Changing the Strength of Implicit Associations Between Males and Science Relative to Females and Science: A Comparison of Alternative Methods

A dissertation submitted in partial fulfilment of the requirements for the Degree of Master of Science in Applied Psychology

by Rebecca Rose James

Department of Psychology

University of Canterbury

2008

Table of Contents

Table of Figures

Table of Tables

Acknowledgements

Enormous thanks go to my supervisor, Associate Professor Lucy Johnston. Her advice and support throughout the project was invaluable. She had a knack for knowing when my understanding of something was shaky (and calling me up on it!), and as a result I have come out the other side with a vastly better understanding of my project than might otherwise have been the case.

Thanks have to go to the APSY crew, for keeping me sane by going crazy right along with me! Special thanks to my officemates for not complaining about my widely scattered messes, and to those of you who worked through the late nights and weekends with me!

Apologies and thanks to my family, friends, and James, for putting up with my short fuse and constant unavailability!

Finally, a very grateful mention must be made of Statistics New Zealand, for giving me leave when it was needed.

Abstract

Two methods of changing the strength of implicit associations between males and science relative to females and science were evaluated in this study. Participants in each of two experimental conditions completed a manipulation in which they sorted photos of students into sciences and arts according to their stated major. One condition exposed participants to equal numbers of stereotypic and counter-stereotypic images, and the other exposed participants to predominantly counter-stereotypic images. Participants' bias was measured with an IAT pre-manipulation, post-manipulation, and twenty-four hours later. No significant changes in IAT scores were observed from pre- to post-manipulation. Reasons for the results are discussed, including the nature of the stereotype and methodological considerations. Finally, future research directions and implications for career decisions are considered.

Women are under-represented in a number of occupational fields, primarily science- and math-based areas. This differential representation is not entirely due to sex differences in ability (Baker & Jones, 1993). One explanation for the difference may be the impact of the prescriptive stereotype that science careers are not appropriate careers for women. Implicit stereotypic associations have been shown to be predictive of behaviours, and are more impervious to faking or issues of social desirability than explicit attitudes (Poehlman, Uhlmann, Greenwald & Banaji, 2005). As sex-based stereotyping is affected by self-presentation biases, studying the strength of implicit stereotypic associations may be useful. Several methods intended to change the strength of implicit stereotypic associations have been successfully used in the past. Two notable examples that have never been directly compared include flooding the individual with counter-stereotypic information (e.g. Dasgupta & Greenwald, 2001), and reducing dependence on categorical information by equal exposure to stereotypic and counter-stereotypic information (e.g. Plant, Peruche, & Butz, 2005). The present research compares the relative effectiveness of these two methods in reducing the implicit association of males and science relative to females and science.

Differential Participation in Science and Math

There is a clear disparity between the sexes in participation rates in areas like physics, engineering, math, and computer science. The US Census (1990) reported that males comprised 85% of architects, 92% of aerospace engineers, 87% of physicists, and 94% of nuclear engineers. In New Zealand, a report by the Ministry of Research Science and Technology (1998) revealed that females made up only 18.9% of all scientists and engineers in 1996. Women were more prevalent in health professions (35.4%) and mathematics/statistics (32.9%), and least represented in engineering (7%) and computing (21.6%).

In an educational context there is a similar participation gap. In US colleges, males make up 66% of physical science, 65% of math and computer science, and 84% of engineering enrolments (Nosek, Banaji & Greenwald, 2002). In New Zealand in 2003, female enrolments in tertiary science equalled that of males in natural and physical sciences, but females were underrepresented in areas like engineering, architecture, and agriculture (Ministry of Research Science & Technology of New Zealand, 2003). For example, in 2006, males obtained 88% of tertiary engineering qualifications (Tertiary Education Statistics, Education Counts NZ, 2006). In fact, differential participation rates often exist as soon as science and math courses become non-compulsory. US research shows male high school students are more likely than female students to take technology, physics, and math papers, with females more likely than males to take modern languages, English and other arts papers (Burr, 1998). In New Zealand, females are similarly more likely than males to take languages, arts, and history, and males are more likely than females to take calculus, physics, and technology (Education Counts New Zealand, 2007).

Differential participation is an issue for various reasons. Diversity in the workplace is a hot topic for businesses today – a survey of 400 human resource professionals in the United States revealed that 76% of the organizations they worked for had specific diversity practices in place, and two-thirds provided training on diversity-related issues (SHRM Report, 2005). The relationship between diversity and organizational performance is not straightforward, but it does appear that diverse organizations can achieve better performance than homogenous organizations under some conditions (e.g. if the organizational culture is inclusive and emphasizes collectivism over individualism; Chatman, Polzer, Barsade & Neale, 1998). As demonstrated above, many science fields are relatively homogenous by virtue of being predominantly male, so would benefit from increased diversity in terms of sex. Interest in science fields is also decreasing generally, meaning that organizations dealing in science must seek out new potential groups of employees

(OECD, 2006). As women are increasing their participation in the workforce (Statistics NZ 2003, 2005), they are a viable source of new employees. Increasing the proportion of women in science fields could facilitate an increase in science participation in general as well as allowing previously homogenous workplaces to reap the benefits of diversity.

Underlying Reasons for Differential Participation

It is generally accepted that environmental factors like stereotyping (rather than simply biologically determined ability) have a strong influence on females' career choices (Baker & Jones, 1993; OECD, 2006). Sex-based stereotyping tends to have a prescriptive component, involving an expectation that group members ought to exhibit traits and behaviours that conform to the stereotype, and a value judgment about individuals who do or do not conform (Prentice & Carranza, 2002). That is, people who conform to sex-based stereotypes are rewarded (by acceptance, positive evaluation and so on), and those who violate the stereotypes are punished (by discrimination, negative evaluation, and so on). This social conditioning teaches individuals what behaviours, traits, and occupations are appropriate for their sex (Prentice & Carranza, 2002). There are a number of areas where the effect of the stereotype associating males with science can be seen (Correll, 2001).

Early in life, individuals begin to experience reminders that science is an inappropriate career choice for women. For example, Crowley, Callanan, Tenebaum and Allen (2001) found that parents were three times more likely to explain science museum exhibits to male children than to female children. In addition, Tiedemann (2000) found that when parents endorsed the stereotype that males are better at science than females, they and their daughters both believed the daughter had less math ability. All sorts of media can serve as reinforcement for the stereotype, including software programs, posters on classroom walls, television programs and advertisements (for a meta-analysis see Oppliger, 2007).

When the sex-gap in participation becomes too extreme, occupations can become sex-typed, and this can lead to the opposite sex being less interested in and less likely to enter that occupation (Heilman, 1979). This is the case with many science and math based fields, which are sex-typed as male (Greene, Sullivan & Beyard-Tyler, 1982). Nosek (1999) argued that the avoidance of opposite sex-typed occupations occurs because of a conflict between gender identity, stereotype endorsement, and interest in science. Most women develop a strong association between self and female long before they have to choose a field of study (Crombie, Sinclair, Silverthorn, Byrne, DuBois & Trineer, 2005). In addition, the stereotype that males are better at science than females creates a strong association between males and science, and a correspondingly strong, but negative, association between females and science (Nosek, 1999). This association also develops early, in children as young as five to eight years old (Stroeher, 1994, Miller & Budd, 1999). The third component of this situation is the strength of association a female develops between herself and science. A woman with a strong female-self association, and a strong negative female-science association, will have a weak or negative self-science association to maintain consistency. This is based on Heider's (1959, as cited in Nosek 1999) balance theory, and has been supported by Nosek et al. (2002). This suggests that some females may choose to avoid science due to beliefs about sex rather than interest. This is a critical area where stereotype reduction research can play a role. If the association between males and science can be weakened, a benefit may be that females no longer have to avoid science to maintain the balance between gender identity and the association of sex and science.

Another area where the influence of the stereotype can be seen is in situations that elicit stereotype threat. Stereotype threat occurs in performance situations when an individual is aware of stereotypes that predict their poor performance (Schmader, Johns & Barquissau, 2004). It has been shown that when stereotypes about females' poor math performance are made salient, females perform worse on math tests (Spencer, Steele & Quinn, 1999; Davies & Spencer, 2005) and show less interest in future participation in math (Davies, Spencer, Quinn & Gerhardstein, 2002). If mere awareness of the stereotype can create these effects on female performance and interest in participation, it follows that decreasing or inhibiting activation of the stereotype, even temporarily, may alleviate these effects. Research into short-term stereotype change may thus have practical implications for stereotype threat reduction.

Those females who do choose science careers may experience sexual harassment and discrimination as a result. A number of researchers have shown that women in predominantly male occupations and workplaces are sexually harassed more than women in balanced or predominantly female occupations / workplaces (Fitzgerald, Drasgow, Hulin, Gelfand & Magley, 1997; Glomb, Munson, Hulin, Bergman & Drasgow 1999; Mansfield, Koch, Henderson, Vicary, Cohn & Young, 1991). Harassment creates working conditions that have been linked to work disruption, loss of job motivation, dread of work, and turnover intentions (Jensen & Gutek, 1982), which suggests that females may be motivated to leave male dominated careers and organizations. The relevance of this to the study of stereotypes should be clear. Women who violate sex-based stereotypes such as the male-science one are more likely to experience sexual harassment than are stereotype-conforming females. Stereotype change research then becomes an important part of reducing the incidence of sexual harassment in the workplace. If science is made into a sex-neutral occupation, then women should not be discriminated against or harassed for choosing that occupation.

