that a completely extinguished CR will reappear at a substantially faster rate if the UCS is ever again paired with the CS. This rapid reacquisition effect has since been demonstrated at both the within subject and between subject levels. At the within subject level, conditioning effects are achieved with fewer CS-UCS pairings after extinction than it takes to create the same effects initially. At the between subject level, conditioning effects are achieved faster with animals in which the current neutral response is a consequence of extinction as compared with animals that were never conditioned in the first place (e.g., Scavio, Ross & McLeod, 1983; see Napier, Macrae, & Kehoe, 1992).

Further, a number of studies have also supported Pavlov's finding that fully extinguished CRs will spontaneously recur even in the absence of any additional CS-UCS pairings. The reinstatement effect, for example, occurs when a completely extinguished CR is challenged by the mere presence of the UCS. In such cases, the CS (which no longer elicits the CR) is simply presented some time after a presentation of the UCS. It has been shown that the even a single presentation of the UCS prior to an exposure of the
extinguished CS can be sufficient to at least partially reinstate the old response (e.g., Bouton & Bolles, 1979; Rescorla & Heth, 1975).

Considering these and the other data they reviewed, Baeyens et al., concluded:

All of the studies reviewed here have in common that they provide independent pieces of evidence to decisively reject the notion that an extinction procedure results in simple unlearning or forgetting of the previously acquired association between the conditioned stimulus (CS) and the unconditioned stimulus (UCS). (p. 127)

Caveats

Are these data consistent with the PAST model? Although the parallels are provocative, there are some problems in generalizing these results to the domain of attitude change. First, we could find no data with human participants that would allow us to see if these results generalize beyond dogs, rats and rabbits. Although a number of studies have looked at the process of extinction with human participants (e.g., Baeyens, Crombez, Van den Bergh, & Eelen, 1988; Miller & Clark, 1969), none have looked for evidence of the prior CR after extinction had been observed (see Baeyens et al, 1995). Rather, these studies have focused on demonstrating the resistance of the CR to be being extinguished in the first place. We are only pointing this out as an obvious caveat and have no intention of addressing the issue of whether or not dogs, rats and rabbits have attitudes (regarding the issue of affect in animals the reader is referred to Darwin, 1872 and LeDoux, 1995).

Perhaps more importantly, extinction research has focused on the process and mechanisms through which a single association is negated rather than replaced with one that diametrically opposes it. In the case of the typical attitude change study, participants are not led to simply reject their current attitude, but are encouraged to adopt a new one.
that will presumably replace it. In fact, one could argue that the extinction data are not incompatible with the AC model. The AC model allows for a single global evaluative association, and does not preclude an individual from deeming that association as incorrect (something akin to Gilbert's model). In such a case, interfering with their inhibition of this single association could produce effects similar to those reported here.

Consequently, we suggest these data should be taken only as suggesting that a rejected attitude might persist intact. A stronger test of the PAST model would require showing an impact of that same rejected association on an evaluative response despite the fact that a new, opposed association had been formed. Only in this case could the AC model not account for it. Next, we review data from the stereotypes and prejudice literature which bring us closer in this regard.

**Stereotypes and prejudice**

A number of researchers have converged on the common conclusion that the strong negative associations held toward minority groups and their members are exceptionally persistent despite conscious rejection or the formation of new beliefs that oppose them (e.g., Banaji & Greenwald, 1995; Devine, 1989; Dovidio, Kawakami, Johnson, Johnson & Howard, 1997; Gaertner & McLaughlin, 1983; Gilbert & Hixon, 1991; cf. Wittenbrink, Judd & Park, 1997).

The common basis for this conclusion is the reliable demonstration of a dissociation between implicit and explicit measures of prejudice toward minorities. Specifically, prejudice measured via self-report is often found to reflect less negative attitudes and, at best, shows a small correlation with attitudes reflected through implicit indicators such as non-verbal behaviour (e.g., Dovidio et al, 1997; Fazio, Jackson, Dunton & Williams,
1995), assimilation of ambiguous information (e.g., Devine, 1989), and response time facilitation (e.g., Dovidio et al, 1997; Fazio et al., 1995; Wittenbrink et al. 1997)

Although the various models proposed to account for this common finding vary, they all seem to share two conceptual features. First, all of the models assert that most individuals will have some degree of association between a given minority group (e.g., blacks) and the characteristics that are stereotypical for that group (e.g., poor, athletic, aggressive). Because these associations are typically expected to have been developed over the course of one’s life, they are expected to be quite firmly established and hence activated automatically upon encountering an individual from the given minority group. Second, they assert that when individuals are given the opportunity to engage in controlled thought between stimulus onset and response, then these automatic associations may be inhibited and are consequently less evident. Therefore, each model asserts that an effective disruption or prevention of one’s ability to engage in controlled processing will allow evidence of the automatic negative associations, if they exist, to be observed.

Devine

The first and most influential model of this type is that described by Devine (1989). Integrating what was known about stereotype processing at the time, Devine argued that most Americans are equally aware of cultural and minority stereotypes but may differ dramatically in the extent to which they feel the stereotypic traits actually describe the relevant group. Devine argued that the stereotypic information remains in tact despite any contrary personal beliefs the individual might form. According to her model, low-prejudice individuals differ from high-prejudice individuals in that they will take any
opportunity to avoid using the automatically activated stereotype in their judgements and behaviour. Consistent with her model, she has found that people who report being low in racial prejudice are less likely to endorse the stereotypic traits, but are as likely as high prejudice individuals to have their implicit responses influenced by them.

If a few critical assumptions are made, Devine’s data appear to support the PAST model. First, if we assume that repeated exposure to stereotypic information facilitates the production of an automatic negative association in those exposed, then most Americans would be unable to avoid the development of a negative attitude toward minority groups. If we also assume that coming to believe in egalitarian principles would facilitate the conscious rejection of this negativity, then we could expect some of these individuals to gradually develop a new positive attitude in place of the rejected negativity. If these assumptions hold true, then we would consider this a case of attitude change and the results of Devine’s research could be evaluated in terms of a strong test of the AC versus PAST model. At the end of this section, we will discuss the tenability of these assumptions, but let’s first assume that we are talking about a case of attitude change.

If so, the AC model can not account for Devine’s data. Individuals who are now unambiguously positive toward blacks continue to show evidence of negativity despite their prior rejection of this association. If the process through which the rejection of negativity and formation of a new positive attitude actually resulted in a literal or even functional change of the attitude, then this should not occur according to the AC model. In contrast and consistent with the PAST model, if the negativity remained to some extent in tact following its rejection and the formation of the new opposing positive
attitude, then these data would be perfectly understandable. High and low prejudice individuals would differ on explicit measures of minority attitudes, since low prejudice individuals would consider any activated negativity as 'false' and reject it. High prejudice individuals, having never experienced attitude change in this regard would not deny their negativity. For high prejudice individuals then, their negative attitude should therefore be reflected in both implicit and explicit measures—being that it’s only attitude that could be reflected. However, the negativity of low-prejudice individuals potentially remaining as a prior attitude after its rejection, could be evidenced though an implicit measure since it could remain directly associated with the attitude object.

Although the PAST model predicts that the prior attitude should still be evident through implicit measures as Devine found, it also implics that if attitude change has actually occurred and is sufficiently strong, then the new positive evaluative association should also have some potential to be evidenced through an implicit measure. In fact, Smith (1998) suggested that this is a potential problem for the Devine model since she clearly argues that there will be no variation in response as measured at the implicit level—despite that new associations may have been formed by some people. He suggested this is problematic because the positive attitudes of low prejudiced individuals should also be available to influence implicit responses. Although the failure to find variation at the implicit level could mean that the new positive attitudes of low-prejudiced individuals are simply of insufficient relative strength to make themselves evident, support for the PAST model would be much clearer if some evidence of both the positive and negative associations could be found at the implicit level.
More Recent Prejudice Models

A number of more recent studies have since demonstrated that prejudicial attitudes measured at the implicit level do in fact seem to vary across individuals (see Wittenbrink et al. 1997 for a methodological critique of Devine, 1989). Using implicit measures of prejudice such as non-verbal behaviour (e.g., Dovidio et al., 1997; Fazio et al., 1995), and response time facilitation (e.g., Dovidio et al., 1997; Fazio et al., 1995; Wittenbrink et al. 1997), it has been demonstrated that some individuals do in fact exhibit less negativity at the implicit level than others. Further, the negativity assessed through non-verbal behaviour and response facilitation have been shown to correlate significantly which would not be expected if negativity at the implicit level was a constant. Although the evidence is not all consistent, it has also been shown that this variable negativity shown at the implicit does vary to a small extent with one's positivity at the explicit level (e.g., Dovidio et al., 1997; Wittenbrink et al. 1997; but see Fazio et al., 1995).

In sum, if we maintain our assumption that being of low prejudice reflects a case of attitude change then these more recent data could be taken as the best support yet for the PAST model. Specifically, low prejudice individuals who have rejected their negativity should show greater evidence of this negativity at the implicit than explicit level, but should evidence less negativity at the implicit level than high prejudice people who have not formed new positive associations. As just described, this appears to be the case.

Caveats

We have asked the reader to suspend evaluation of some critical assumptions in making these arguments. Perhaps the most important assumption is that the prejudice researchers have been dealing with attitudes (at both the implicit and explicit levels) and
that low prejudice individuals necessarily reflect cases of attitude change. A number of conceptual and methodological issues prevent us from drawing any strong conclusions in this regard, however.

Perhaps most importantly, in none of the prejudice studies in these lines of research were individuals' evaluative histories toward the target group actually known or manipulated. Consequently, it is impossible to draw any satisfactory conclusions about the impact of prior attitudes on current responding toward the group in question since there is no way of knowing whether or not people used to feel differently than at the time the study was conducted. In fact, if Devine (1989) was correct, it is theoretically possible that the dissociation effects reported in these studies could occur without anyone ever having to have held a negative attitude toward blacks. From her perspective, an individual could always have held a neutral or positive attitude and yet still evidence prejudice on implicit measures simply because of semantic associations (e.g., lazy, poor) that were never endorsed. The variations in this regard found in more recent research could simply reflect this to a greater or lesser degree in certain individuals. To make confident conclusions regarding the impact of prior attitudes, one must at least know that the prior attitude was in fact a current attitude at some point in time.

Secondly, the explicit attitude measures used in these studies (e.g., MRS; McConahay, 1986) have been questioned with regard to construct validity due to the potential impact of socially desirable responding. In order to make a meaningful comparison between prior and current attitudes in terms of their respective impact, we need to be confident that each reflects a genuine evaluative association with the object in memory. Since self-report measures of attitudes toward stereotyped groups can reflect
attempts to appear unprejudiced (see especially Fazio et al., 1995) in addition to genuine prejudice (e.g., see Wittenbrink et al, 1997), it does not seem fair to compare the impact of these self-report attitudes with implicit measures that are not also subject to conscious fabrication or construction. For our purposes, the MRS is also problematic because it assesses racial beliefs rather than direct evaluative responses (e.g., Over the past few years, Blacks have gotten more economically than they deserve). Consequently, one could argue that even if social desirability were not an issue, that such a measure would provide only a partial representation of one’s explicit attitude toward blacks (e.g., due to the interactive relation between belief and desirability; e.g., Fishbein, 1967; Fishbein & Ajzen, 1975; or due to the attributions one makes to understand the beliefs, e.g., is this problem actually caused by blacks themselves or by poor government? see Heider, 1958).