Stereotype Change Research

It is evident that there are very good reasons why more women should enter science careers, and several stereotype-based reasons why they do not. Stereotyping can be beneficial as it saves time and effort in interpersonal interactions, by allowing inferences to be made about people based on their group membership (Patterson & Bigler, 2006). However, this particular stereotype is problematic because it influences the career choices of women in restrictive ways. Until the stereotype is lessened or eliminated, there will still be barriers to entry into science careers for women. It is important to conduct research on how this stereotype might be changed or its impact on women reduced. The examples above highlight some of the varied areas that stereotypes affect, including sex-typing, stereotype threat and sexual harassment. Research into how best to change stereotypes may therefore be an important step towards encouraging more women into science.

Stereotypes can be measured explicitly and implicitly (Rydell, McConnell, Mackie, & Strain, 2006). An individual is aware of their explicit stereotypes, and can manipulate their expression. Most measurement of explicit stereotypes consists of asking individuals to self-report their level of endorsement of a particular stereotype. Because individuals can manipulate the expression of their explicit stereotypes, their expression is often subject to a number of moderating factors, including self-presentation biases (Gregg, Banaji, & Seibt, 2006). In studies looking at socially sensitive topics, individuals may report less stereotype endorsement than they actually have, which will affect the ability to predict behaviour based on expressed explicit attitudes. In contrast, individuals cannot directly report or manipulate the implicit expression of stereotypes, often because they are unaware of the strength of their implicit associations (Fazio & Olson, 2003). Because of this, indirect measurement techniques may be a better means of capturing the strength of an individual's stereotypes, especially for stereotypes that are subject to social desirability forces. Implicit and explicit attitudes correlate quite highly under some conditions (e.g. domains that are not socially sensitive, such as a preference for flowers versus insects, Greenwald, McGhee & Schwartz, 1998). Further (and discussed below), implicit measures of stereotyping have been shown to be predictive of behaviours beyond those predicted by explicit stereotype endorsement. Implicit measures of stereotype strength therefore have valuable research implications for topics like sex-based stereotypes, and for that reason, the remainder of the discussion here focuses on implicit stereotypes.

There is a wide range of measures of the strength of implicit stereotypes, but probably the most widely known and frequently used measure today (Fazio & Olson, 2003) is the Implicit Associations Test (IAT; Greenwald et al 1998). The IAT traditionally measures the strength of a person's associations between two contrasting categories and an attribute dimension (e.g. the concepts of Blacks and Whites and a positive/negative attitude), but has also been used to measure the strength of association between two sets of categories, such as male/female and arts/science (e.g. Nosek et al 2002). That is, an IAT might measure how strong an association a person has between males and science and females and arts, relative to how strong their association is between females and science and males and arts. This strength of association is based on response times to stereotypic and counter-stereotypic information. An IAT score then represents the 'bias' a person has, where a larger score represents a larger difference in relative association strength between the two alternate pairings of categories. Greater detail about the procedure of the IAT is described in the method section of this thesis.

IATs are popular because they are easy to use with a variety of topics, have good reliability and are particularly difficult to fake on (Poehlman et al 2005; Nosek, Smyth, Hansen, Devos, Lindner, Ranganath et al, 2007). Cunningham, Preacher and Banaji (2001) and Nosek and Banaji (2001) found strong correlations between IAT scores and various other established implicit stereotype measures. The relationship between implicit and explicit attitudes and stereotype endorsement is moderated by a number of variables such as self-presentation and strength of endorsement (Nosek, 2005). However, a large study by Nosek (2005) examining the relationship between IAT scores and explicit stereotype endorsement in 57 content domains (including relative liking of presidential candidates Gore versus Bush, hip hop versus classical music, and Whites versus Blacks) found quite strong correlations in some domains, and a reasonably strong average correlation across all 57 domains.

A recent meta-analysis of implicit attitude studies showed that scores on an IAT predicted a wide range of outcomes including social judgments, physiological responses, and social behaviours (Poehlman et al, 2005). Specific examples from this meta-analysis include measures of implicit racial stereotypes predicting individuals' preferences for a White over a Black partner on an intellectual task (Ashburn-Nardo, Knowles & Monteith, 2003) and negative nonverbal behaviours towards a Black confederate (McConnell & Leibold, 2001). There are not many studies specific to the stereotype that males are better suited to science than females. However, Nosek et al. (2002) found that stronger stereotypic IAT scores were associated with higher relative performance on the standardized US college admission tests (SATs) for math ability than for verbal ability for males, but worse relative performance for females. In addition, Kiefer and Sekaquaptewa (2007) showed that for females, having a stronger implicit association of males and science relative to females and science predicted worse performance on the final exam for a calculus course and less interest in a math career. These findings held even after controlling for SAT scores and prior course performance. Explicit stereotype endorsement did not predict performance or career interest.

Research on a direct link between changing implicit attitudes and resulting changes in behaviour is sparse. In a series of studies, Kawakami, Phills, Steele, and Dovidio (2007) established that a training program, which required participants to push a joystick away in response to an image of a White individual, and pull the joystick towards them in response to an image of a Black person, resulted in lower implicit negative attitudes towards Blacks. They then showed that after participating in the training task, participants showed less stereotyped nonverbal behaviours in an interaction with a Black confederate than did a control group. The combination of these findings suggests that the training task may have changed participants' implicit attitudes and thus their behaviour, however no firm conclusions can be made as this link was not directly established.

In general, Poehlman et al. (2005) showed that both explicit and implicit attitudes successfully predicted behaviours in a number of areas. Explicit attitudes were a better predictor than implicit attitudes in areas like political preferences and brand-related choices, but implicit attitudes were a better predictor of stereotype related behaviour. The predictive value of explicit attitudes was weaker in socially sensitive situations and where predicted behaviours were hard to control, whereas implicit attitudes predicted equally well across socially sensitive and non-sensitive areas, and across a range of controllable and uncontrollable behaviours. The above evidence shows that the strength of implicit stereotypic associations relative to the strength of counter-stereotypic associations is measurable by IATs, and that measure is predictive of a range of behaviours. These findings, especially the studies that establish predictive validity of the male-science stereotype, support the notion that changing implicit stereotype strength is an important part of opening the field of science to females.

Both stereotype malleability (temporary change) and permanent or long-term change are important areas of research. Reducing or eliminating the stereotype permanently could allow people to choose their career based on interest rather than beliefs about the appropriateness of that career for their sex. This may involve other mechanisms beyond what is required to temporarily change stereotypes. However, progress on complete elimination of the stereotype is likely to be slow. Spence and Buckner (2000) found very similar endorsement of sex stereotypes about instrumentality and expressivity in 1974 and 1996, despite undeniable cultural and legislative changes in that time period. Stereotype malleability is useful in the interim as it creates an immediate change in the stereotype, which may result in temporary behaviour changes. Most research has focused on immediate outcomes – for example, Blair's (2002) review of fifty-one studies of stereotype malleability included only three studies that measured implicit attitudes a day or more later. In circumstances where stereotype minimisation is important (e.g. in a job interview or where stereotype threat is an issue), methods of temporary change may be of great value. Two successful approaches to affecting the expression of implicit stereotypes in the short-term will be discussed below and used in the present research.

Counter Stereotypic Exposure

The first approach considered in the present research is the use of counter-stereotypic exposure to change the relative strength of stereotypic and counter-stereotypic associations. Following the environmental association model (Karpinski & Hilton, 2001), implicit attitudes are considered to be a result of the associations an individual experiences throughout their life. These associations can come in various forms, from seeing more male than female scientists, to hearing more stereotypic than counter-stereotypic comments, jokes, and so on. An individual who experiences more male-science pairings than female-science pairings will thus have a stronger implicit stereotype than an individual who has had a more stereotype-neutral set of experiences. Following this logic, flooding an individual with counter-stereotypic experiences until the number and strength of associations becomes balanced ought to eliminate the stereotype. Empirical evidence has generally been favourable towards this method, and several examples are given below to demonstrate its success.

Karpinski and Hilton (2001) used the stereotype that older people are viewed more negatively than young people. They gave participants an IAT to measure implicit negative attitudes about the elderly relative to youth, showed participants a series of either all counter-stereotypic or all stereotypic word pairs, and then gave them a second IAT. The IAT scores of participants who had been exposed to stereotypic pairings did not change over time, but IAT scores for those exposed to counter-stereotypic information decreased significantly from baseline to after the manipulation. This suggests that the strength of their implicit negative associations towards the elderly had been at least temporarily reduced.

Dasgupta and Greenwald (2001) assessed whether exposure to counter-stereotypic images reduced implicit negative attitudes towards Blacks, and whether the reduction would persist a day later. This is one of the few studies that have looked at whether a change in stereotypes persists over any length of time. They put participants in one of three groups – a pro-White group that saw images of admired Whites and disliked Blacks, a pro-Black group that saw images of admired Blacks and disliked Whites, and a control group that saw images of flowers and insects. After repeated exposure to the images, participants' implicit negative attitude towards Blacks relative to Whites was measured with an IAT. Results showed that immediately after exposure, the pro-Black group had significantly lower negative attitudes towards Blacks relative to Whites than the control group, which had significantly lower negative attitudes towards Blacks relative to Whites than the pro-White group. Twenty-four hours later, there was no significant difference between the control and pro-White groups, but the pro-Black group maintained significantly lower negative attitudes towards Blacks than both of the other groups.

There is minimal research on changing sex-based stereotypes using counter-stereotypic exposure. However, Dasgupta and Asgari (2004) examined the effect of counter-stereotypic exposure on the negative association between females and leadership in two studies, one short-term and one over a longer period. Participants in the first study were exposed to pictures and descriptions of either flowers (control) or famous women leaders (counter-stereotypic exposure). Participants then completed an IAT measuring the strength of association between women and leadership compared to between women and supportiveness. Participants exposed to counter-stereotypic information were faster at responding to counter-stereotypic pairs within the IAT than were the control group, and were faster at responding to counter-stereotypic pairs than to stereotypic pairs. This study suggests that the stereotype of women being more strongly associated with supportiveness than with leadership can be manipulated by repeated counter-stereotypic exposure in a laboratory based manipulation.

The second study was a longitudinal comparison of a group of female-only college students with a group of female students at a coeducational college. Participants completed an IAT measuring the strength of association between women and leadership two months into their first year of college and another IAT a year later. At time 1, both groups showed a stronger association between males and leadership than between females and leadership. After a year, participants from the female-only college had reduced their bias so they showed no differences in the association between males and leadership relative to females and leadership, and the co-ed college students had an even stronger association between males and science relative to females and science than they did at time 1. Dasgupta and Asgari (2004) conducted a series of regressions on their data, and found that the amount of exposure to female faculty a student had significantly predicted their IAT scores at time 2, such that more exposure to female faculty predicted lower bias on the IAT. This factor completely mediated any effect of college type or choice of courses on IAT scores at time 2.

Most, if not all research that has employed counter-stereotypic exposure to create change in implicit associations has used 100% counter-stereotypic information. Plant et al. (2005) believed that proportions other than 100% counter-stereotypic exposure could successfully affect the expression of stereotypes. Their rationale was that while effective, the counter-stereotypic method was flawed by its continuing emphasis on the relationship between the category (e.g. sex) and the stereotyped trait (e.g. science aptitude). Plant et al.'s (2005) proposed solution to this dilemma is the second change manipulation discussed and used in the present research.

Equal Frequency Exposure

Plant et al. (2005) proposed that if an individual was engaged in a task where race (or other categorical variables) did not aid performance, they would either not activate or would inhibit the activation of the category. Their proposed method was a stimulus-sorting manipulation that removed the association between the category (race) and the trait (violence) by using a number of stimuli that were evenly divided between stereotypic and counter-stereotypic in nature. Race was therefore irrelevant to performance on the task, and participants ought to have been motivated to ignore it. Plant et al. (2005) proposed that improvement on the manipulation might be a result of one or both of two possible routes.

The first explanation was that participants who endorsed the stereotype would probably initially notice the race of the target individuals, and respond in a stereotypic manner. That is, they would be slower to respond to, and would make more errors on, counter-stereotypic pairs compared to stereotypic pairs. Attention to race and the subsequent influence of stereotypes on response patterns are not, however, inevitable. The application of a stereotype depends on the individual's goals and cognitive resources, among other things (Kunda & Spencer, 2003). Bodenhausen and Macrae (1998) argued that individuals will disregard or inhibit interfering information in person perception tasks. Kunda and Spencer (2003) also reviewed evidence that attention to categories or stereotypes can either be inhibited initially, or attention can dissipate over time as required by the individual's goals and circumstances. Participants were told to perform as quickly and accurately as possible. In addition, participants were made aware of their accuracy by an error message that appeared after every mistake. Responding in a manner that was influenced stereotypically by the race of target individuals would hurt performance on the manipulation, and therefore individuals would inhibit (or not attend to) the category of race to improve performance. Plant et al. (2005) predicted that participants in their experiment might initially show stereotyped response patterns, but participants would successfully ignore race after some practice on the experimental manipulation.

The second explanation they proposed was that an equal frequency manipulation might affect the strength of stereotypic associations in the same way that a counter-stereotypic manipulation would. Counter-stereotypic exposure is thought to work by partially counter-balancing the previous imbalance in an individual's total associations. Plant et al. (2005) believed that for an individual with a set of stereotyped associations that was larger than their set of counter-stereotypic associations, exposure to equal frequencies of stereotypic and counter-stereotypic pairs might also partially balance out the total of previous associations. If this were the case, however, the equal frequencies method might be expected to be somewhat less effective than the counter-stereotypic method, as both the proportion and absolute number of counter-stereotypic associations provided would be less.

The stereotype that Plant et al. (2005) assessed was that Blacks are more likely to be violent criminals than Whites are. They conducted a series of four related studies, which all used one or more variations of a basic manipulation program. The basic program used in all studies displayed a series of pictures of a Black or a White face, with either a gun or a tool superimposed. These combinations represented either violent criminals or innocent bystanders, and were either stereotypic (innocent Whites and violent Black criminals) or counter-stereotypic (innocent Blacks and violent White criminals). Participants were instructed to press one button to 'shoot' faces with guns superimposed, and a second button to 'not shoot' faces with tools superimposed. The four combinations occurred equally often so there was no benefit for participants who used race as a diagnostic cue. If participants used race as a cue, they would make more errors and be quicker to shoot innocent Blacks and not shoot violent White criminals than the reverse.

In the first study, all participants completed 160 trials of the basic program. This study was intended to assess whether biases existed, and whether these biases could be reduced with the basic program. In the first half of trials, participants were significantly more likely to shoot at Black than White targets holding tools, and more likely to choose not to shoot at White than Black targets holding guns – that is, they responded in a stereotypic manner. This effect disappeared by the end of the experiment so that there were no significant differences in reaction time or error rate between the four pair types in the second half of the experimental trials.

The second study assessed whether the influence of the manipulation would persist over time. Half the participants completed the basic program, and then returned to the laboratory twenty-four hours later to complete the basic program again. The second half of participants completed a similar program on the first visit, where participants had to 'swat' insects and 'not swat' birds. These participants also returned to the lab twenty-four hours later and completed the basic program. Participants who completed the basic program at both times showed bias on the first half of the first visit, but their bias had been eliminated by the second half of the first visit, and they remained non-biased on both halves of the second visit. Participants who completed the insect/bird program on the first visit showed bias on the first half of the second visit, which was not exhibited on the second half of the second visit. This showed both that the reduction in bias persisted over time, and that mere practice with a similarly structured program did not result in reduced bias 24 hours later.

In a third study, Plant et al. (2005) then assessed whether improvements in performance were the result of practice effects with the format of the computer program. Participants completed the same program but with 70% stereotypic and 30% counter-stereotypic pairs. If effects in the first two studies were simply due to increased familiarity with the shoot/don't shoot principle of the program, this group should have shown a similar pattern of improvement. There were no significant differences in response time between the first half and second half of trials, and participants made significantly more errors on the second half than they did on the first half. Participants made more errors on counter-stereotypic pairs than on stereotypic pairs at both times. These results indicate that it is not simple practice with the program's format that is causing the elimination of bias.

The fourth study examined whether decreases in bias were due to the category of race not being activated by testing category activation immediately after the computer program. Participants completed either a full version of the basic program (160 trials), a short version (40 trials) or did not complete the program at all. All participants then filled out a word completion task that included 18 word fragments, 10 of which could be completed with race related words (e.g. R_C_, _ _ACK). Participants who had undergone the full session responded with fewer race-related words than any other group. This suggests that the category of race was either not activated at all, or was inhibited during the course of the manipulation. The short version and control groups were not significantly different in terms of the number of race-related words produced. These results, however, do not allow conclusions to be drawn about effects of equal frequencies exposure on stereotype strength.

Plant and Peruche (2005) had police officers complete the same manipulation, and showed the same pattern of results – stereotyped response patterns on the first half of trials, and successful elimination of this response pattern on the second half of trials, so that participants took equal time and made equal numbers of errors on stereotypic and counter-stereotypic trials alike. Peruche and Plant (2006) applied the equal frequency method to the stereotype that Blacks are more athletic than Whites and found initial racial stereotyping in terms of more errors on counter-stereotypic trials than stereotypic trials, which disappeared by the end of the experiment. These studies, and that of Plant et al. (2005), are the only ones which have explicitly tried to use this method to reduce bias, and for the most part have only looked at performance on the experimental manipulation itself (with the exception of Study 4 from Plant et al, 2005). None of the experiments assessed whether participants' stereotype endorsement or implicit associations had changed as a result of the manipulation. However, Plant et al. (2005) believed that their manipulation could potentially reduce the strength of association between categories in the same way that the counter-stereotypic method does. Unfortunately, none of the three studies discussed here included an appropriate measure to test that possibility. Further studies must be done to replicate these successes and extend understanding of what is happening to participants' stereotypes after undergoing the manipulation. In addition, while Plant et al.'s (2005) logic for using equal frequencies rather than counter-stereotypic frequencies is reasonable, their comparison of the two manipulations has not been tested empirically. Empirical evidence is needed to suggest that one is truly superior to the other (or that both work equally well). Research that directly compares the two will enable further understanding of each and allow recommendations to be made for which is superior.

The Present Research

The present research compared the effects on implicit stereotypic associations of Plant et al.'s (2005) equal frequencies manipulation with the effects of a counter-stereotypic manipulation, using the stereotype that males are more strongly associated with science than females. The IAT was used to measure relative strength of implicit association of stereotypic and counter-stereotypic category pairs, allowing changes to be compared between conditions.

The experimental manipulation consisted of a series of male and female faces paired with science and arts subject titles. This manipulation is referred to hereafter as the subject-sorting task (SST). Participants sorted each pairing into arts and science categories based on the subject in the label. The SST involved 240 trials (120 unique photos each viewed twice). Performance on this manipulation, like Plant et al.'s (2005), shows the influence of stereotypes (or lack of influence). The equal frequencies proposed by Plant et al. (2005) formed one condition of the experiment (referred to as the 'equal frequencies' condition or EF). Participants in this condition saw equal numbers of stereotypic and counter-stereotypic pairings. A second condition was included where participants saw 70% counter-stereotypic pairings (referred to as the 'counter-stereotypic condition, or CS). Previous studies have usually employed exposure to entirely counter-stereotypic information, but a 70/30 split was chosen instead to create an increased likelihood that participants would believe the labels. That is, it was expected that a group of students where 30% were behaving in a stereotypic manner would be more realistic than a group where every student was behaving in a counter-stereotypic manner.