Third, as with the extinction research, it is possible that the prejudice research involves cases in which a single negative attitude has been rejected, rather than changed to be positive. In the case of the extinction research, we knew the evaluative histories of the participants, but these histories did not include the formation of a new opposing attitude. A controlled inhibition of the single existing attitude could possibly account for the data. In the case of the prejudice research, it seems more plausible that low prejudice individuals have formed new opposing attitudes, but there is no way to know. In fact both Devine (1989) and Dovidio et al. (1997) have explicitly argued that when low-prejudice individuals are given the opportunity to engage in controlled processing, their relatively positive attitudes may simply be reflecting a successful inhibition of their negative associations.
Finally, even if we assume that in some cases the automatic response of a low-prejudiced individual represents a prior attitude, there remains the problem that, in all cases, prior vs. current has been confounded with negative vs. positive. Both conceptually and methodologically the stereotyping literature has focused on cases in which the prior attitude was assumed to be negative and the where the current explicit attitude, if different from the prior attitude, was assumed to be positive. Consequently, we can not know whether the differential impact of the prior and current attitudes on evaluative responding has been a partial consequence of the valence of the associations (i.e., that positivity can not overcome negativity). Further, even if we assume the consequences reported in these studies were in no part a result of the valence confound, we still do not know whether these effects would generalize to cases where the prior attitude was positive and the current attitude was negative. The importance of the valence confound is accentuated by the growing body of evidence demonstrating the often asymmetrical impact of negativity and positivity on evaluative responding (e.g., see Cacioppo & Bernston, 1994).

**Personality**

Finally, research from the personality domain offers provocative evidence suggesting that changes in one’s personality through the life span do not necessarily reflect the dissociation of prior predispositions, but rather the application of new associations in the self-concept. Like the rest of the evidence reviewed so far, data from this domain suggest that under specifiable conditions, implicit measures can reveal the persistence and continuing impact of what explicit measures would not communicate.
McClelland and his colleagues (e.g., McClelland et al., 1989; Weinberger & McClelland, 1990) drew a distinction between self-attributed and implicit personality motives that is strikingly similar to that of the explicit and implicit attitudes distinction (e.g., see Greenwald & Banaji, 1995). Although personality motives have traditionally been conceptualized as distinct from attitudes, we feel that McClelland's logic and findings are provocative and that the potential parallels may prove useful.

McClelland et al. noted that numerous personality theorists were arguing over the value of self-report measures of motives versus less obtrusive implicit measures (e.g., the Thematic Apperception Test). The primary source of this conflict was the recurrent finding that motives measured through self-report and the same motives measured through implicit tests were uncorrelated. This caused many researchers to claim that one or the other method must clearly be invalid.

As an alternative to arguing over which method was providing a measure of the "true" trait, McClelland et al. suggested people might in fact have two distinct motives associated with any given behaviour—something akin to an explicit and an implicit. First, self-attributed motives are viewed as components of the self-concept. They are the motives that an individual believes he or she possesses. According to McClelland, self-attributed motives impact on behaviour when situational variables make the relevant component of the self salient, but they otherwise lay dormant and are uninfluential in determining behaviour. McClelland et al. argued that self-attributed motives rarely predict behaviour when the task contains no motivationally relevant cues. For example, Weinberger and McClelland (1990) noted that in one study (Patten & White, 1977), a self-report measure of achievement motivation predicted performance on a laboratory
task only when there was some external incentive for achievement. When no such incentive was present, self-attributed need for achievement did not relate to performance.

In contrast to self-attributed motives which McClelland argued are unique to humans and are acquired after language development, implicit motives are argued to be either biologically based or learned in the prelinguistic period. Furthermore, in contrast to self-attributed motives that predict behaviour in structured situations where the motive is salient, implicit motives are thought to predict behaviour in unstructured situations where the motive is not explicitly cued (Weinberger & McClelland, 1990).

As with the stereotyping literature, if we assume a parallel between these constructs the data would seem to be consistent with the PAST model. Specifically, implicit motives share with prior attitudes the fact that they temporally precede any self-attributed motive that would come to be perceived as the individual’s current motive. Ergo, self-attributed motives share with current attitudes the fact that they are perceived as the current motive despite the persistence of the individual’s implicit motive.

This parallel is especially nice because the data underlying it are not subject to some of the issues associated with the other domains we have discussed so far. For example, this research is not subject to the valence confound discussed in the stereotypes literature, since the implicit motive is not presumed to be negative, nor is it presumed to reflect a high or low level of the motive. For example, people may be high or low in achievement motivation at the implicit level and either high or low at the explicit level. The samples in these studies have not been restricted to a single content configuration of the independent motives. Additionally, these data seem less subject to the social desirability
issues discussed earlier simply as a function of their relatively less sensitive nature (i.e., as opposed to self-reported prejudice).

Caveats

Because McClelland has argued that implicit personality traits are either biologically based or learned in the pre-linguistic period, we need to use caution in generalizing from these data to attitudes in the general sense. For example, it would not be so surprising to find that biologically based attitudes remained in tact despite other opposing associations formed through one’s life (e.g., see Tesser, 1992).

Again, perhaps the most important caveat is that as with the prior research reviewed so far, in none of these studies were individuals’ implicit and self-attributed motive histories actually known or manipulated (if that were even possible). Consequently, although the relative impact of each can be assessed to some extent using implicit and explicit measure of the target motive, it is impossible to draw any satisfactory conclusions regarding change since there is no way of knowing whether or not people ever held varying levels of either motive than at the time the studies were conducted.

Finally, evaluative associations again were not the dependent variable of interest in these studies.

Additional Conceptual Distinctions

In this final section, we outline how the conceptual contribution of this research can be seen as an advance beyond our current understanding of attitude systems. To do this, we will draw what we feel to be the critical conceptual distinctions between the PAST model and other related concepts in the attitudes literature that we have not addressed so far.
Ambivalence and Affective-Cognitive Consistency

A great deal of research has been directed at understanding the dynamics of simultaneous positivity and negativity toward the same attitude object (e.g., Kaplan, 1972; Thompson, Zanna & Griffin, 1995; Preister & Petty 1996; Cacioppo & Bernston, 1994). Across its various conceptualizations, ambivalence essentially represents a simultaneous state of both positivity and negativity toward the same object resulting in a conflicted global evaluative response (or attitude when it occurs reliably). In cases of extreme ambivalence, it is not clear to the individual whether to approach or avoid the attitude object in question because the perceived costs approach equality with the value of the perceived benefits.

Although the PAST model also allows for simultaneous positivity and negativity, we would argue that it remains distinct from ambivalence for the following reason. Ambivalence has traditionally been conceptualized as the conflict between current feelings about a given object. The individual not only experiences both positive and negative reactions to the same object, but also does not reject the associations of either valence—all associations are perceived as currently valid. In contrast, the PAST model focuses on cases where the person's current feelings conflict with how he or she used to feel. Consequently, the PAST model allows for cases where positivity and negativity co-occur without a simultaneous feeling of conflict or need to resolve one's feelings. Nevertheless, we will have to demonstrate that our findings can not be accounted for by an ambivalence explanation since recent research has suggested that the positive and negative components of a single ambivalent attitude can have differential impact when one is more highly accessible than the other (e.g., see Thompson, Zanna & Griffin, 1995).
We use a similar rational to distinguish the PAST model from the concept of affective-cognitive consistency (e.g., see Chaiken, Pomerantz, & Giner-Sorolla, 1995). Despite that affective and relatively cognitive components of an attitude may differ in valence, they are still conceptualized as being components of one's current attitude. The PAST model again, is referring to cases where two people can share the same current attitude (including their degree of ambivalence and affective-cognitive consistency) and differ only in terms of how they used to feel.

**Context Dependent Attitudes**

Quite a number of studies have demonstrated that attitudes frequently evidence context dependency such that they will predict behaviour better in some situations than others. Attitudes tend to predict best when they are measured at a level of specificity that matches that of the context in which the behaviour will be observed. (e.g., Davidson & Jaccard, 1979; Kraus, 1995; Fishbein & Ajzen, 1975; see also McConnell, Leibold & Sherman, 1997; Smith, Fazio & Cejka, 1996). Although the PAST model also allows for variability in evaluative response, we are conceptually constraining the model to deal with a constant attitude object. In contrast, context dependency implies different attitude objects by definition (e.g. having a Christmas tree in your house in December vs. July). The PAST model would address the variability within each of these attitudes whereas the context dependency research address the variability between them.

**Implicit vs. Explicit Attitudes and Measures**

The distinction between implicit and explicit in the attitudes literature has recently been gaining widespread attention (e.g., Dovidio et al. 1997; Fazio, et al. 1995; Greenwald and Banaji, 1995; Smith, 1998). These terms are also used widely in the
current paper. Exactly how the PAST model differs from this idea depends on whether the distinction is being used to refer to the attitude itself or the measure used to assess it. In some cases, researchers state they are using an implicit attitude measure in order to assess the evaluative response while minimizing the participant’s chances of engaging in controlled thought. Explicit measures in contrast allow for the participant to think about their evaluative response and to respond in any way desired. When the distinction is stated this way, there is virtually no conceptual overlap with the PAST model. Implicit measures simply serve as a tool with which to assess any persistence of evaluative associations that could otherwise be rejected at a conscious level. Explicit measures in contrast serve as a tool with which to primarily assess an individual’s current attitude, since any prior attitude can be consciously rejected at this level.

However, the implicit/explicit extinction is sometimes used to refer to the attitudes themselves. Some might see the prior attitude for example as the implicit attitude, whereas the current attitude represents the explicit attitude. We definitely want to make a distinction here. Most importantly, we would not use the term implicit attitude synonymously with prior attitude. This is because an individual’s current attitude, if sufficiently strong, could also influence an implicit attitude measure (see Smith, 1998). Consequently, an implicit attitude measure could simultaneously reflect the individuals current attitude while still being influenced by their prior attitude. Therefore if one equates the implicit attitude to that which is implied by the implicit measure, then it could represent the current attitude, the prior attitude, or any combination of the two (or more if multiple prior attitudes exist). Under the AC model, this is not a problem because the prior attitude is presumed unimportant. However, under the PAST model, the concept of
an 'implicit attitude' becomes a less informative construct. As we will discuss later, our preference is to avoid the use of implicit and explicit to describe attitudes themselves, and use the terms simply to refer to the measures used for assessment.

Goals of the Current Research

In order to further our understanding of how prior attitudes can impact on current evaluative responding, we attempted to design a program of research that would allow us to draw stronger causal inferences regarding the conceptual outline of the PAST model than provided by any currently existing data. Keeping the caveats and critical conceptual distinctions we have discussed throughout our review in mind, our primary goals were to:

1. Work specifically and unambiguously with attitudes defined as global evaluative associations in memory.
2. Specifically address what happens when attitudes are not simply rejected, but purportedly replaced with a new opposing attitude.
3. Create a paradigm in which we manipulate and therefore have knowledge of each participant's evaluative history with the target objects.
4. Before looking for evidence of prior attitudes, demonstrate the equality of current attitudes in terms of extremity, ambivalence and accessibility such that our test groups differ only in terms of their prior attitudes and not in the strength of their current attitudes.
5. Ensure that the prior vs. current factor is not confounded with negative vs. positive.
6. Eliminate social desirability as a plausible alternative explanation for dissociation. That is, use attitude objects that are not socially sensitive as well as demonstrate the results to be relatively symmetrical across valence manipulations.
In our attempt to achieve these goals, we first developed a paradigm to create initial attitudes using a classical conditioning paradigm targeted at creating positive or negative attitudes toward initially neutral faces. We then developed a similarity manipulation through which to change the attitudes of some individuals such that their new attitudes would match those of people induced to feel that way but *without change*. Our intention was to demonstrate that the attitudes of both groups at this point were now equal as evidenced by measures of attitude extremity, accessibility, ambivalence and behaviour. In other words, through a thorough conventional assessment, we attempted to demonstrate clear evidence of attitude change. Finally, we attempted to examine whether people who had undergone attitude change would produce markedly different patterns of evaluative response on an implicit attitude task than those who had not. This would allow us to determine whether or not knowing individuals’ prior attitudes could help to account for variance in evaluative responses that could not have otherwise been explained.
CHAPTER 2

STUDY 1: CREATING INITIAL ATTITUDES

In our first study, we developed and tested a classical conditioning paradigm that was designed to create the initial (to later be considered ‘prior’) positive and negative attitudes. Participants experienced a 30-minute conditioning session in which 24 neutral faces were repeatedly paired with positive, negative or neutral images. Imbedded within these trials were 6 target faces, and 18 ‘filler’ faces. Two of the target faces were paired 40 times each with different positive images; two were paired with negative images, and two were paired with neutral images. In a subsequent and ostensibly unrelated ‘rating’ task, participants were asked to report their evaluative impressions of each man. The purpose of this first study was to establish that the conditioning task was sufficient to induce positive and negative attitudes as a function of the paired association trials.