Response latencies and error rates were recorded for both stereotypic and counter-stereotypic pair types throughout the SST. Plant et al. (2005) used a median split in their analyses. However, the present research uses the first third (80 trials) and the last third, discarding the middle third. We were interested in two distinct performances – participant performance at the beginning of the experiment, and participant performance at the end of the experiment, after they had been exposed to the entire manipulation. As we were interested in two distinct end points, it made the most sense to capture the end points and exclude as much of the transitional learning phase as possible.

An IAT was used to measure the relative strength of association between stereotyped concepts initially, immediately after the SST, and twenty-four hours later. A third group acted as a control and completed three IATs at the same intervals but without completing the SST.

Hypotheses

Hypothesis 1: Participants will show a stronger association between males and science than females and science on the first IAT and will show stereotyped responding on the first third of the SST.

That is, all participants will have significant bias on the first IAT, and all participants who complete the first third of the SST will take longer to respond to, and make more errors on, counter-stereotypic pairings than they will on stereotypic pairings.

Hypothesis 2: Participants' IAT scores will be reduced by both versions of the SST.

The counter-stereotypic method has been successful in past research, and Plant et al. (2005) postulated that equal frequency exposure would work similarly. It is therefore expected that both methods will be effective in the current research. Scores on the second IAT should be less biased than the first IAT scores for both experimental conditions. In addition, the control condition's first and second IAT scores should not differ from each other. The second IAT score for the control condition should be significantly more biased than the second IAT scores for either of the experimental conditions.

Hypothesis 3: The counter-stereotypic condition will be a more effective method for reducing IAT scores than the equal frequencies condition.

The equal frequency method has been shown to potentially inhibit the concept of race, and its effectiveness at reducing IAT scores has not been assessed. Plant et al. (2005) argued that their manipulation could have reduced endorsement of the stereotype as well as inhibiting the category of race, but they did not measure stereotype change in their studies. They proposed that if the equal frequencies method was to change the strength of the stereotype, it would do so through the same route as the counter-stereotypic method. As discussed above, if this is the case the equal frequencies method is likely to be a weaker version of the counter-stereotypic method (due to the lower absolute number of counter-stereotypic associations). While Plant et al. (2005) believed their method was potentially more useful than counter-stereotypic exposure for reducing the impact of stereotypes through category inhibition, they did not measure or postulate how effective the method would be for reducing implicit bias. Further, the IAT forces participants to think in terms of categories as it provides the categories of interest, so category inhibition should not affect IAT scores.

This hypothesis will be assessed by comparing the second IAT scores between the two experimental conditions, assuming that initial IAT scores are not significantly different between the two. Additionally, response time and error rates on counter-stereotypic pairs of the late block of the SST will be compared between the experimental conditions.

> **Hypothesis 4**: The effect of the experimental manipulations will be maintained over the 24 hours, and bias will still be reduced on the third IAT.

Following the findings of Dasgupta and Greenwald (2001), whose participants showed similar scores immediately after a counter-stereotypic intervention and 24 hours later, it is expected that the scores will remain stable from IAT 2 to IAT 3 for all three conditions. The control condition is expected to have stable scores across all three IATs, and the experimental conditions should each have similar IAT scores at time 2 and 3.

> **Hypothesis 5**: There will be a positive correlation between the change in counter-stereotypic response times on the SST across time and the change in IAT scores from pre-SST to post-SST.

The SST was designed for this study in an attempt to create a manipulation that reduced the bias of IAT scores. It is hypothesized that this will be successful. Logically, if participating in one of the SST versions results in lower IAT scores afterwards, changes on each will be correlated.

Method

Participants

Participants were 100 students from the University of Canterbury, who volunteered to participate in return for a $5 voucher and entry into a prize draw. There were fifty-nine female and forty-one male participants. The mean age was 23 years (ranging from 18 to 52 years). Participants were from a range of disciplines – 35 science (19 females, 16 males), 31 arts (21 females, 10 males), 15 commerce (13 females, 2 males), 1 law (female), 12 engineering (all male), and 4 doing both a science and an arts degree (3 females, 1 male). The average university tenure was three years.

Materials

Implicit Association Test

Implicit associations between males and science relative to females and science were tested using a custom-written PC version (Walton, 2003) of the Implicit Association Test (IAT; Greenwald et al 1998). In this version of the IAT the stimuli came from two sets of categories – subject domain (arts/science) and sex (male/female). There were ten words for each category (see Appendix A for a list of all 40 target words).

The target words for the science and arts categories were determined by a pilot study. Six students from the University of Canterbury were given a list of all the arts and science subjects offered at the University (except subjects offered as both arts and science). The students were instructed to select the ten subjects that they thought were most obviously arts subjects and the ten which were most obviously science subjects. The subjects selected for inclusion in the experiment were those with the highest agreement by the six students. Five subjects were selected by all students, six by five students, five by four students, and two by three students. Two subjects that were not in the top-ranked list (math and computer science) were included, as much of the previous literature has focused on these two subjects in particular. They replaced cell biology and evolutionary biology as there was already adequate representation of the biological sciences and because previous literature has shown that the biological sciences are more equally populated by males and females.

The IAT consisted of seven blocks of trials. The category labels appeared in the upper left and right hand sides of the screen during all trials. On each trial a target word appeared in the centre of the screen. The participant was instructed to press one key (on the left-hand side of the keyboard) if the word presented belonged to the category shown on the left of the screen, and a second key (on the right-hand side) if the word belonged to the category on the right of the screen. If the participant pressed the correct button a green tick appeared below the centre word and then the next word was displayed. If the participant pressed the wrong button a red cross appeared and the participant had to press the correct button in order to move on to the next word. For each target word the response latency for the first key press was recorded, as well as whether the correct key was pressed first.

In three of the seven blocks (blocks one, two, and five) only one set of category labels appeared onscreen –either males/females or arts/sciences) and the words presented in the centre of the screen came only from that set of categories. In the remaining blocks both sets of category labels were shown and words from both sets of categories appeared during these blocks. For example, male and science appeared on the left hand side, sharing a response key, and female and arts on the right, again sharing a response key. For two consecutive blocks, the pairs were stereotypic (male and science on the same side, female and arts on the same side), and for the other two blocks the pairs were counter-stereotypic (female and science on the same side, male and arts on the same side). Blocks 3 and 6 were practice blocks and Blocks 4 and 7 the critical blocks. Following recommendations by Greenwald, Nosek, and Banaji (2003), all four of these blocks were used to generate the overall IAT score. The short practice blocks (3 and 6) contained 40 trials in each block, and the critical blocks were twice as long, with 80 trials each. To minimize order effects, half the participants completed the stereotypic blocks first followed by the counter-stereotypic blocks (SCS version), and half completed the counter-stereotypic block first (CSS version).

Subject-Sorting Task

The subject-sorting task (SST) was conducted using custom-written computer software (Walton, 2007). Participants were shown images of faces labelled with subject majors, and were asked to classify the subject into one of two categories – sciences and arts. The images (a face with a subject major beneath it; see Appendix B for examples) were presented one at a time in the centre of the computer screen, in a unique random order for each participant. Throughout the task the subject domains (arts/sciences) were displayed in the upper right and left hand sides of the screen. For each image the participant had to respond by pressing one of two buttons to indicate which domain (arts/science) the subject major belonged to. If the participant pressed the correct button a green tick appeared and the next image was displayed. If the participant pressed the wrong button a red cross appeared and the participant had to press the correct button in order to move on to the next image. For each image the program recorded the response latency for the first key press and whether the correct key was pressed first.

A practice block consisted of ten black and white photos of individual faces, five male and five female, taken from a dataset used in previous research at the University of Canterbury (Johnston, Arden, Macrae & Grace, 2003). One hundred and eleven colour photos of individual faces (55 male, 56 female) previously used in research at the University of Canterbury (Brinsmead-Stockham, Johnston, Miles, & Macrae, in press) were used for the experimental manipulation, along with nine extra photos of student taken against a similar background, bringing the total number to 120. The photos were resized and edited to include a subject major underneath the photo. The 20 subjects (10 arts and 10 sciences) used in the IAT were also used in the SST. These were randomly allocated to pictures – the method for subject allocation differed for each version of the SST created, and is described below.

Two versions of the SST were created – one where target sex and subject domain were statistically unrelated (SST-EF) and one where target sex and subject domain were related in a counter-stereotypic way (i.e. a higher proportion of arts than science subjects being paired with photographs of males and vice-versa with photographs of females; SST-CS). For each condition, two versions of the SST were created. This was to ensure that error was not due to specific faces being more readily classifiable into one or the other subjects (e.g. a feminine male face may be more easily sorted into arts than a masculine male face).

SST-EF

The SST-EF contained equal frequencies of all combinations – 30 males in science, 30 males in arts, 30 females in science, and 30 males in arts. For Version 1 the 120 faces were randomly allocated to either science or arts and then randomly assigned a specific subject within that domain. To create Version 2 the set of 'science' faces from Version 1 were randomly allocated an arts subject, and the 'arts' faces from Version 1 were randomly allocated a science subject.

SST-CS

In this condition, 70% of the male faces were paired with an arts subject and 30% with a science subject, and for females 30% of the faces were paired with an art subject and 70% with a science subject. For Version 1 the faces were randomly split into arts and science groups with the correct proportions for each sex. Faces in the science group were randomly assigned a science subject, and in the arts group faces randomly received an arts subject. Version 2 was created by randomly selecting 18 male/arts pairs and 18 female/science pairs and swapping subjects between these faces and the faces that had stereotypic subjects in Version 1. This ensured that the majority of faces were counterbalanced.

Procedure

Participants were tested individually at a time that they could attend on two consecutive days. Upon booking their time, they were assigned to one of the three conditions (equal frequency, counter-stereotypic, or control), one of the two versions of the SST if relevant, and one of the two versions of the IAT. This allocation to condition was controlled to ensure that there were equal numbers of males and females in each condition, and equal numbers of participants in each condition in each version of the SST and the IAT.

Upon arrival, the participant was given a copy of an information sheet to read, a demographics sheet to complete and a consent form to sign. They were not given any verbal instructions on how to complete the tasks, simply told that all instructions would be presented on the computer screen. The instruction screens were identical across all versions of the IAT. The instruction screens for the four versions of the SST were also all identical. Both sets of instructions can be seen in Appendix C.

Each participant completed the first IAT while the experimenter waited outside. Once the first IAT was completed, if the participant was in the control condition they were told that the program was temperamental, and that their data had to be saved right away, so they would have to wait a few minutes while the experimenter took the laptop out of the room to save their data. They were then left alone in the room for five minutes. If the participant was in either of the experimental conditions, once they had finished the first IAT, the experimenter set up and began the relevant SST version. The participant was told to read the onscreen instructions and to let the experimenter know when they had finished. The experimenter then left the room while the participant completed the SST.

After either the SST was completed or the five minutes waiting were over, the experimenter started the same version of the IAT again for the participant, explained that it was a similar task to the previous ones, and left the room again. When the participant finished the second IAT, they informed the experimenter. The experimenter thanked them for participating and reminded them to come back twenty-four hours later.

During the second session each participant arrived and completed the third and final IAT. After its completion, the experimenter fully debriefed the participant. The participant was given a $5 voucher and entered into the draw to win one of three $50 vouchers.

This project was reviewed and approved by the University of Canterbury Human Ethics Committee.

Results

Data Coding

The SST

Prior to analyses, SST data was split into thirds (80 trials in each third) for each participant. The data on the first third was used as the measure of initial SST performance, and data from the final third was used as the measure of final SST performance. These two thirds are referred to hereafter as the early block and late block respectively[1].

The distribution of the raw response time data on the SST was examined and found to be positively skewed, as is usual for response time data. To address this, median rather than mean response times were computed, and these were found to be normally distributed. Median responses times were computed for each participant for each combination of block and pair type (giving four scores, one each for early block stereotypic pairs, early block counter-stereotypic pairs, late block stereotypic pairs, and late block counter-stereotypic pairs). Trials on which the initial response of the participant was incorrect were excluded from the calculation of median response times. An error rate was also calculated for each participant for each combination of block and pair type (giving an error rate for early block stereotypic pairs, early block counter-stereotypic pairs, late block stereotypic pairs, and late block counter-stereotypic pairs).

The IAT

A macro was used to generate overall IAT bias scores for each participant. This macro followed the procedure recommended by Greenwald et al. (2003). The IAT bias score reflects the relative strength of association between the stereotypic and the counter-stereotypic pairings in the IAT task. That is, the greater the deviation from zero, the greater the difference in relative

[1] Analyses were also run using a median split on the SST trials, as used by Plant et al. (2005). This analysis revealed the same pattern of results as that reported here based on a thirds-split.

strength of association between stereotyped and counter-stereotyped concepts. To ensure that the IAT bias scores from each experimental condition could be interpreted in the same way, the IAT bias scores on the CSS version (counter-stereotypic block presented first) were reverse signed. This resulted in all positive IAT bias scores reflecting greater strength of association between stereotyped categories than between counter-stereotyped categories, and all negative IAT bias scores reflecting greater strength of association between counter-stereotyped categories than stereotyped categories. That is, in this research a positive IAT score indicates a stronger association between males and science than between females and science.

Following Greenwald et al. (2003), the details of the procedure used to calculate IAT scores were as follows. First, all trials with response times greater than 10 seconds were eliminated (23 trials, 0.03% of all trials). Greenwald et al. (2003) then recommend removal of any participants who had greater than 10% of responses faster than 300 ms, but there were no such participants in the present study. With the remaining data, the mean for each block of correct latencies was computed, as well as individual block standard deviations and pooled standard deviations for blocks 3 and 6 (practice blocks) and 4 and 7 (critical blocks). Response time for errors was replaced with the individual block mean plus two individual block standard deviations. There was an average of 11.45 errors made per IAT per person (4.8% error rate). Corrected means were calculated for each block including the replaced error latencies. Two differences were computed: block 6 – block 3, and block 7 – block 4. These differences were divided by the respective pooled block standard deviation. Finally, these two quotients were averaged to provide a final IAT score.

Analyses

The SST: Response Times.

The mean median response times (in seconds) on the SST are presented in Table 1 as a function of condition, block and pair type.

Table 1:
Mean median response times on the SST as a function of condition, block and pair type

		EF Condition	CS Condition	Both experimental conditions
Early Block	Stereotypic pairs	.8315	.8540	.8426
	Counter-stereotypic pairs	.8560	.8276	. 8420
	Overall	.8438	.8408	.8423
Late Block	Stereotypic pairs	.7256	.6758	.7011
	Counter-stereotypic pairs	.7292	.6710	.7006
	Overall	.7274	.6734	.7009

A 2 (block: early, late) x 2 (condition: EF, CS) x 2 (pair type: stereotypic, counter-stereotypic) x 2 (sex: male, female) x 2 (degree: science, non-science) mixed-model ANOVA with repeated measures on the first and third factors was computed on response times. The study focused on strength of association of sex with science, so non-science degrees were combined for the analyses to illustrate whether SST scores were related to doing a science degree. The analysis revealed a main effect of block, such that participants were faster on the late block than on the early block, $F_{(1,59)} = 54.62, p < .001$ (early block mean 0.84 seconds, late block mean 0.70 seconds).

The SST: Error Rates

The mean error rates on the SST are presented in Table 2 as a function of condition, block and pair type.

A 2 (block: early, late) x 2(condition: equal frequencies, counter-stereotypic) x 2 (pair type: stereotypic, counter-stereotypic) x 2 (sex: male, female) x 2 (degree: science, non-science) mixed-model ANOVA with repeated measures on the first and third factors was conducted on error rates. There was only a significant main effect of pair type, $F_{(1, 59)} = 5.59, p < 0.05$. Participants made more errors on counter-stereotypic pairs (mean 2.70 errors) than on stereotypic pairs (mean 2.09 errors).

Table 2:
Mean errors on the SST as a function of condition, block and pair type.

		EF Condition	CS Condition	Both experimental conditions
Stereotypic Pairs	Early Block	2.205	1.606	1.906
	Late Block	2.500	1.455	1.977
	Overall	2.353	1.530	1.942
Counter-stereotypic Pairs	Early Block	2.147	2.667	2.407
	Late Block	2.676	3.061	2.869
	Overall	2.412	2.864	2.638

The IAT

Preliminary analyses revealed no effects of sex or degree on any of the IAT scores and hence these factors were not considered further. Mean IAT bias scores as a function of condition and time can be seen below in Table 3. The IAT bias scores were compared to zero (no bias) separately for each condition using single sample t-tests ($p < .01$). All nine IAT scores were significantly different from zero, indicating significant levels of bias for all conditions on all three IATs. That is, participants at all times and in all conditions had a stronger association between males and science than between females and science.

Table 3:
Mean bias scores on the IAT as a function of condition and time

	IAT 1	IAT 2	IAT 3
Control	.254	.172	.248
Equal Frequencies	.237	.251	.261
Counter-stereotypic	.192	.280	.394

A 3 (condition: control, EF, CS) x 3 (time: 1, 2, 3) mixed model ANOVA with repeated measures on the second factor was computed on the IAT bias scores. There was a significant main effect of time, $F_{(2, 184)} = 3.70, p < .05$. There was also a significant interaction between time and condition, $F_{(4, 184)} = 2.39, p < 0.05$. The IAT scores as a function of time and condition can be seen in Figure 1 below. To capture the meaning of this interaction, post-hoc tests were conducted to check for significant differences within- or across-conditions. There were no significant differences between conditions on any of the IATs. The only significant within-condition difference that emerged was in the counter-stereotypic condition, between IAT 1 (mean score 0.17) and IAT 3 (mean score 0.39).

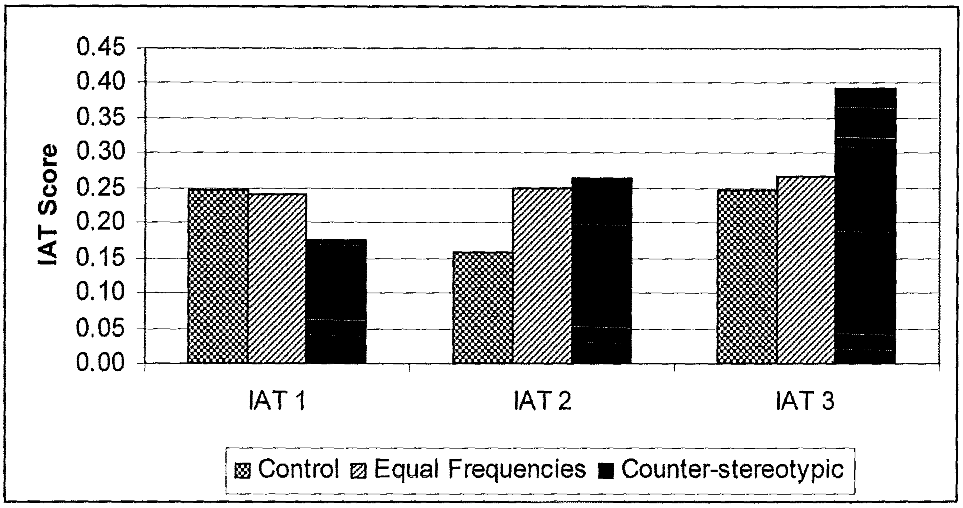


Figure 1: IAT bias scores as a function of condition and time

Relationship between Measures: SST and IAT

Change on the SST was captured by four scores. Counter-stereotypic change scores for response time (SST CS RT) and error rates (SST CS error) were calculated by subtracting the response time (or error rate) on the late block for counter-stereotypic pairs from the respective score on the early block. Stereotypic change scores were calculated for response times (SST S RT) and error rates (SST S error) by subtracting the response time score (or error rate) on the stereotypic pairs on the late block from the respective score on the early block. A positive score on each of the response time measures reflected a shorter mean response time on the late block than the early block. A positive score on the error rate change scores reflected less errors made on the late block than on the early block.