Method

Participants

Participants in the first study were 47 undergraduates enrolled in an introductory psychology course at Ohio State University. Sessions were conducted in a single room with four partitioned work areas. Up to four people participated simultaneously during any given session. All participants received course credit.
Procedure

In order to draw participants attention away from the purpose of the conditioning task, they were informed that the purpose of the study was to examine issues regarding visual memory (see Zajonc, 1968). After welcoming participants to our lab, the experimenter introduced himself and gave the following verbal instructions:

What you are going to be doing here today is helping us with some issues regarding visual memory. To do this we are going to be showing you a lot of different images over the course of the next half hour.

This is a very difficult and challenging experiment since it is very hard to focus your attention on one thing for a half hour straight. This is why we simply ask that you do your best to keep your attention focused on the monitor for the full half hour. And after that's over, we'll be looking at a number of the issues that we're interested in.

We also need to forewarn you that some of the images you'll be seeing are of a negative nature—that is they're not pleasant to look at. It's important for you to know that you have the right to discontinue your participation if these images make you feel too uncomfortable. Does everybody understand that?

Participants were then given the same instructions again on their computers, along with the following additional information to explain the presence of the faces before each image:

Important: You may or may not see the word “focus” or a small black and white picture before each image. We are examining how these affect your visual memory for the color photographs. So please be sure to pay careful attention to the screen at all times.

Unconditioned Stimuli

The UCS images (see Appendices B, C & D) we used to elicit evaluative reactions were chosen from the International Affective Picture System (IAPS; Lang, Bradley &
Cuthbert, 1995). The IAPS is a set of approximately 500 digital color images spanning the full range of evaluative content from extremely negative (e.g., mutilated bodies, dead animals) to extremely positive (e.g., puppies, babies). Based on normative data obtained in our own pretesting as well as by Ito, Cacioppo & Lang (in press), we selected 70 of the most uniformly negative, neutral and positive images for a total set of 210. Since we had reaction time data from our own pretesting, we also avoided images that tended to elicit a relatively delayed reaction (> 2 standard deviations above the mean reaction time). Since both sets of pre-test data examined positivity and negativity separately, we were able to avoid pictures that elicited ambivalent responses (see Ito et al., 1997).* To reduce computer load and presentation error, all images were sized to a standard 640 × 480 resolution, and were saved in an 8-bit colour bitmap format (custom colour palettes for each image allowed them to retain their full colour appearance).

**Conditioned Stimuli**

The faces used in the conditioning trials were a set of 24 heterogeneous white male faces ranging moderately on various dimensions of appearance such as age, attractiveness, apparent socio-economic status, and so forth. This set was selected from an initial pool of 120 that we had pretested in an earlier study. The 24 faces we chose were rated as the most evaluatively neutral of the set (all ranging from 3.75 to 4.25 on a 7 point scale). Positivity and negativity were both assessed using the same measures as described shortly, and only those faces that scored neutrally on both dimensions were used. All images were cropped to a head and shoulders shot and were sized to a standard 320 × 240 resolution, and were saved in an 8-bit black and white bitmap format (many
of the original images were black and white, so we converted all of them for consistency).

**Conditioning Task**

Table 1 presents the breakdown of the trials for Study 1. From the set of 24 faces, 6 were randomly chosen for each participant to be the 'target' faces, and 18 were randomly chosen to be 'filler' faces. For each of the three valence conditions (negative—neutral—positive), two faces were used to provide a within subject replication. Consequently, two of the target faces were paired 40 times each with different negative images; two were paired with neutral images, and two were paired with positive images. Importantly, each of the 6 target faces was presented 40 times, each time being paired with a different image from its assigned UCS condition. Therefore, each participant saw two faces each paired with forty different negative images, two paired with forty different neutral images, and two paired with forty different positive images. This resulted in 240 target trials (2 faces × 40 negative UCS + 2 × 40 neutral + 2 × 40 positive = 240).

To reduce awareness of the CS-UCS contingencies, the remaining 18 faces were used for 'filler' trials. These 'filler' faces ranged in frequency from 5 to 20 exposures but otherwise were identical in format to the target trials. The filler trials were balanced such that an equal number of filler faces were seen 5, 10 and 20 times each, and an equal number were paired with negative, neutral, and positive images. This resulted in 210 filler trials (i.e., 6 faces × 5 exposures + 6 × 10 + 6 × 20 = 210). Since trials
### Target Trials (n=240)

<table>
<thead>
<tr>
<th>UCS</th>
<th>Faces</th>
<th>Exposures</th>
<th>Trials</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>2</td>
<td>40</td>
<td>80</td>
<td>1</td>
</tr>
<tr>
<td>Neutral</td>
<td>2</td>
<td>40</td>
<td>80</td>
<td>2</td>
</tr>
<tr>
<td>Positive</td>
<td>2</td>
<td>40</td>
<td>80</td>
<td>3</td>
</tr>
</tbody>
</table>

### Filler Trials (n = 210) – not analyzed

<table>
<thead>
<tr>
<th>UCS</th>
<th>Faces</th>
<th>Exposures</th>
<th>Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>2</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Neutral</td>
<td>2</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Positive</td>
<td>2</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Negative</td>
<td>2</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Neutral</td>
<td>2</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Positive</td>
<td>2</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Negative</td>
<td>2</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>Neutral</td>
<td>2</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>Positive</td>
<td>2</td>
<td>20</td>
<td>40</td>
</tr>
</tbody>
</table>

Table 1. Breakdown of Trials for Study 1
associated with the filler faces were used to distract participants from the main purpose of the experiment, we will not refer to them in the results (see Endnote 3).

All 450 trials (240 target trials + 210 filler trials) were presented in a random order with the only constraint being that the same face would not appear within any run of three trials. On each trial, participants saw a the word ‘FOCUS’ centered on the screen for approximately 500 milliseconds (ms). Next, they saw a face presented for approximately 600 ms, the UCS image for 1000 ms, and then a blank screen for 1000 ms. This conditioning task was accomplished using 90 MHz Pentium class PC’s with standard multi-media capabilities. Display resolutions were set at 640x480. The software used to accomplish the presentation of stimuli was an alpha release of MediaLabRT 1.0 (Jarvis, 1997).

**Dependent Measures**

Following the conditioning trials, participants were informed that we were going to give the information a chance to ‘settle in’ before we started the visual memory task:

The photograph exposure task is complete! We will begin the memory test in about ten minutes after the information has had a chance to “settle in”. In the meantime, we are going to have you rate some of the pictures you saw (and possibly some new ones) for how they make you feel.

Participants then received instructions that they were to rate each of the faces to follow in terms of how good a man they thought he might be, and how bad a man they thought he might be. All faces were rated together in a random order on one dimension, and then again on the other. The order of the two sets of ratings was random for each participant. The instructions for the good and bad rating sets, respectively were as follows:
For each of the faces that follow, we would like to know how good a man you think he might be. By good, we mean that he probably is good natured, helps people, and cares about other people. A scale will be provided for you. Simply indicate your response by typing the number that corresponds to your response.

For each of the faces that follow, we would like to know how bad a man you think he might be. By bad, we mean that he probably is bad natured, does not help people, and does not care about other people. A scale will be provided for you. Simply indicate your response by typing the number that corresponds to your response.

For each set, the faces were rated sequentially on the computer, with the face displayed beside the vertical rating scale. Figure 1 presents a sample screen from the rating task. All ratings were on a seven point scale ranging from 1—not at all good (bad) to 7—very good (bad). The software used to accomplish the presentation and ratings of the stimuli was the beta release of MediaLab 3.0 (Jarvis, 1997). Following the two sets of ratings, participants were thanked and debriefed.

Results

Data Transformation & Design

Since each UCS condition (negative—neutral—positive) had a within subject replication built into the design, we first averaged the responses across the two faces in each valence condition. This resulted in three sets of “good” and “bad” judgements for each participant rather than six. Each of the three sets represented their reaction to faces associated with one of the UCS valence conditions (i.e., negative, neutral, and positive).
How good a man do you think this man might be?

- Not at all good
- Average good
- Very Good

Figure 1: Sample Screen from Rating Task
Analysis of "Bad" and "Good"

Figure 2 presents both the average ratings of "How Bad" and "How Good" participants judged each man to be as a function of prior conditioning. These self-report evaluations of the target faces were first analyzed using a 3 (Conditioning: Negative—Neutral—Positive) × 2 (Attitude Scale: Bad—Good) analysis of variance (ANOVA) on with both factors being treated as within-subject variables. This initial analysis yielded a significant 2-way, Conditioning × Attitude Scale interaction, $F(2,92) = 10.41, p < .001$. The pattern of this interaction suggested that positive conditioning produced the most positive attitudes overall, and that the negative conditioning produced the least positive attitudes overall. As Figure 2 suggests, judged "badness" tended to increase with negative conditioning, whereas judged 'goodness' tended to increase with positive conditioning. We will describe our exploration of the nature of these relations in detail below. Finally, neither of the main effects was reliable, although the general tendency to rate the men as being more good ($M = 4.0$) than bad ($M = 3.5$) was of marginal significance ($F(1,46) = 3.12, p < .10$).

Positive and Negative Conditioning vs. Neutral Baseline

Since we were interested in creating positive and negative attitudes, and not just in a significant difference caused by the manipulation (e.g., varying degrees of positivity), we proceeded by examining the effects of positive and negative conditioning relative to the effects of neutral conditioning. To do this, we calculated participants' judgements in these conditions relative to their judgements of faces that had been paired with neutral stimuli. For each participant, we calculated the difference between how 'bad' they rated faces paired with neutral stimuli and how 'bad' they rated faces paired with positive or
Figure 2: Raw Attitude Scores as a Function of Prior Conditioning
negative images. Similarly, we calculated the difference between how 'good' they rated faces paired with neutral stimuli and how 'good' they rated faces paired with positive or negative images. We then separately standardized these relative 'good' and relative 'bad' scores.

Analyzing these relative judgement scores again revealed the 2-way Conditioning × Attitude Scale interaction, $F(1,46) = 14.49$, $p < .0001$. 'Goodness' ratings were higher when faces had been paired with positive images ($M = .30$) than when paired with negative images ($M = -.34$). In contrast, 'badness' ratings were higher when faces had been paired with negatives images ($M = .41$) than when paired with positive images ($M = -.34$). Again, neither of the main effects was reliable (Conditioning $F = 1.06$, n.s.; Attitude Scale $F = 0.06$, n.s.).

**Source of 2-Way Interaction**

Given that the prior conditioning manipulation was expected to influence the 'goodness' and 'badness' judgements in opposite directions, this alone could account for the 2-way interactions reported above. However, another possibility is that the conditioning manipulation may have affected one judgement more than the other. To examine this possibility, we reverse scored participants' judgements of 'badness' such that higher scores would reflect relatively more positive evaluations. By doing this, we would now expect only a main effect for Conditioning if 'goodness' and 'badness' were being similarly affected. In contrast, a 2-way interaction between Conditioning and Attitude Scale and would occur only if 'Goodness' and 'Badness' were exhibiting independent consequences.
Figure 3 presents the standardized judgement ratings with those for 'badness' reverse scored to reflect relative positivity with higher numbers. As expected, the analysis of these data yielded a significant main effect for Conditioning, $F(2,92) = 10.40, p < .001$. Faces previously paired with positive images were rated more favourably ($M = .32$; relative 'good' averaged with reverse-scored relative 'bad') than those previously paired with negative images ($M = -.37$). With 'badness' judgements reverse scored, the 2-way interaction between Conditioning and Attitude Scale was no longer reliable, $F(2,92) = 0.59$, n.s., indicating that conditioning had a similar effect on both judgements.