Change on the IAT was captured by two change scores. Change between IAT 1 and IAT 2 (IAT 1-2 Bias) was calculated by subtracting the IAT 2 score from the IAT 1 score. Change between IAT 2 and IAT 3 (IAT 2-3 Bias) was calculated by subtracting the IAT 3 score from the IAT 2 score. A positive IAT change score is indicative of less bias on the later IAT than the earlier IAT, that is, a positive IAT 2-3 bias score means that the participant had more bias at time 2 than they did at time 3.

Significant product-moment correlations between the change scores can be seen Table 4 below. Bold figures indicate correlations that are significant, $p < .05$. A positive correlation emerged between change on counter-stereotypic response times on the SST and stereotypic response times on the SST, $r_{(64)} = .755, p < .05$. This could indicate an individual difference variable related to general speed on the task, as an increase in speed on stereotypic trials was associated with an increase in speed on counter-stereotypic trials. Additionally, there was a significant negative correlation between change in SST stereotypic error and change in SST counter-stereotypic error, $r_{(64)} = -0.291, p < .05$. That is, an increase in error on stereotypic pairs was related to a decrease in error on counter-stereotypic pairs.

Table 4:
Correlation matrix of change scores.

	SST CS RT	SST S RT	IAT 1-2 Bias	IAT 2-3 Bias	SST S error	SST CS error
SST CS RT	1.000					
SST S RT	**0.755**	1.000				
IAT 1-2 Bias	0.134	0.127	1.000			
IAT 2-3 Bias	0.005	-0.120	-0.199	1.000		
SST S error	0.029	0.012	0.029	0.169	1.000	
SST CS error	0.115	0.078	-0.056	-0.029	**-0.291**	1.000

Discussion

Summary of Findings

The results of this experiment offered minimal support for the hypotheses. Partial support only was obtained for hypothesis 1, and the rest of the hypotheses were not supported. Each hypothesis and related results are discussed briefly, followed by a number of considerations which may have had an impact on the lack of support.

Hypothesis 1 was that participants would initially show significant levels of bias toward associating males with science on the IAT (i.e., a positive IAT score significantly different from zero) and on the SST (i.e., faster response times and less errors on stereotypic pairs than counter-stereotypic pairs). This was partially supported, as participants in all three conditions had a significantly stronger association between males and science than females and science on the initial IAT. This fits with previous research demonstrating the stereotype in multiple countries (Nosek et al 2007), including New Zealand (Malinen & Johnston, 2007). However, the prediction was also made that participants would show stereotypic responding on the first block of trials on the SST. This was not supported as there were no significant differences in either error rates or response times between the stereotypic and counter-stereotypic pairs on the early block of the SST.

The pattern of responding found on the SST indicates that participants' responses on the SST were not being influenced by their stereotype. This differs from previous research, for example Plant et al. (2005), as the participants who went through their manipulation (similar to the SST) initially made more errors on stereotypic than counter-stereotypic trials.

Hypothesis 2 predicted that the relative strength of participants' association between males and science would be reduced by both versions of the SST. Lower scores on the second IAT than the first IAT for both experimental conditions would have supported this hypothesis. There were no significant differences in IAT scores between IAT 1 and IAT 2, which gives no support to the hypothesis. With regards to the SST, there was no difference in error rates between the early and late blocks of the SST. Analyses revealed significantly faster reaction times on the late block than early block, but this decrease in reaction time was found for both stereotypic and counter-stereotypic pairs, and there was no difference between pair types on either the early or late block, or between conditions. The SST failed to have an effect on IAT scores for both experimental conditions.

The third hypothesis was that the counter-stereotypic condition would be a more effective method for changing stereotypes than the equal frequencies condition. Given the lack of impact of either condition on performance on the SST, however, such a comparison could not be tested.

Hypothesis 4 predicted that the effect of the experimental manipulations would remain stable over the twenty-four hour period such that bias would still be reduced on the third IAT. This would have been indicated by scores on the second and third IATs that were not significantly different from each other, for either or both of the experimental conditions. However, it would be illogical to say that effects persisted when no effects were found in the first instance. The current results prevent any conclusions being drawn about the comparative effectiveness of either condition at reducing the strength of implicit stereotypic associations.

Hypothesis 5 predicted a positive correlation between change on the SST and change on the IAT. The only significant correlations were a positive correlation between response time change on the stereotypic and the counter-stereotypic pairs of the SST, and a negative correlation between change in error rate on the stereotypic and counter-stereotypic pairs on the SST. This second correlation could indicate that some change in the pattern of responding on the SST was occurring.

One result emerged which was not predicted by any hypothesis. Post-hoc tests on the IAT scores showed that the counter-stereotypic condition had significantly higher bias at time 3 than at time 1. There was no significant change from IAT 1 to IAT 2, and no significant change from IAT 2 to IAT 3, but in both time periods participants in the counter-stereotypic condition increased their bias so that overall IAT 1 and IAT 3 were significantly different. This finding is difficult to interpret meaningfully in light of the limited support for the hypotheses. As such, further replication of this effect is necessary before interpretation is attempted.

Overall, results were not supportive of the hypotheses. There are a number of reasons why effects did not emerge in this study, and a selection of these are discussed in the following sections.

Image Placement

Plant et al. (2005) varied the placement of their facial image and the superimposed object so that the image appeared in one of nine locations onscreen, and the object appeared in one of nine locations on the face. This ensured that participants had to scan the entire screen and face to see the information they wanted, making it more likely that they processed characteristics of the target individual like race.

In contrast, the text label in the present study contained all the necessary information for completing the SST, and the location of the text label (on the screen or relative to the photo) did not change. Additionally, the instructions for the task made no reference to looking at the photos, simply instructed participants to respond as quickly and accurately as they could. Thus participants had no incentive to shift their gaze from the location of the text box. This is likely to have minimized the amount of unintentional facial processing they engaged in and therefore also minimized attention to the sex of the target. Macrae, Bodenhausen, Milne, Thorn and Castelli (1997) showed that unless participants were processing an image semantically, they did not show stereotype activation. It is unlikely that participants processed the photos enough to interfere with their performance, as there was no benefit to performance for looking at the image, no cues to suggest they ought to, and no mechanisms in place to ensure they would.

The Amount of Counter-stereotypic Information

All of the counter-stereotypic studies discussed in the introduction employed a manipulation that consisted entirely of counter-stereotypic information. In addition, Blair's (2002) review of stereotype malleability research showed all ten of the studies that attempted to create change via promoting counter-stereotypes employed 100% counter-stereotypic information. The 70/30 split was used here to create increased realism, but perhaps it simply rendered the manipulation too weak instead.

Dasgupta and Asgari (2004) examined the relationship between the number of female faculty students had over the course of a year, and their resulting scores on an IAT at the end of that year (on an IAT measuring associations between males/females and leadership/supportiveness), and found that more exposure to female faculty significantly predicted less stereotypic IAT scores at the end of the year. This suggests there may be a relationship between the proportion of counter-stereotypic exposure and change in strength of stereotypic associations. To this end, perhaps a

70/30 or 50/50 split in a manipulation that only lasted for five to ten minutes may not have provided enough counter-stereotypic information to successfully change participants' bias.

Limitations and Future Research

All research that deals with human participants runs the risk that features of the experiment will influence participants in ways that were not intended. The biggest limitation of the current research is thought to be the placement of the image and label within the SST. No change can be expected if participants did not process the sex of the faces during the task. Replicating this study with variation in the location of the image and label may produce better results.

In addition, a potential reason why no effects were found may have been that neither 70% nor 50% counter-stereotypic information was enough to cause changes in the strength of stereotypic implicit associations. Further research could investigate the amount of counter-stereotypic exposure required to reduce the strength of association between stereotypic categories relative to the strength of association between counter-stereotypic categories.

Measuring whether each of the experimental conditions in the present research inhibited the category of sex would also be useful for drawing conclusions about the effects of the SST on the IAT scores. That is, Plant et al. (2005) did not use a measure of implicit associations in their study. Their results indicated that performance on the SST improved, and they demonstrated that the category of race had been inhibited. While no significant reduction in stereotyped responding on the SST was found here, it could still be possible that participants had managed to refrain from activating the category of sex. Future research might also include a measure of category inhibition such as a word fragment completion task. Measurement of category activation could help elucidate processes underlying performance on the SST, for example, whether category activation occurs across a range of amounts of counter-stereotypic information, or whether it only occurs when there is no relationship between the target categories and the sex of the target individual.

There are a number of other areas of research tangential to the current study that could be pursued. For example, studying how long changes in implicit associations persist for could aid the development of practically useful manipulations. That is, a manipulation that changes an individual's stereotypic associations for twenty-four hours has reasonably limited practical use, but one that has a more lasting effect could be used in situations where temporary stereotype reduction is beneficial (e.g. situations where stereotype threat is a concern). Another area of potential study that has received very little attention to date is the evaluation of whether changes in implicit association strength produce corresponding changes in behaviour. This again is an area with great potential for developing practical applications to address situations affected by stereotypes.