**Effect Sizes Relative to Neutral**

The final question we wanted to address was whether positive and negative conditioning both were effective at creating attitudes, or whether the main effect just described was being driven more by one than the other. To examine this, we created a new variable that represented the absolute value of conditioning-consistent attitude change induced by both positive and negative conditioning procedures. For faces that were negatively conditioned, we did this by reverse scoring the relative to neutral 'goodness' judgements. For faces that were positively conditioned, we reverse scored the relative to neutral 'badness' judgements. We analyzed the relative data in the standard 2 (Conditioning: Negative—Positive) $\times$ 2 (Attitude Scale: Bad—Good) ANOVA. The results of this analysis revealed that, relative to neutral conditioning, positive ($M = .32$) and negative ($M = .37$) conditioning did not differ in the extent of their effects, $F(1,46) = 0.06$, n.s. The 2-way interaction with Attitude Scale, again was not significant, $F(1,46) = 0.02$, n.s., indicating that the extent of conditioning relative to neutral also was similar for both judgements of 'goodness and 'badness'.
Figure 3: Transformed Standardized Attitude Scores as a Function of Prior Conditioning
CHAPTER 3

STUDY 2: ARE THESE CONDITIONED ATTITUDES REAL?

In the first study, we found that our classical conditioning paradigm was capable of creating positive and negative attitudes of approximately equal extremity toward initially neutral faces. The purpose for conducting the second study was to develop an implicit attitude measure that could be capable of demonstrating the same evaluative response as the explicit self-report measure described above. We had two reasons for this. First, we needed a measure that would stand a good chance of showing evidence of the attitude if it still remained after a subsequent attitude change manipulation. Second, we reasoned that if we could show an implicit attitude measure to reveal the same attitude as the explicit self-report measure used in Study 1, then it would be difficult for one to argue that the classical conditioning effects were simply due to demand or social desirability effects (e.g., see Page, 1970). If participants in Study 1 were simply constructing their self-reported attitudes at the time of response to follow from the conditioning contingencies, then we should not find the same effects when measuring the attitude without their awareness (see Fazio et al, 1995 for a complete discussion of this rationale). Therefore, we chose to model our methodology after Fazio’s evaluative priming paradigm (e.g., Fazio et al., 1995). This paradigm has become a standard for assessing object-evaluation associations under conditions that minimize the potential for contamination from controlled thought.
Method

Participants

Participants in this study were 36 undergraduates enrolled in an introductory psychology course at Ohio State University. Again, sessions were conducted in a single room with four partitioned work areas. Up to four people participated during any given session. All participants received course credit.

Procedure

Participants completed the same conditioning procedure as in Study 1. The only difference in the methodology occurred after the conditioning trials were over. Instead of receiving the self-report ratings as the bogus filler task, participants were given our new priming procedure. When the conditioning trials were over, participants were again informed:

The photograph exposure task is complete! We will begin the memory test in about ten minutes after the information has had a chance to “settle in”. In the meantime, we are going to have you rate some of the pictures you saw (and possibly some new ones) for how they make you feel.

Instead of rating the faces though, participants were to rate the positive and negative UCS images as being either “good” or “bad” as quickly as possible. Reaction times were recorded. To turn the task into an implicit attitude measure, we briefly presented one of the conditioned faces before each trial to assess it’s impact on response times (see detail below). The basic rationale being that if a person really had a positive evaluation associated with the face in memory, then seeing that face briefly should facilitate the identification of a subsequent positive image as “good”, and relatively inhibit the
identification of a subsequent negative picture as “bad.” The opposite pattern would be expected for a face associated with a negative evaluation (see Fazio et al., 1995, for additional details). Before the priming task began, all participants received the following additional instructions:

Does the picture make you feel good or bad? In a few moments, we will show you a number of images and we want to know if they make you feel good or if they make you feel bad. If the image makes you feel good, press the button marked “GOOD” at the top left corner of your keyboard. If the image makes you feel bad, then press the button marked “BAD” at the top right corner of your keyboard. Remember, there are no right or wrong answers! We are interested in how the images make you feel.

We are most interested in your first impression and immediate reaction to the images. For that reason we want you to respond as quickly as possible to each image. It will help you to respond more quickly if you keep one finger from each hand on the “good” and “bad” buttons at all times. Also, remember to pay close attention to the screen at all times.

Remember to be fast and accurate!! Remember to respond to quickly and as accurately as you can. Remember to make your judgment based on your first impression. Remember to pay close attention to the screen at all times. To help you respond faster, you will again see a face before each image to help you focus your attention. Be sure to pay attention to the face as it will help you respond faster.

**Priming Task Design**

The stimuli used in the priming task were drawn from the same pool of images used in the previous conditioning trials. Recall that in the conditioning paradigm, six faces served as the target stimuli. Two of the faces were repeatedly paired with negative images, two were paired with neutral and 2 were paired with positive images. In the priming task, each of these 6 faces served as a prime 6 times—3 times followed by a positive image (P), and 3 times followed by a negative image (N). One face from each
pair served as a prime for target images in the order NPNPNP, and the other in the order PNPNPN. This design resulted in a total of 36 trials (see Table 2).

<table>
<thead>
<tr>
<th>Prime Valence (CS)</th>
<th>Target Valence by Trial Block</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Negative 1</td>
<td>P</td>
</tr>
<tr>
<td>Negative 2</td>
<td>N</td>
</tr>
<tr>
<td>Neutral 1</td>
<td>P</td>
</tr>
<tr>
<td>Neutral 2</td>
<td>N</td>
</tr>
<tr>
<td>Positive 1</td>
<td>P</td>
</tr>
<tr>
<td>Positive 2</td>
<td>N</td>
</tr>
</tbody>
</table>

Table 2. Prime-Target Sequences for Six Faces Across Six trial Blocks

Again, to reduce participants focus on the 6 primary faces, the remaining 18 faces were used for 'filler' trials. The 'filler' face trials were identical in format to that just described for the 6 primary faces. This resulted in 108 filler trials (i.e., 18 faces × 6 trials = 108). Again, since trials associated with the filler faces were used to distract participants from the main purpose of the experiment, we will not refer to them in the results (see Endnote 3).
The six blocks of trials were presented in 3 sets of 2, with a 30 second rest break occurring after the second and fourth blocks. Target images for the priming trials also were assigned randomly with the constraint that no target image would be repeated within a set of two blocks.

**Priming Trials**

Each trial was preceded by a screen on which the word 'FOCUS' was centered and displayed for a 2.5-s interval. Following Fazio et al. (1995), all primes were presented for 315 ms, followed by a 135-ms interval before onset of the target images. Consequently, the time between prime onset and target onset (stimulus onset asynchrony or SOA) was 450 ms. As described in detail above, six blocks of 30 trials were presented, with each face appearing once in each block. Each face served as a prime for negative images in half the trials, and as a prime for positive images in the remaining trials.

To encourage both speed and accuracy, participants were shown a prompt screen reading 'Please try to respond faster' if their response time was greater than 1500 ms on any given trial, and 'Please wait for the image to appear' if their response time was less than 200 ms. In such cases, the prompt screen was displayed for 2.5 s before the trials resumed. The software used to present the priming trials and assess reaction times was an alpha release of MediaLabRT 1.0 (Jarvis, 1997).
Results

Data Transformation & Design

Before submitting the raw data to analysis, we assigned missing values to response times greater than 1500 ms, and to those less than 200 ms. Together, such occurrences accounted for less than one per cent of all trials (0.8%). Additionally, we excluded response times for trials on which participants responded with the incorrect key (e.g., indicating 'good' to a 'bad' item or vice versa). Consistent with the error rate reported by Fazio et al. (1995), such trials account for approximately 4% of the total.

Since each Prime-Target combination had a within subject replication built into the design (differing only in sequence), we first averaged the responses across the two faces in each Prime-Target combination. This resulted in three sets of “good” and “bad” response times for each participant rather than six. Each of the three sets represented their reaction times to target images associated with one of the three prime conditions (i.e., negative, neutral, and positive).

Analysis of Means

Figure 4 presents the response times for judging good and bad target stimuli as a function of the conditioned valence of the prime. The raw response times were first analyzed using a 3 (Prime: Negative—Neutral—Positive) × 2 (Target: Bad—Good) ANOVA with both factors being treated as within-subject variables. This initial analysis yielded a significant 2-way, Prime × Target interaction, \( F(2,68) = 4.45, p < .02 \). The pattern of this interaction again suggested that faces conditioned with positive images
Figure 4: Reaction Times for Judging Good and Bad Stimuli as a Function of Prime Prior Conditioning.
produced the most positive attitudes overall, and that faces conditioned with negative images produced the least positive attitudes overall. As indicated in Figure 4, response times to 'bad' target images decreased with negative primes (positive prime $M = 641$; neutral $M = 600$; negative $M = 583$), whereas response times to 'good' target images tended to decrease with positive primes (positive prime $M = 592$; neutral $M = 602$; negative $M = 620$). Neither of the main effects was reliable (Prime $F = 1.04$, n.s.; Target $F = 0.12$, n.s.).

**Positive and Negative Conditioning vs. Neutral Baseline**

To further explore the nature of the interaction, we followed the same basic rationale and procedures outlined in our examination of self-report attitudes (see Results, Study 1, above). Consequently, we proceeded by examining the effects of the positive and negative primes *relative* to the effects of the neutral 'baseline' primes. For trials on which 'bad' targets were preceded by positive or negative primes, we subtracted each participant’s average response time to 'bad' targets preceded by neutral primes. Similarly, for trials on which 'good' targets were preceded by positive or negative primes, we subtracted the average response time to 'good' targets preceded by neutral primes. We then standardized the resulting scores separately for good and bad targets.

Analyzing these relative judgement scores again revealed the 2-way Prime $\times$ Target interaction, $F(1,34) = 5.95$, $p < .02$. Relative to neutral primes, good targets were judged faster when preceded by positive primes ($M = -.17$) than when preceded by negative primes ($M = .09$). In contrast, bad targets were judged faster when preceded by negative
primes ($M = -.24$) than when preceded by positive primes ($M = .31$). Again, neither of the main effects was reliable (Prime $F = 1.61$, n.s.; Target $F = 0.21$, n.s.).

**Source of 2-Way Interaction**

Again, since the positive and negative conditioning manipulations were each expected to influence response times to 'good' and 'bad' targets in opposite directions, we reverse scored participants' standardized reaction times to the bad targets such that higher scores would reflect more positive evaluations of the prime. We refer to this new set of scores as the 'positive impact' attributable to the prime. Reflecting positive attitudes, high positive impact scores are attained by a prime to the extent that it speeds responding to positive targets, and slows responding to negative targets. In contrast, negative attitudes are reflected by low positive impact scores which are attained to the extent that the prime slows responding to positive targets, and speeds responding to negative targets.

Submitting the 'positive impact' scores to the standard ANOVA, we would now expect only a main effect for Prime if the impact on good and bad targets was similar.

Figure 5 presents the 'positive impact' scores for positive and negative primes. As expected, the analysis of these data yielded a significant main effect for Prime, $F(1,34) = 5.95, p < .02$. Positive primes ($M = .17$) produced higher positive impact scores for 'good' targets than negative primes ($M = -.09$). Similarly, positive primes produced higher positive impact scores for 'bad' targets ($M = .31$) than negative primes ($M = -.24$). Importantly, the main effect of Prime was no longer qualified by the 2-way interaction with Target, $F(2,92) = 1.61$, n.s., indicating that the conditioning manipulation had a similar effect on both good and bad targets.
Figure 5: Standardized ‘Positive Impact’ Scores Attributable to Prime as a Function of Prior Conditioning
Effect Sizes Relative to Neutral

Finally, to examine again whether the positive and negative conditioning procedures were comparably effective at creating attitudes, we calculated a new variable which was the absolute value of the positive impact scores described above. We analyzed these data using the standard 2 (Prime: Negative—Positive) × 2 (Target: Bad—Good) ANOVA. The results of this analysis revealed that, relative to neutral conditioning, positively ($M = .24$) and negatively ($M = .16$) conditioned primes did not differ in the extremity of their effects, $F(1,34) = 0.21$, n.s.
CHAPTER 4

STUDY 3. ATTITUDE CHANGE MANIPULATION

In Study 2, the response time facilitation task revealed the same pattern of conditioning effects evidenced by the self-report measures used in Study 1. When serving as primes, faces previously paired with negatives images were found to facilitate responding to negative relative to positive stimuli, whereas faces previously paired with positive images were found to facilitate responding to positive relative to negative stimuli. As discussed earlier, the fact that these data paralleled the self-report data observed in Study 1, makes a demand or social desirability explanation of the conditioning effects difficult to argue (see Fazio et al, 1995). But most importantly, in contrast to the self-report measures, the implicit nature of the response time facilitation task now offers us a means for showing evidence of these initial attitudes following an attitude change manipulation.

The next step in developing a test of the PAST model, was to develop a paradigm to change the conditioned attitudes in order to examine the extent to which they retain their influence. Consequently, the purpose of Study 3 was to create a powerful attitude manipulation that could at least conceivably lead to true attitude change, as assessed by typical attitude measures.

Upon consideration of previous work investigating the formation of attitudes toward individuals, the similarity-attraction relationship appeared to be one of the most reliable
findings in this arena (see Byrne, 1971; Newcomb, 1961; and Berscheid and Walster, 1969). One of the strongest predictors of liking for a target individual is the degree to which that individual is in some way similar to the perceiver. Specifically, people tend to like similar others and dislike dissimilar others. The near ubiquity of this relationship has been demonstrated through experimental explorations in a variety of contexts. The relationship has been found, for example, across gender, age, race, culture, occupational status, socio-economic status, etc. (Byrne, 1971; Byrne et al., 1971).

We designed a similarity manipulation based on the Byrne paradigm and tested its effects on attitudes toward two selected target individuals. In this study, participants first reported their own attitudes toward a number of social and political issues. In a subsequent impression formation task, participants learned about the attitudes of two individuals named Eddie and Phil. They also saw pictures of each individual (selected from the faces used in Studies 1 and 2). Based on their own personal attitude data, participants learned that one of the individuals shared similar attitudes and that one held very discrepant attitudes. Following the similarity manipulation, participants rated their liking for each individual and then selected which of the two they wanted to work with in the next experimental session.

**Method**

**Participants**

Participants in the Study 3 were 27 undergraduates enrolled in an introductory psychology course at Ohio State University. Sessions were conducted in a single room.
with four partitioned work areas. Up to four people participated simultaneously during any given session. All participants received course credit.

**Procedure**

Initial instructions asked participants to answer a series of questions regarding their feelings and opinions about a number of social and political issues. Following these instructions, participants reported their attitudes toward several issues on a 7-point scale, ranging from “very much against” to “very much in favor.” They responded to the following 12 issues: Fraternities and sororities; Belief in God; Money as life’s ultimate goal; Smoking in public; Premarital sex; Political party affiliation; Affirmative action; Homosexuals in the military; Death penalty; Freedom of choice with abortion; Spending tax money to protect the environment; and Sex education in school. This questionnaire was administered via computer using a beta version of MediaLab 3.0 (Jarvis, 1997).

Responses were written to a data file that was accessible for the similarity manipulation to follow.

After completing this brief questionnaire, participants read a paragraph describing the ostensible purpose of the experiment. This paragraph read as follows:

The purpose of this experiment is to determine the kinds of things that affect the impressions we form about others. Last year we carried out other studies of this sort. In these studies, several faculty, staff, and students wrote down all sorts of information about themselves, including some of the important events in their lives. We gave this information to experimental participants, along with a picture of the person, and asked participants to form an impression about the people based on this real life information. In the next task we are interested in finding out what you think of them when you find out about their personal attitudes toward social and political issues.
Following this brief overview, participants were told that they would be reading over survey items, much like the ones they themselves had just completed, that had been filled out by two individuals. They were asked to try to form impressions of these people after reading over their responses. Participants were given 60 seconds to view the picture and survey information for each person, which was displayed on the computer as answers to the 12 survey issues described above.

Participants read over this information and were then asked a series of questions designed to assess their attitudes toward the two target individuals. Upon completion of the target ratings, participants were given a bogus debriefing, which thanked them for their time and explained that the pictures and survey information were actually those provided by two of our research assistants. These assistants, participants were told, would be conducting the follow-up interviews for the study, and that participants were free to select whichever assistant they would prefer to work with. Choices in this situation represented the final dependent measure. Subsequent to making this selection, participants were thoroughly and accurately debriefed, and then dismissed.

**Targets**

All participants rated two target individuals in this experiment. The two faces chosen from the set used in Studies 1 and 2 were referred to as 'Eddie' and 'Phil' (see Figure 6). The faces serving as Eddie and Phil were chosen for two primary reasons. One was on the basis that they demonstrated the prototypical pattern of effects found in Studies 1 and 2. Both Eddie and Phil showed the standard pattern of symmetrical conditioning whether assessed with the implicit or the explicit measures. A second reason was to maintain the plausibility of our cover story. Since participants were to be told that Eddie and Phil
Figure 6: Two Target Faces Used in Studies 3 and 4
were actually our lab assistants and the set of faces ranged widely in age, we also chose these two faces because it was plausible that they were the faces of college aged students.

**Similarity Condition**

Participants were randomly assigned to one of two counterbalanced similarity conditions. In both conditions, participants learned that one man was similar and the other was dissimilar. For half of the participants, the similar man was Eddie and for the remaining participants, the similar man was Phil. In both cases, the other face was that of the dissimilar target. The order in which participants learned about Eddie and Phil was randomized. For each man, participants were presented with the picture of his face and a complete list of his responses to the same survey items the participants had completed earlier.

Following the Byrne paradigm, when the target was similar, his responses matched 8 of the participants attitudes exactly. However, unlike Byrne, four non-identical responses were matched in valence but shifted one point on the scale (in the more extreme direction if possible or else in the less extreme direction). It was thought that this occasional shift would make the high degree of similarity seem more plausible than if every target response had been identical to their own.

Also following Byrne, when the target was dissimilar, his responses mirrored 8 of the participants attitudes exactly (e.g., if the participant had reported a -2, then the target’s attitude would be reported as +2). However, to again avoid a suspicious looking set of responses, the four remaining responses were generated in a similar manner but were
shifted an additional point in the opposite direction if possible (e.g., -2 became +3), or in the less extreme direction if not (e.g., -3 became +2).

**Attitude Index**

Participants' attitudes toward Eddie and Phil following the similarity manipulation were assessed with 7 questions. Three of these questions were designed to determine participants' overall evaluations of Eddie and Phil. Specifically, participants were asked to rate them using two 7-point scales: 'how good a person this man might be' ('not at all good' to 'very good'), and 'how bad a person this man might be' ('not at all bad' to 'very bad').

**Behavioural Choice**

Finally, a forced choice measure was included at the conclusion of the experiment. This assessment was to serve as a proxy for a behavioural measure of participants' evaluations of Eddie and Phil. Specifically, participants were told that they were to select either Eddie or Phil as an interviewer for a post-experiment interview session.

**Results**

**Attitudes**

As in Studies 1 and 2, the ratings of 'goodness' and 'badness' were combined to form a single attitude measure (inter-item correlation after reverse scoring 'badness' $r = .839$). The attitude measure was analyzed using a 2 (Target: Phil vs. Eddie) $\times$ 2 (Condition: Eddie-Similar vs. Phil-Similar) ANOVA. The results of this analysis revealed a two-way Target $\times$ Similarity interaction, $F(1,25) = 38.58, p < .001$. Participants liked Eddie ($M =$
5.6) more than Phil ($M = 3.4$) in the Eddie-Similar condition, whereas they liked Phil ($M = 6.0$) more than Eddie ($M = 3.1$) in the Phil-Similar condition. As expected, neither of the main effects was significant ($F_s < 1$) suggesting that neither Eddie nor Phil was preferred more in general, and that the counterbalancing factor had no reliable impact on its own.

**Choice of Partner**

Serving as a behavioural proxy measure, participants were asked at the close of the experiment to select either Eddie or Phil as the interviewer for a follow-up session. These data were submitted to a 2 (Target: Phil vs. Eddie) $\times$ 2 (Condition: Eddie-Similar vs. Phil-Similar) Chi-square analysis. The results of this analysis were consistent with the attitude data. In the Eddie-Similar condition, all 12 participants selected Eddie as their interviewer. Similarly, in the Phil-Similar condition, 14 out of 15 participants selected Phil as their interviewer, $X^2(1) = 23.26$, $p < .001$. 

58
CHAPTER 5

STUDY 4: TESTING THE PAST MODEL

To briefly review, in Studies 1 and 2 we developed a classical conditioning paradigm targeted at creating positive or negative attitudes toward initially neutral faces. These attitudes would later serve as participants' 'prior' attitudes, once changed. Then, in Study 3, we developed a similarity manipulation that would perhaps be capable of changing the attitudes created through the conditioning procedure. Our next goal then was to combine these two paradigms in a single study. In Study 4, our intention was to use the similarity manipulation to change the initial attitudes (created through the conditioning paradigm) of some individuals such that their new attitudes would match those of people induced to feel that way but without change. It was also our intent to demonstrate that the attitudes of both groups at this point were now equal in valence and strength as evidenced by measures of attitude extremity, accessibility, ambivalence and behaviour. In other words, through a thorough conventional assessment, we attempted to demonstrate clear evidence of attitude change. Finally, we attempted to examine whether people who had undergone attitude change would produce markedly different patterns of evaluative response on an implicit attitude task than those who had not. This would allow us to determine whether or not knowing individuals' prior attitudes could help to account for variance in evaluative responses that could not have otherwise been explained.
Method

Participants

Participants in the second study were 58 undergraduates enrolled in an introductory psychology course at Ohio State University. Sessions were conducted in a single room with four partitioned work areas. Up to four people participated during any given session. All participants received course credit.

Procedure

When participants first arrived in the lab, they were instructed that we were conducting a variety of tasks, the first of which was a basic questionnaire regarding various social and political issues. This questionnaire was the one described in Study 3 above. Unbeknownst to them, their responses to this questionnaire would later be used in the attitude change manipulation (described below).

Initial 'Prior' Attitude Induction

After completing the short attitude questionnaire, we instructed participants that the primary purpose of the session was to examine issues having to do with visual memory. The instructions for this part of the session were identical to those given in Study 1. To create the initial attitudes, we used the same conditioning procedure as before except that, rather than randomly assigning all faces to the various conditions, we constrained the selection such that two of them (“Eddie” and “Phil”) served as primary targets for all participants (the remaining faces were randomly assigned to complete the design). With this constraint, the faces of Eddie and Phil were always paired 40 times with either positive or negative images. For half of the participants, Eddie was paired with negative
images and Phil was paired with positive images. This assignment was counterbalanced for the other half of the participants. Aside from this, the conditioning paradigm was identical to that described in Study 1.

As in Studies 1 and 2, participants were instructed at the end of the conditioning trials that we needed to give the information a chance to ‘settle in’ before the visual memory test. During this time, participants were administered the rating task in which they judged how ‘good’ a man and how ‘bad’ a man they thought each of the 24 men might be.

‘New Attitude’ Manipulation

Following these ratings, participants were given additional instructions that led them into the attitude change manipulation procedure described in Study 3 above. It was explained to participants that we were interested in knowing what kind of impressions they would form about these men if they had some personal information upon which they could base their judgements. This ‘personal’ information was in the form of how these men felt about various social and political issues. The instructions were modified slightly in order to explain the preceding task in which they rated all 24 faces (which was not part of Study 3):

The purpose of this part of the experiment is to determine the kinds of things that affect the impressions we form about others. Last year we carried out other studies of this sort. In these studies, several faculty, staff, and students wrote down all sorts of information about themselves, including some of the important events in their lives. We gave this information to experimental participants, along with a picture of the person, and asked participants to form an impression about the people based on this real-life information. The pictures you just saw were the pictures of these people. The task you just completed will help us to see what impressions people have of these individuals when they know nothing about them at all. In the next
task, we are going to give you some actual information about some of these people.

We are interested in finding out what you think of these people when you find out about their personal attitudes toward social and political issues.

The surveys they filled out are similar to the one you filled out at the beginning of this experiment. You will now be allowed to see the surveys of two other people who have filled out this questionnaire, along with their pictures. Please read their answers carefully and try to form an impression of each of the two people. You will have 60 seconds to form your impression of each person.