The Wider Picture

This research was embedded in two major areas of research – implicit stereotyping and the effect of stereotypes on female participation in science and math based disciplines. Each of these areas, while already encompassing much promising research, still requires work to develop a clear understanding of the concepts involved.

There are some fairly unclear areas within the domain of implicit attitudes. There are a number of papers debating the theoretical basis and actual meaning of IAT scores (Greenwald, Nosek, Banaji & Klauer 2004, Nosek & Sriram, 2007, Blanton, Jaccard, Christie & Gonzales 2007). There is dissention regarding whether explicit and implicit measures of stereotypes are two different measures of the same stereotype, or whether they capture two distinct attitudes (Fazio & Olson, 2003; Nosek & Smyth, 2007). Further, the IAT, the most popular method of measuring implicit attitudes (Fazio & Olson, 2003), is not a measure of strength of association between males and science, but a relative strength of association between males and science (and females and arts) compared to females and science (and males and arts). The introduction of the Go-No Go Association Test (GNAT; Nosek & Banaji 2001) and similar tasks is beneficial for the field, as it allows exploration of the strength of association between, say, males and science relative to females and science, without the need for a contrasting category to science. However, given the proliferation of studies that use the IAT, exploration into the relationship between scores on an IAT and scores on a GNAT will be of great importance for future attempts to draw the implicit attitude literature together into a meaningful framework.

The second area of psychological research that this study links to is the study of the stereotype that males are better suited to science than females, and how the strength of this stereotype affects the career choices of women. Reducing the strength of this stereotype may make females more likely to choose science careers, if there is indeed a direct link between the prevalence of the stereotype and the career choices of females. This approach to the problem of differential participation by sex cannot be developed in isolation of other considerations. For example, making science more accessible to females may have the added effect of making it less desirable for males (Heilman, 1979). Changing the strength of the association between males and science may simply be one necessary facet of a multifaceted approach to encouraging individuals to choose occupations based on their interests and abilities, rather than their sex. The lack of stereotyped responding on the SST in the present research shows that individuals can ignore information that is irrelevant to performance (see also Bodenhausen & Macrae, 1998), which could be another important way to address stereotyping issues.

Conclusion

This study attempted to change the implicit strength of stereotypic associations between males and science relative to females and science. For a combination of reasons discussed above, the manipulation used failed to generate any change in implicit stereotypes as measured by an IAT. Significant levels of bias on the IAT indicate that the sample showed a stronger association between males and science than females and science, and it is likely that features of the SST were a main contribution to the lack of significant findings. A number of avenues for future research were proposed which will aid comprehension of the conditions under which stereotype change is possible.

References

Ashburn-Nardo, L., Knowles, M. L., & Monteith, M. J. (2003). Black Americans' implicit racial associations and their implications for intergroup judgment. *Social Cognition, 21 (1),* 61 – 87.

Baker, D. P. & Jones, D. P. (1993). Creating gender equality: Cross-national gender stratification and mathematical performance. *Sociology of Education, 66 (2*), 91 – 103.

Berdahl, J. L. (2007). The sexual harassment of uppity women. *Journal of Applied Psychology, 92 (2),* 425 – 437.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology* Review, 6 *(3),* 242 – 261.

Blanton, H., Jaccard, J., Christie, C., Gonzales, P. M. (2007). Plausible assumptions, questionable assumptions and post hoc rationalizations: Will the real IAT, please stand up? *Journal of Experimental Social Psychology, 43 (3)*, 399 – 409.

Bodenhausen, G. V. & Macrae, C. N. (1998). Stereotype activation and inhibition. In R. S. Wyer (Ed.), *Stereotype Activation and Inhibition* (pp 243 – 256). Mahwah, NJ: Erlbaum.

Brinsmead-Stockham, K., Johnston, L., Miles, L., & Macrae, C. N. (in press). Female sexual orientation and menstrual influences on person perception. *Journal of Experimental Social Psychology.*

Burr, V. (1998). *Gender and Social Psychology.* London: Routledge.

Chatman, J. A., Polzer, J. T., Barsade, S. G. & Neale, M. A. (1998). Being different yet feeling similar: The influence of demographic composition and organizational culture on work processes and outcomes. *Administrative Science Quarterly, 43 (4),* 749 – 780.

Correll, S. J. (2001). Gender and the career choice process: The role of biased self-assessments. *The American Journal of Sociology, 106 (6),* 1691 – 1730.

Crombie, G., Sinclair, N., Silverthorn, N., Byrne, B. M., DuBois, D. L., & Trineer, A. (2005). Predictors of young adolescents' math grades and course enrollment intentions: Gender similarities and differences. *Sex Roles, 52 (5/6),* 351 – 367.

Crowley, K., Callanan, M. A., Tenenbaum, H. R., & Allen, E. (2001). Parents explain more often to boys than to girls during shared scientific thinking. *Psychological Science, 12 (3),* 258 - 261.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 12 (2),* 163 – 170.

Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology, 40,* 642 – 658.

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

Davies, P. G., & Spencer, S. J. (2005). The gender-gap artefact: Women's underperformance in quantitative domains through the lens of stereotype threat. (Pp. 172 – 188.). In A. M. Gallagher & J. C. Kaufman (Eds.). *Gender differences in mathematics: An integrative psychological approach.* New York: Cambridge University Press.

Davies, P. G., Spencer, S. J., Quinn, D. M. & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin, 28 (12),* 1615 – 1628.

Education Counts – Ministry of Education (2007). School Roll Summary Report: July 2007: Subject Enrolments as at 1 July 2007. Retrieved 19 January 2008 from http://www.educationcounts.govt.nz/statistics/schooling/july_school_roll_returns/subject_enrolment/subject_enrolments_as_at_1_july_2007

Education Counts – Ministry of Education (2006). Provider-based enrolments. Retrieved 1 December 2007 from http://www.educationcounts.govt.nz/statistics/tertiary_education/participation

Fazio, R. H. & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297 – 327.

Fitzgerald, L. F., Drasgow, F., Hulin, C. L., Gelfand, M. J., Magley, V. J. (1997). Antecedents and consequences of sexual harassment in organizations: A test of an integrated model. *Journal of Applied Psychology, 82 (4),* 578 – 589.

Glomb, T. M., Munson, L. J., Hulin, C. L., Bergman, M. E. & Drasgow, F. (1999). Structural equation models of sexual harassment: Longitudinal explorations and cross-sectional generalizations. *Journal of Applied Psychology, 84 (1),* 14 – 28.

Greene, A. L., Sullivan, H. J., & Beyard-Tyler, K. (1982). Attitudinal effects of the use of role models in information about sex-typed careers. *Journal of Educational Psychology, 74 (3),* 393 – 398.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology,* 74 (6), 1464 – 1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85 (2),* 197 – 216.

Greenwald, A. G., Nosek, B. A., Banaji, M. R., Klauer, K. C. (2004). Validity of the salience asymmetry interpretation of the Implicit Association Test: Comment on Rothermund and Wentura (2004). *Journal of Experimental Psychology, 134 (3),* 420 – 425.

Gregg, A. P., Banaji, M. R. & Seibt, B. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology, 90 (1),* 1 – 20.

Heilman, M. E. (1979). High school students' occupational interest as a function of projected sex ratios in male-dominated occupations. *Journal of Applied Psychology, 64 (3),* 275 – 279.

Jensen, I. W. & Gutek, B. A. (1982). Attributions and assignment of responsibility in sexual harassment. *Journal of Social Issues, 38 (4),* 121 – 136.

Johnston, L., Arden, K., Macrae, C. N., & Grace, R. C. (2003). The need for speed: The menstrual cycle and person construal. *Social Cognition, 21 (2),* 89 – 100.

Karpinski, A. & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81 (5)*, 774 – 788.

Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder: Improving implicit racial attitudes and interracial interactions through approach behaviours. *Journal of Personality and Social Psychology, 92 (6),* 957 – 971.

Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes, gender identification, and math-related outcomes: A prospective study of female college students. *Psychological Science, 18 (1)*, 13 – 18.

Kunda, Z. & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin, 129 (4)*, 522 – 544.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., Thorn, T. M. J., & Castelli, L. (1997). On the activation of social stereotypes: The moderating role of processing objectives. *Journal of Experimental Social Psychology, 33,* 471 – 489.

Malinen, S. & Johnston, L. (2007). The influence of an equity statement on perceivers' implicit and explicit associations between males and science. *New Zealand Journal of Psychology, 36 (1), 18 – 24.*

Mansfield, P. K., Koch, P. B., Henderson, J., Vicary, J. R., Cohn, M., & Young, E. W. (1991). The job climate for women in traditionally male blue-collar occupations. *Sex Roles, 25,* 63 – 79.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behaviour, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37(5),* 435 – 442.

Miller, L. & Budd, J. (1999). The development of occupational sex-role stereotypes, occupational preferences and academic subject preferences in children at ages 8, 12, and 16. *Educational Psychology, 19 (1).*

Ministry of Research Science and Technology (1998). Human resources in science and technology in New Zealand. Retrieved 19 January 2008, from http://www.morst.govt.nz/Documents/publications/statistics/HRST-report.pdf

Ministry of Research Science and Technology of New Zealand (2003). Women and Scientific Careers Report. Retrieved 19 January 2008, from http://morst.govt.nz/Documents/publications/researchreports/Women-and-Scientific-Careers.pdf

Nosek, B. A. (1999). Math is hard! Gender, mathematics, and implicit social cognition. Paper presented at the 1999 Graduate Student Conference, New York University, November 19, 1999.

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology, 135 (4),* 565 – 584.

Nosek, B. A. & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition, 19 (6*), 625 – 666.