After these instructions were presented, the methodology was identical to the procedure used in Study 3 above. Participants were presented with pictures of the two target individuals (Eddie and Phil) along with opinion data that were similar to their own for one and dissimilar for the other. Following the similarity manipulation, participants were asked to rate each of the target individuals on the same set of rating scales used in Study 3. Of primary interest were the two standard evaluation scales asking participants to rate how 'good' and how 'bad' a man they thought each target might be.

Assignment of which target was similar and which was dissimilar depended on the initial attitudes induced through the conditioning procedure. Half of the participants were assigned to the Congruent Prior Attitude condition in which the similar target was the one who had previously been associated with positive images and the dissimilar target was the one who had been associated with negative images. The other half of the participants were assigned to the Incongruent Prior Attitude condition in which the similar target had been conditioned with negative images, and the dissimilar with positive. This design is presented below in Tables 3 and 4.
Design

Table 3 presents the full 2 (Prior Attitude: Congruent vs. Incongruent) × 2 (Conditioning: Negative vs. Positive) × 2 (Similarity: Similar vs. Dissimilar) × 2 (Counterbalancing: Ed Positive-Phil Negative vs. Phil Positive-Ed Negative) factors. Since the Conditioning and Similarity factors were manipulated within subject, this resulted in four between subject conditions as indicated by the column labelled ‘Between.’ In two of these conditions (1 & 2), participants’ new attitudes toward Ed and Phil were congruent with their prior attitudes. In the other two conditions (3 & 4), participants’ new attitudes toward Ed and Phil were incongruent with their prior attitudes.

<table>
<thead>
<tr>
<th>Prior Attitude</th>
<th>Between</th>
<th>Ed</th>
<th>Phil</th>
<th>New Attitude</th>
<th>Via Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Prior Attitude</td>
<td>Via Conditioning</td>
</tr>
<tr>
<td>Congruent</td>
<td>1</td>
<td>+Positive</td>
<td>-Negative</td>
<td>+Similar</td>
<td>-Dissimilar</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-Negative</td>
<td>+Positive</td>
<td>-Dissimilar</td>
<td>+Similar</td>
</tr>
<tr>
<td>Incongruent</td>
<td>3</td>
<td>+Positive</td>
<td>-Negative</td>
<td>-Dissimilar</td>
<td>+Similar</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-Negative</td>
<td>+Positive</td>
<td>+Similar</td>
<td>-Dissimilar</td>
</tr>
</tbody>
</table>

Table 3. Design for Study 4 Including Counter-balancing
However, since we were not interested in differences between Phil and Ed per se, collapsing across the Counterbalancing factor produced the following simplified 2 (Prior Attitude: Congruent vs. Incongruent) × 2 (Conditioning: Negative vs. Positive) × 2 (Similarity: Similar vs. Dissimilar) design:

<table>
<thead>
<tr>
<th>Prior Attitude</th>
<th>Between</th>
<th>Prior Attitude Via Conditioning</th>
<th>New Attitude Via Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Target 1</td>
<td>Target 2</td>
</tr>
<tr>
<td>Congruent</td>
<td>1</td>
<td>+Positive</td>
<td>-Negative</td>
</tr>
<tr>
<td>Incongruent</td>
<td>2</td>
<td>-Negative</td>
<td>+Positive</td>
</tr>
</tbody>
</table>

Table 4. Simplified Design for Study 4

*Implicit Attitude Measure*

Following the attitude change manipulation and self-report attitude assessment, participants were instructed that it was now time to begin the ‘visual memory task.’ This task was identical in all respects to the design of the priming procedure used in Study 2. As in Study 2, all faces (including Eddie and Phil) served as a prime a total of 6 times—3 times followed by a positive image, and 3 times followed by a negative image.

*Choice Measure*

Finally, the same forced choice measure used in Study 3 was included at the conclusion of the experiment. This assessment was again to serve as a proxy for a
behavioural measure of participants' evaluations of Eddie and Phil. Specifically, participants were told that they were to select either Eddie or Phil as an interviewer for a post-experiment interview session.

Results

Figures 7 and 8 present participants' attitudes toward the similar and dissimilar targets immediately before and immediately following the similarity manipulation. Figure 7 presents the means for participants in the Congruent condition in which the similar (dissimilar) target was also the target for which positive (negative) attitudes were induced through the conditioning paradigm. Figure 8 presents the means for participants in the Incongruent condition in which the similar (dissimilar) target was the target for which negative (positive) attitudes were induced through the conditioning paradigm. In other words, Figure 7 represents the attitudes of participants whose prior attitudes were congruent with their new attitudes, whereas Figure 8 represents the attitudes of participants whose prior attitudes were incongruent with their new attitudes.

Initial (Pre-test) Attitudes

We first needed to ensure that our Congruent and Incongruent samples were equally affected by the conditioning manipulation. In order to examine this, we first calculated participants initial attitudes toward ‘Eddie’ and ‘Phil’ by averaging the two attitude scales (“How good..” and “How bad..” with the negative judgement reverse scored to reflect more positive judgements). Attitudes were then standardized to a mean of 0 with a
Figure 7: Attitudes as a Function of Similarity Manipulation for Participants with Congruent Prior Attitudes.
Figure 8. Attitudes as a Function of Similarity Manipulation for Participants with incongruent Prior Attitudes.
standard deviation of 1. We then assigned a value to each of the two resulting attitudes which corresponded to whether the face served as the similar or dissimilar target in the subsequent similarity manipulation. The resulting design was a 2 (Conditioning: Negative vs. Positive) × 2 (Prior Attitude: Congruent vs. Incongruent) × 2 (Similarity: Similar vs. Dissimilar), with the second factor being treated as a within subjects.

Consistent with the results of Study 1, the ANOVA revealed a main effect for Conditioning such that negative conditioning ($M = -.41$) produced more negative attitudes than positive conditioning ($M = .46$), $F(1,56) = 49.44$, $p < .001$. This main effect was not qualified by any interactions (all $F$s < 1). Importantly, this implied that the positive attitudes and negative attitudes created by the conditioning paradigm were equivalent across our two between subject experimental groups (i.e., Congruent and Incongruent).

Post-Test: Attitude Change via Similarity Manipulation

Next, we need to ensure that the subsequent similarity manipulation was effective in changing the attitudes of participants in the Incongruent condition, such that their new attitudes appeared to be equivalent to those of participants in the Congruent condition. To do this, we first calculated participants new attitudes toward ‘Eddie’ and ‘Phil’ after the similarity manipulation. Again, this was done by averaging the two attitude scales (“How good..” and “How bad..” with the negative judgement reverse scored to reflect more positive judgements) and then standardizing the resulting attitudes to a mean of 0 with a standard deviation of 1. We again assigned a value to each of the two resulting attitudes which corresponded to whether the face served as the similar or dissimilar target. The resulting design was a 2 (Prior Attitude: Congruent vs. Incongruent) × 2
(Similarity: Similar vs. Dissimilar), with the second factor being treated as a within subjects.

Consistent with the results of Study 3, the ANOVA revealed a main effect for Similarity such that the dissimilar target ($M = -.72$) produced more negative attitudes than the similar target ($M = .72$), $F(1,56) = 123.83, p < .001$. Importantly, this main effect was not qualified by a Prior Attitude interaction, $F(1,56) = 1.37, n.s..$ This main effect with no qualifying interaction indicated that the similar target was viewed more positively than the dissimilar target, regardless of whether participant held a congruent or incongruent prior attitude.

Other Attitude Strength Indicators

Although we demonstrated that the current attitudes of participants were equivalent in terms of attitude extremity, it seemed possible that an incongruent prior attitude could influence other attitude strength indicators that could potentially account for our effects. Thus, in order to further support that these attitudes were in fact equivalent by a standard attitude assessment, we examined these two possibilities in this regard.

Although there exist numerous attitude strength indicators (e.g., see Petty & Krosnick, 1995), we chose two for our purposes that seemed especially relevant for our paradigm: attitude accessibility and attitudinal ambivalence. Fazio (1993) demonstrated that attitudes of comparable extremity could produce differential priming patterns if they varied in their accessibility. Specifically, an attitude that comes to mind more quickly will produce a stronger priming effect than one which does not. Therefore, we wanted to examine whether having an incongruent prior attitude had an impact on the accessibility of current attitudes. If having an incongruent prior attitude impacts on the ease with
which a current attitude comes to mind, then this could potentially account for our priming patterns.

To examine accessibility, participants' log-transformed response times to the two attitude scales (i.e., 'how good' and 'how bad') were standardized and then averaged to serve as an indicator of attitude accessibility (e.g., see Fazio, 1990a). The resulting accessibility scores reflected greater accessibility (i.e., shorter response times) with lower numbers. Using the same ANOVA design as for attitude extremity above, this analysis revealed no main effect or interaction. Participants with Incongruent prior attitudes ($M = -.04$) evidenced equivalent attitude accessibility scores as those with Congruent prior attitudes ($M = .04$).

Secondly, ambivalence represents a simultaneous state of both positivity and negativity toward the same object. Recent research has also suggested that the positive and negative components of a single ambivalent attitude can have differential impact when one is more highly accessible than the other (e.g., see Thompson, Zanna & Griffin, 1995). Since two attitudes can be of equal extremity yet differ in their extent of ambivalence, we thought it was important to demonstrate that participants' current attitudes did not differ in this regard.

To examine ambivalence, we used a variety of calculations used to approximate subjective ambivalence (e.g., Kaplan, 1972; Priester & Petty, 1996; Thompson et al. 1995). Because the results were highly consistent across these various measures of ambivalence, we report the results here using the Kaplan (1972) measure for illustrative purposes in which the lower of the two attitude extremity judgements served as the index of ambivalence. For example, if a participant rated the target as 7 on 'how good', and 4
on 'how bad', then this participant’s ambivalence score would be 4. If the same participant rated the target as 2 on 'how bad', then this would reflect the lesser degree of ambivalence with a score of 2. Again, using the same ANOVA design as for attitude extremity and accessibility analyses, this analysis revealed no main effect or interaction. Participants with Incongruent prior attitudes ($M = 2.43$) evidenced equivalent ambivalence scores as those with Congruent prior attitudes ($M = 2.40$).

**Implicit Attitude Measure**

Given that participants with Incongruent prior attitudes now evidenced the same attitude extremity, accessibility, ambivalence, and choice behaviour as those with Congruent prior attitudes, it seems reasonable to argue that by a standard attitude assessment, we had successfully changed their prior attitudes. The Congruent and Incongruent prior attitude groups no longer differed as they did in the past.

To now examine whether evidence of the prior attitude could still be found, we looked at the response time facilitation data. To create attitude indices as reflected by the response time task, we followed the same procedure used in Study 2. Before submitting the raw data to analysis, we assigned missing values to response times greater than 1500 ms, and to those less than 200 ms. Together, such occurrences accounted for less than one per cent of all trials (0.7%). Additionally, we excluded response times for trials on which participants responded with the incorrect key (e.g., indicating ‘good’ to a ‘bad’ item or vice versa). Again, this was consistent with the error rate reported by Fazio et al. (1995), such trials account for approximately 4% of the total.

For each of our two faces, we had three "good" and three "bad" response times. Since positive and negative attitudes were each expected to influence response times to 'good'
and 'bad' targets in opposite directions, we again reverse scored participants' standardized reaction times to the bad targets such that higher scores would reflect more positive evaluations of the prime. We then averaged the six response times from each participant for each face. This new variable now reflected greater positivity toward the prime with high values, and greater negativity with low values.

Figure 9 presents the attitudes toward the Similar and Dissimilar targets separately for participants in the Congruent and Incongruent prior attitudes conditions. The attitude data were submitted to a 2 (Prior Attitude: Congruent vs. Incongruent) × 2 (Prime: Similar vs. Dissimilar) ANOVA with the second factor being treated as within subjects. The results of this analysis revealed a main effect such that the Similar prime (M = .11) produced more positive attitudes than the Dissimilar prime (M = -.14), F(1,56) = 5.03, p < .03. However, this main effect was qualified by a Prior Attitude interaction, F(1,56) = 3.88, p < .05. Examining the simple main effects, it was evident that participants with Congruent prior attitudes showed greater positivity toward the Similar prime (M = .18) than toward the Dissimilar prime (M = -.31), F(1,28) = 6.58, p < .01. However, participants with Incongruent prior attitudes showed no greater positivity toward the Similar prime (M = .02) than toward the Dissimilar prime (M = .05), F(1,28) = 0.06, n.s.

Choice Measure

In addition to intrinsic indicators, another way to assess the relative strength of two attitudes is via their consequences. For example, two attitudes equal in extremity may produce different behavioural consequences if they differ in strength (e.g., see Petty & Cacioppo, 1986). Therefore, in addition to assessing strength via intrinsic indicators such as accessibility and ambivalence, we also examined participants' choices regarding
Figure 9: Implicit Attitudes as a Function of Similarity and Prior Attitude
with whom they would prefer to work. We reasoned that similar behavioural outcomes would suggest further evidence of equivalence in strength between our experimental groups. Additionally, since the behavioural measure was to follow the priming procedure, we expected this would serve as additional evidence that participants’ current attitudes had in fact persisted through the duration of the priming task.

Again, following the implicit attitude assessment, participants were told that Eddie and Phil were actually lab assistants, and that they could choose which of the two they would like to work with in a follow-up interview. As shown in Figure 10, participants were equally likely to choose the similar target (93%) over the dissimilar target (7%), regardless of whether they held a Congruent or Incongruent prior attitude.
Figure 10: Choice of Partner as a Function of Manipulated Similarity and Prior Attitude
CHAPTER 6

GENERAL DISCUSSION

The current studies appear to have provided initial support for the PAST model. The results of Study 4 in particular demonstrated that different people can appear to hold similar attitudes as assessed by traditional measures, yet evidence significant discrepancies in their evaluative responses when assessment occurs via an implicit attitude measure. Because of the methodological structure of the current studies, we were able to plausibly explain this variance as being a function of people’s prior attitudes—a possibility not permitted by a model which presumes literal attitude change.

Although a number of previous studies were argued to have provided data that make literal attitude change seem implausible, our primary goal in this research was to design a paradigm that would allow us explicitly test this hypothesis and to make stronger causal inferences than that provided by any existing data.

In Studies 1 and 2, we developed a classical conditioning paradigm to create the initial attitudes. We demonstrated that this paradigm was capable of creating positive and negative attitudes toward initially neutral targets, and that these attitudes were consistent across both implicit and explicit measures of attitude. Then, in Study 3, we developed a similarity-based attitude change manipulation and demonstrated it to be effective in creating strong positive and negative attitudes toward our neutral target objects.
Finally, in Study 4, we combined the paradigms. After creating initial attitudes with the conditioning paradigm, we demonstrated that with the similarity paradigm we were able to change the attitudes of some participants such that their new attitudes ‘matched’ those of participants who had been induced to feel the same way but without change.

Importantly, we demonstrated that the attitudes of both groups at this point were now equivalent as evidenced by traditional attitude measures. The attitudes were also equivalent in strength as assessed by measures of accessibility, ambivalence and impact on choice. We therefore suggested that according to conventional assessment techniques, we had demonstrated clear evidence of equivalent attitude change. The critical test of the PAST vs. AC models at this point then involved a test to determine whether or not evidence of people’s prior attitudes could still be found.

To this end, we examined whether participants who had undergone attitude change would produce markedly different patterns of evaluative response on an implicit attitude measure than those who had not. The results of Study 4 demonstrated that people who appeared to feel the same way by conventional measures reacted quite differently when their attitudes were assessed with an implicit attitude measure. Most importantly, this variance at the implicit level could be accounted for by knowing participants’ prior attitudes. At this point, we discuss what we see to be the limitations, implications and future directions that follow from these results.

**Methodological limitations**

As discussed in the introduction, we attempted to remedy a number of factors that seemed to limit the inferences that could be drawn from prior research regarding the
plausibility of the AC model. We briefly discuss a few issues here that bear on the success of these efforts.

Were We Working With Attitudes? Our intention was to work directly and unambiguously with attitudes defined as relatively enduring, global, and direct evaluative associations with an object in memory. This goal was an attempt to alleviate the problem of generalizing from data associated with other constructs (e.g., stereotypes, prejudice, personality, affective responses, temporary evaluative constructions, and so forth). When compared to these alternative constructs, it may seem obvious that we were working more directly with attitudes in the current research. However, if we consider the strict definition of attitudes (above), and whether or not we met those criteria, there are some potential issues.

For example, rather than creating attitudes toward our neutral stimuli with the conditioning procedure, we could potentially have produced indirect associations between the neutral faces and people’s unconditioned responses to the positive and negative images. In fact some classical conditioning theorists argue that conditioned responses do not reflect direct associations between CS and CR, but rather an indirect link between the CS and UCR, mediated by the CS-induced activation of the UCS (see Baeyens et al., 1995). By our strict definition of an attitude as a direct evaluative association, an evaluative response resulting from such an indirect link would not qualify as an attitude.

We argue though that this would seem much more plausible if we had paired each face with the same UCS each time. Rather, we used a different UCS on each conditioning trial. This makes it very difficult to suggest that the evaluative response elicited by any
given face was in fact the unconditioned response elicited by a highly associated UCS—simply because there was never a case of a highly associated UCS. However, we still have to acknowledge this or some related mechanism as a possibility (e.g., mediation through multiple UCS activations, or even through an activation of the UCS category, e.g., “phil” -> “negative pictures” -> “bad”).

Similarly, one could argue that the similarity manipulation might not have created an attitude, but rather a constructed and temporary evaluative response. If this is the case, then the AC model could account for the results with the argument that attitude change did not actually occur (e.g., because the new response was a temporary construction and not an attitude per se, it could simply have decayed before the response time facilitation task). We offer two responses to this suggestion. First, the attitudes induced through the similarity manipulation were highly predictive of the behavioural choice measures assessed following the response time facilitation task. Therefore, one would have to assume 1) the initial response was a temporary construction, 2) that it decayed before the response time task, and 3) that it was then was reconstructed for the choice task. Second, even if this is assumed, one would still have to account for the fact that the attitudes induced through the similarity paradigm were evident in the response time task (i.e., the prior and current attitudes appeared to cancel each other out). Since the response time task is presumed to prevent the influence of controlled or constructed evaluative thought, the possibility of the similarity based attitudes being temporary constructions seems limited.

**What does the response facilitation task really reflect?** In the current studies, the evidence we provided for the persistence of prior attitudes came from the fact that people
with discrepant current and prior attitudes could be identified through their patterns of response time facilitation on an evaluative judgement task. What if an existing association could, if activated, impact on other judgements without impacting the person's evaluation of the target object itself? One might then expect to find response time facilitation effects but no evidence of the prior attitude with respect to how the target object itself is actually evaluated (at the implicit or explicit level).

This could potentially be an issue because we did not show direct evidence of the prior attitude influencing current evaluative responses toward the same object. Rather we demonstrated evidence that the prior association can have an impact on judgements toward other objects (i.e., by speeding or slowing the judgement latency). This does not seem like a critical issue, but we would suggest that in future research it would be useful to also examine the potential impact of prior attitudes on implicit measures that more directly reflect how the target object itself is evaluated (e.g., subtle non-verbal behaviour toward the object, evaluative responses toward the target object reported under cognitive load).

*Were the Attitudes Really the Same at Time 2?* We argued that our participants who had experienced attitude change following the similarity task appeared to have equivalent attitudes to those who did not. We demonstrated this by showing that our attitude change group did not differ from the other participants in terms of the valence or strength of their current attitudes (i.e., via measures of extremity, accessibility, ambivalence, and choice behaviour). This was a critical demonstration so that we could rule out the strength of participants' current attitudes as the cause of the response time effects. For example, if attitude change participants experienced literal attitude change but felt more ambivalent,
or had current attitudes that were less accessible, these factors could potentially have accounted for the reduced priming effects. Although we showed that our groups did not differ on any of these current attitude measures (all F's < 1), we need to emphasize that our claim of 'equivalence' is based on null findings. Consequently, there are a number of factors that could have influenced this conclusion (e.g., insufficiently sensitive measures, insufficient statistical power). Additionally, one must consider that attitudes can vary on other dimensions which we did not assess (e.g., confidence, affective-cognitive consistency, importance; see Petty & Krosnick, 1995).

Other Directions for Future Research

**Dual Memory Systems.** A number of recent reviews strongly implicate at least two separate memory systems that could be relevant to understanding our current findings (e.g., see McClelland, McNaughton & O’Reilly, 1995; Sloman, 1996; Smith, 1998; Smolensky, 1988). Across various conceptualizations, one memory system is generally considered to *associative* in nature in that it represents the cumulative experience between concepts, and the other is generally considered to be interpretive or rule-based in nature and can represent learning from as little as a single experience.

With respect to attitudes, the associative memory system may become involved when there are many trials linking an object to an evaluation. Since evaluations based on associative memory are a consequence of many experiences, they are activated automatically, but are slow to change (e.g., see Smith, 1998). In contrast, interpretive or rule-based memory can learn rapidly from a single symbolic experience. Since this memory is interpretive rather than associative it is generally expected to require more effortful and conscious processes.
We think this conception is quite compatible with the PAST model, and that it suggests some interesting directions for future research. First, if one assumes that prior attitudes will often be based on more trials or experiences than any given “new” attitude, then prior attitudes could plausibly remain based on associative memory while relatively new attitudes could be based on rule-based memory. In our study, we seem to have paralleled this idea. The initial attitudes were based on 40 conditioning trials, whereas the relatively new attitudes were based on a single event in which participants learned more about the target.

We point this out because our findings could potentially be dependent on this difference. If these two memory systems are to some extent independent, as Smith suggests, then evidence of the prior attitude may depend on it being in a different memory system. In future research it would be useful to determine whether prior attitudes can be found to co-exist with current attitudes within the same memory system. For example, one could use a counter-conditioning paradigm in which attitude change would occur through a process of conditioning over many trials. This would parallel the concept of a new attitude becoming highly practised over time and consequently represented in associative rather than rule-based memory. The question then would be whether a prior attitude could still impact on evaluative responding when the new attitude reaches some criterion level of association. If not, then perhaps it is at this time that attitude change could finally be considered to have occurred (i.e., the old attitude, at least within one memory system, replaced with a new one).

Conditions Facilitating the Impact of Prior Attitudes. Early in the introduction, we suggested the influence of prior attitudes might be facilitated by: 1) a failure to explicitly
reject an existing attitude when a new opposing one is formed (i.e., failing to tag it false), 2) a failure to *recall* the ‘false’ tag if and when the prior attitude subsequently comes to mind, and 3) an inability to inhibit the application of the prior attitude even if it is recognized as false.

By which of these mechanisms might our priming paradigm have facilitated the use of participants’ prior attitudes? At this point, we can’t say. It seems likely that the high speed nature of the task could have interfered with the recollection of any ‘false’ tag associated with a given prior attitude. However, it also seems likely that the same time pressure could have interfered with the ability to inhibit the old attitude if in fact a ‘false’ tag *was* activated with the attitude. Additionally, it seems plausible that the conditioned attitudes may never have been explicitly tagged as false when the new attitudes were formed. In future research, it would be useful to tease these possibilities apart and determine the extent to which each of them facilitates the impact of prior attitudes.

We think it would be particularly useful to examine the role of explicit rejection of the prior attitude at the time of attitude change. For example, prior to our response time task, imagine that we had reminded participants of their prior attitudes and asked them whether or not they still felt that way. By doing this, participants would have been given the opportunity to explicitly reject their prior attitudes, in essence, tagging them as false. This creates an interesting scenario since the accessibility of the prior attitude would be increased, but so would the ‘false’ association tagged with that attitude. Although we know that accessibility facilitates the potential for the impact of any given attitude, the rehearsal of its rejection should have an opposing effect through its facilitating the automatic recognition of it as false when it does come to mind. To the extent that prior
attitudes do remain in tact, the accessibility of the tags marking them false might be as important for determining some behaviour as the accessibility of the new attitude—especially when the prior attitude remains more accessible than the current attitude.

**Attitude-Behaviour Consistency.** In the last three decades, a number of models have developed to account for the frequently observed discrepancy between attitudes and behaviour. As a result, when attitudes do not guide behaviour, investigators tend to cite research that describes how an attitude can sometimes become an impotent antecedent of behaviour (e.g., the attitude did not come to mind, e.g., Fazio, Powell & Herr, 1983; powerful normative or situational factors were prepotent, e.g., Fishbein & Ajzen, 1975; Fazio, 1990b; the attitude measure was susceptible to demand or social desirability, e.g., Fazio et al., 1995; the attitude was not measured at a sufficiently specific level, e.g., Davidson & Jaccard, 1979).

In contrast, the PAST model provides a complementary account of attitude behaviour discrepancy that does not diminish the power of the attitude construct. Specifically, if prior attitudes can remain potent after new attitudes have been formed, then behaviour becomes a potential consequence of both prior and current attitudes. Under conditions where prior attitudes are influential but only current attitudes have been measured, it seems unfair to attribute a low attitude-behavior correlation to an inability of attitudes to predict behaviour. The problem in this case is that the influential prior attitude was simply not assessed.

Surprisingly, we found no studies in which prior attitudes were examined for their power to predict behaviour above and beyond that of participants’ current attitudes. This again, we think, reflects the implicit assumption of the AC model in that prior attitudes
are generally assumed to be unimportant except as a literal referent to how people used to feel. However, a number of researchers have suggested that past behaviour, and in particular habit, are important determinants of current behaviour. In fact, a number of researchers have criticized attitude models such as the theory of reasoned action (Fishbein & Ajzen, 1975) because past behaviour has been found to predict current behaviour above and beyond participants' behavioural intentions (e.g., Bagozzi, 1981; Bentler & Speckart, 1979; 1981; Fredericks & Dossett, 1983; Sutton & Hallet, 1989; Eiser, Morgan, Gammage, & Gray, 1989; Whittenbraker, Gibbs, & Kahle, 1983). As a result, the unmediated influence of old behaviour is argued to weaken the importance of attitudes and consequent intentions as the primary determinants of current behaviour.

Does the predictive power of past behaviour above and beyond current attitudes and intentions weaken the attitude construct? We suggest this claim rests on the implicit assumption that the impact of previous behaviour on current behaviour is not mediated though the intact prior attitude which influenced the prior behaviour in the first place. For example, Sutton and Hallet (1989) found that following attitude change toward seatbelt use, participants' old seat belt wearing habits remained an important influence above and beyond participants' new attitudes. If the PAST model is correct, then it at least seems possible that participants' prior attitudes—and not their prior behaviour per se—could be responsible for this effect (e.g., since prior behaviour would likely have been correlated with prior attitudes). However, since prior attitudes were not assessed, we can not tell. In future research, it would be useful to examine whether past behaviour can still predict current behaviour when prior attitudes are first partialled out of the equation.
Implications for 'Implicit Attitude' Research

In their appeal for researchers to take greater advantage of implicit attitude measures, Greenwald and Banaji (1995) stated:

The evaluative content of this implicit attitude may disagree with results from a direct measure of attitude ... such disagreement, referred to as dissociation of implicit and explicit attitudes, is especially interesting and perhaps most dramatically indicates the value of the implicit attitude construct. (p. 8)

and ultimately concluded, "Methodologically, this review calls for increased use of indirect measures..." (p. 4). In line with Greenwald and Banaji's conclusion, Fazio et al. (1995) also stated that response latency techniques provide "an indirect, unobtrusive measure of attitude," and more directly, that "the priming procedure appears to be a bona fide pipeline for attitude measurement."

To evaluate the tenability of these suggestions, one needs to consider whether the PAST or AC model is assumed. We suggest that the enthusiasm for implicit measures is reasonable under the AC model but becomes much more difficult to accept if one believes that prior attitudes can sometimes remain intact. In fact, the PAST model suggests that in some circumstances, self-report measures could actually yield attitudes that would be more meaningful than those obtained through implicit measures.

If a prior attitude were to exist intact, then it too would presumably remain as information directly associated with the attitude object in memory. So if one assumes that prior attitudes can remain intact, then without knowing a person's evaluative history with the object, there is no way to know whether a response latency or other implicit measure was tapping the prior attitude, the current attitude, or a combination of current and prior attitudes. Again, this would not be an issue under the AC model, because prior
attitudes are not potential determinants of implicit responses under this model. Under the AC model, implicit measures should provide a meaningful assessment of people's 'true' attitudes under most conditions. This is because implicit measures can help to circumvent social desirability and other self-report issues. And when such factors are not relevant, implicit measures should reveal the same attitude as a self-report assessment. Again, the problem occurs only once we assume that prior attitudes might still remain intact.

Under the PAST model, implicit measures provide a meaningful assessment of people's 'true' attitudes when no prior attitudes exist, or when any prior attitudes that do remain intact do not have sufficient relative strength to cloud the assessment of the current attitude. In cases where prior attitudes remain intact and are of sufficient associative strength, implicit measures become much more difficult to interpret. In such cases, it might in fact make more sense to rely on explicit self-report measures to assess the 'true' attitude since people can indicate which evaluative association is actually endorsed. One needs only to consider a smoker of 20 years who had quit for 6 months and now holds a genuinely negative attitude toward smoking. It would be entirely reasonable to expect that 20 years of positivity toward the object could have a significant impact on the person's implicit evaluative response, yet no evidence of positivity whatsoever would appear at the explicit level. If the results of such an examination were to parallel our results, the explicit measure would reflect negativity and would correlate highly with subsequent behaviour -- at least in contexts where a person could reflect on his or her attitude. In contrast, an implicit assessment could reveal negativity, positivity or even an apparently neutral attitude if both the new negative attitude and old positive
attitude were to result in a combined impact. Although one could argue as to the
definition of what a 'true' attitude really means, this potential for confusion certainly
warrants caution when interpreting attitudes assessed at the implicit level.

Finally, as Fazio et al (1995) and many others have pointed out, when factors such as
social desirability and introspective problems are an issue, explicit self-report measures
are also likely to be problematic and not especially meaningful. Accordingly, under the
PAST model, when the individual has had a prior attitude and these factors are relevant,
then both implicit and explicit measures could fail to be highly informative.

Why Wouldn’t Attitudes Just Change?

As a final thought, we speculate as to whether some adaptive function might be served
by the persistence of prior attitudes. Upon initial reflection, one might think that our
inability to immediately replace old attitudes with new ones reflects a flaw or
maladaptive function of the human mind. After all, adaptivity requires a flexibility of
response to one's changing environment—new attitudes must be formed to replace those
that are no longer deemed appropriate or useful. However, consider that new attitudes
are typically based on new information. This new information will have contradicted the
previous cumulative experience associated with the object. As a general rule, should an
attitude which follows from cumulative experience be deleted when information at a
given moment suggests that it's wrong?

Although a new attitude holds the promise of adaptation to a current context, old
attitudes hold the benefit of reliability—they are the consequence of cumulative past
experience and knowledge. For example, consider that the new information is revealed
to be incorrect, or that it is qualified by some as yet unrealized contextual factor. If the
new evaluative association can not be relied upon, or if some experience causes us to question its validity in a critical situation, is it best to have nothing else to rely on, or to have an intact prior association that got us by as long as it did?

If prior attitudes serve the adaptive function of such a backup system, there is of course still a cost. As with information stored on a computer, a backup system creates the potential to mistake old files for new. In contrast, without a backup system, there would be no such risk. One therefore needs to weigh the potential cost of suffering the mistakes of a backup system with the potential cost of having no backup system at all. If all your attitudes were stored on your computer, would want such a backup system? Or would you prefer to delete your old attitudes whenever you created new ones? To answer this question in the 'adaptive' way, one needs to know how often and under what circumstances having this backup system would cause old attitudes to be mistaken for new. When future research can show us how and when such mistakes happen, perhaps then we will be able to better address whether the persistence of prior attitudes can be considered adaptive. For now, we are simply suggesting it would be premature to assume that such a mechanism would be maladaptive.

Conclusion

The primary purpose of the current research was to formally challenge and directly test the implicit assumption that when attitudes change, the old attitude is literally replaced with the newer attitude. We suggested that in many cases, prior attitudes might remain at least partially intact and therefore influence current evaluative responding under specifiable conditions. The data reported here seem to suggest that the PAST model remains a tenable position. Our intention was not to suggest that attitudes never
change or that prior attitudes will forever remain impactful. Rather, our intention was to demonstrate the danger of presuming that apparent attitude change necessarily reflects a literal changing, updating, or destruction of the prior attitude, and that in some cases current evaluative responding can be explained with the knowledge of how people felt in the past.

We also wanted to demonstrate that implicit measures can sometimes provide an attitudinal assessment that could be misleading if prior attitudes are not taken into account. In Study 4, participants who had experienced attitude change toward another person appeared neutral on an implicit attitude assessment despite a near perfect correlation between their relatively extreme, self-reported current attitudes and subsequent choice behaviour. We suggest that this should call into question the idea that the neutral responses evidenced by the implicit measure reflected 'true' attitudes.
ENDNOTES

1 Baeyens et al. review a much wider array of evidence supporting the permanence of conditioned responses that does not directly lend itself to the current issue and so we do not discuss it here.

2 Consistent with McClelland et al., we use the term motive here rather than trait. McClelland et al.'s conceptualization is consistent with a recent contrast discussed by Winter, John, Stewart, Klohnen & Duncan, 1998. They conceptualize motives as being the desire underlying a behaviour, which parallels the attitude construct quite nicely. In contrast, they suggest that traits can better be thought of as the consistency in one's behaviour that follows from the reliable impact of other psychological mechanisms in addition to the motive. Essentially then, a motive is what a person consistently desires, whereas a trait is what he or she consistently does.

3 As with Study 2, this study represents a partial data set from a series of experiments conducted to examine a number of issues associated with classical conditioning (e.g., the interactive roles of mere exposure and classical conditioning on attitudes toward neutral stimuli; see Jarvis, Petty & Tormala, 1998). The design of these studies was such that specific conditions were able to serve both research purposes. Methodology and results discussed here will be limited to those that were designed specifically for the current topic. This explains, for example, why the CS-UCS pairings in each study occurred in frequencies of 0, 5, 10, 20 and 40-exposures. For the purpose of the current research, the only condition of interest involved faces from the 40-exposure condition. All other CS-UCS pairing conditions were consequently categorized as filler or 'distraction' trials for the studies reported here.

4 To help accommodate any desire to withdraw, experimental credit was always given before the session began, freedom to leave at any time was stressed, and the experimenter left the room during the sessions so that participants could feel more comfortable with 'slipping out.' Two participants did decline to finish the session. The sample size of 47 does not include these two participants. Through the debriefing and feedback procedure, we learned that most participants thought the purpose of the research justified any discomfort caused by the images.

5 In the Ito et al. (1997) study, positivity and negativity were assessed separately by asking participants to rate the extent to which each of a series of adjectives described their positive (e.g., delighted, approving, pleasant) and negative (e.g., unlikable,
unattractive, unpleasant) reactions to the slide on a five-point scale ranging from very-slightly/not at all to extremely. In our own pretesting, participants were asked to rate their immediate reaction to the image on two 7-point scales ranging from not at all positive (negative) to extremely positive (negative).

6 This counter-balancing factor did not reliably interact with any of the main effects or interactions presented in the Results section (i.e., all effects were comparable for Eddie and Phil regardless of which face was conditioned with positive images and which was conditioned with negative images.

7 In a subsequent replication of this study, we used an alternative operationalization of attitude in which participants rated the men in terms of 'Overall, how do you feel about this man?' on a scale from 1 (Like very much) to 7 (Dislike very much). This operationalization was used for both the initial attitude induction as well as for the subsequent attitude change manipulation. All results were identical to those reported here.
REFERENCES


93


Byrne, D., Griffitt, W., & Stefaniak, D. (1967). Attraction and similarity of personality characteristics. *Journal of Personality and Social Psychology, 5*, 82-90.


APPENDIX A

Positive Images Used as Unconditioned Stimuli
APPENDIX B

Neutral Images Used as Unconditioned Stimuli
APPENDIX C

Negative Images Used as Unconditioned Stimuli
APPENDIX D

Images of Neutral Faces Used as Conditioned Stimuli