Nosek, B. A., Banaji, M. R. & Greenwald, A. G. (2002). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology, 83 (1),* 44 – 59.

Nosek, B. A. & Smyth, F. L. (2007). A multitrait-multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs. *Experimental Psychology, 54(1),* 14 – 29.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18,* 36 – 88.

Nosek, B. A., & Sriram, N. (2007). Faulty assumptions: A comment on Blanton, Jaccard, Gonzales and Christie (2006*). Journal of Experimental Social Psychology, 43*, 393 – 398.

Oppliger, P. (2007). Effects of gender stereotyping on socialization. In R. W. Preiss, B. M. Gayle, N. Burrell, M. Allen, & J. Bryant (Eds.). *Mass Media Effects Research: Advances through Meta-analysis.* Routledge.

Organization for Economic Co-operation and Development (2006). Report on the Evolution of Student Interest in Science and Technology Studies. Retrieved 1 November 2007 from http://www.oecd.org/dataoecd/16/30/36645825.pdf

Patterson, M. M., & Bigler, R. S. (2006). Preschool children's attention to environmental messages about groups: Social categorization and the origins of intergroup bias. *Child Development, 77 (4),* 847 – 860.

Plant, E. A., & Peruche, B. M. (2005). The consequences of race for police officers' responses to criminal suspects. *Psychological Science, 16 (3),* 180 – 183.

Plant, E. A., Peruche, B. M. & Butz, D. A. (2005). Eliminating implicit racial bias: Making race nondiagnostic. *Journal of Experimental Social Psychology, 41 (2),* 141 – 156.

Peruche, B. M. & Plant, E. A. (2006). Racial bias in perceptions of athleticism: The role of motivation in the elimination of bias. *Social Cognition, 24 (4),* 438.

Poehlman, T. A., Uhlmann, E. L., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. Unpublished manuscript. Retrieved 15 May 2007 from http://faculty.washington.edu/agg/pdf/IAT.Meta-analysis.16Sep05.pdf.

Prentice, D. A., & Carranza, E. (2002). What women and men should be, shouldn't be, are allowed to be, and don't have to be: The contents of prescriptive gender stereotypes. *Psychology of Women Quarterly, 26,* 269 - 281.

Rydell, R. J., McConnell, A. R., Mackie, D. M., & Strain, L. M. (2006). Of two minds: Forming and changing valence-inconsistent implicit and explicit attitudes. *Psychological Science, 17 (11),* 954 – 958.

Schmader, T., Johns, M., & Barquissau, M. (2004). The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. *Sex Roles, 50 (11- 12)*, 835 – 850.

Society for Human Resource Management (SHRM) (2005). 2005 Workplace diversity practices survey report. Available from http://www.shrm.org/.

Spence, J. T., & Buckner, C. E. (2000). Instrumental and expressive traits, trait stereotypes, and sexist attitudes. *Psychology of Women Quarterly, 24 (1),* 44 – 62.

Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35 (1),* 4 – 28.

Statistics New Zealand (2003). Work. Retrieved 19 January 2008 from http://www.stats.govt.nz/NR/rdonlyres/D8E93A92-AB8B-4E49-9DD7-E35F74589485/0/Work.pdf

Statistics New Zealand (2005). Focusing on Women 2005. Retrieved 19 January 2008 from http://www.stats.govt.nz/NR/rdonlyres/FC1BDDF4-F40F-4B05-B00C-104AC9D38D78/0/FocusingOnWomen2005.pdf

Stroeher, S. K. (1994). Sixteen kindergartners' gender-related views of careers. *The Elementary School Journal, 95 (1),* 95 – 103.

Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational Psychology, 92 (1),* 144 – 151.

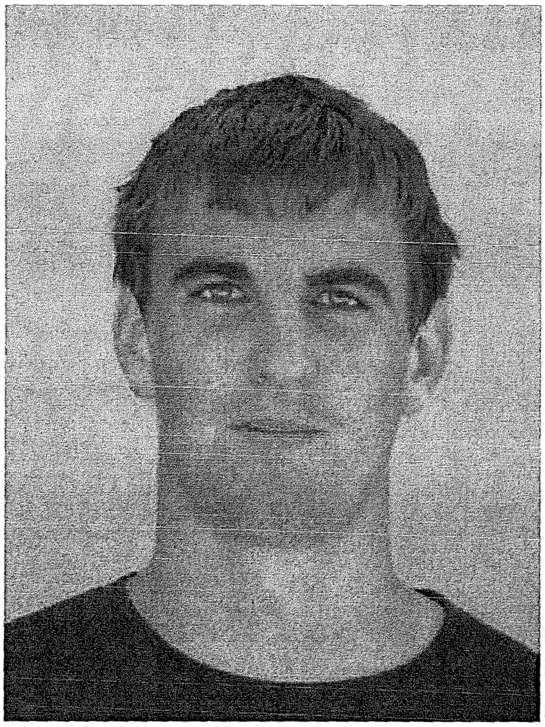
Walton, P. R. (2003). An Implicit Associations Test. [Computer software]. Christchurch, New Zealand: Dexterware.

Walton, P. R. (2003). A stimulus sorting task. [Computer software]. Christchurch, New Zealand: Dexterware.

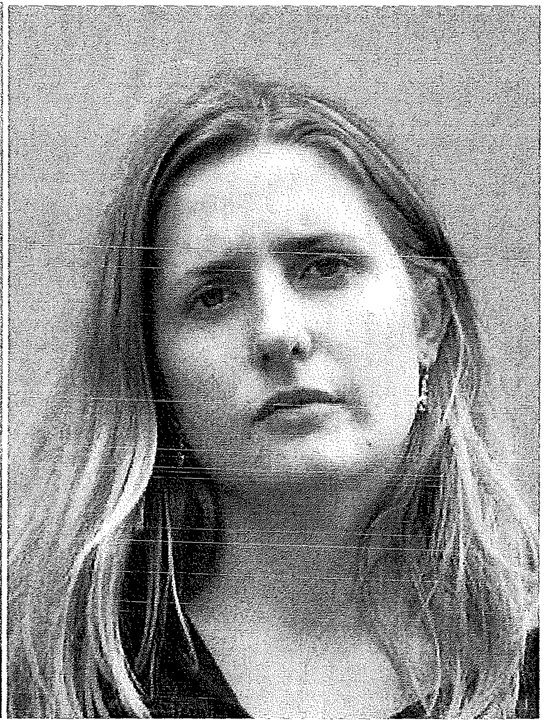
Appendix A: Words Used in the IAT and SST

Male	Female	Science	Arts
Boy	Girl	Astronomy	American Studies
Father	Mother	Biochemistry	Art History
Uncle	Aunt	Biology	Classics
Him	Her	Chemistry	Cultural Studies
Man	Woman	Computer Science	English
Male	Female	Geology	Gender Studies
Son	Daughter	Mathematics	History
Grandfather	Grandmother	Microbiology	Music
His	Hers	Molecular Genetics	Religious Studies
He	She	Physics	Sociology

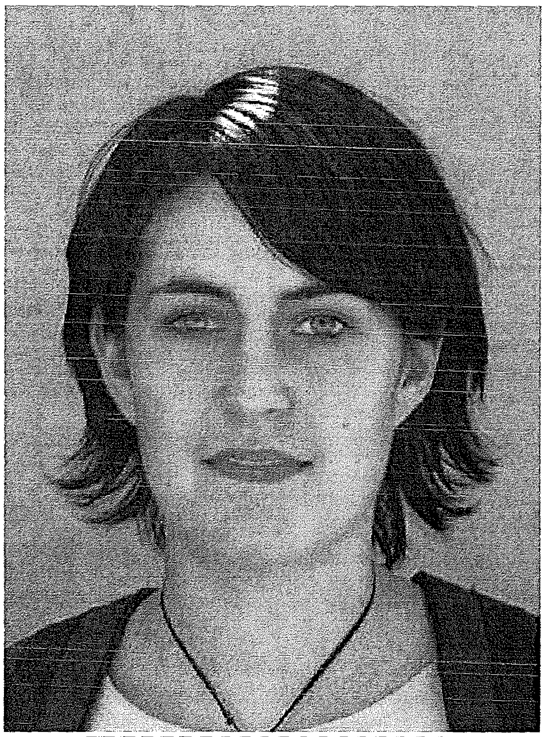
Appendix B: Sample pictures used in the SST



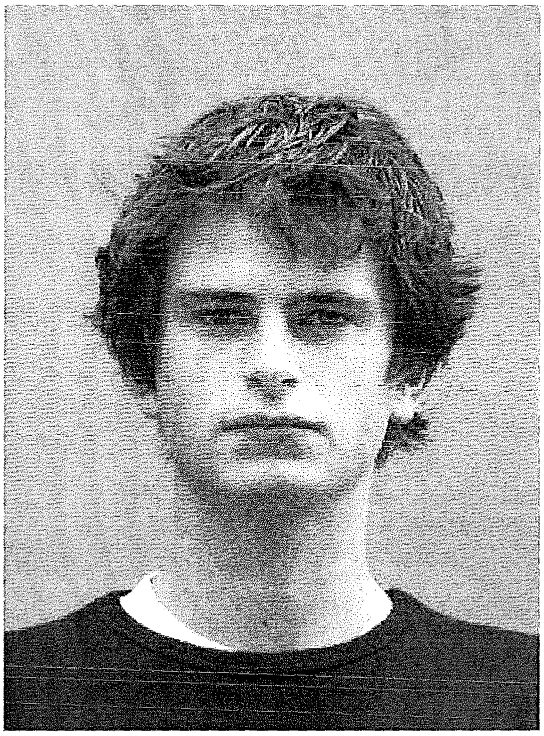
What is your major?
Biochemistry



What is your major?
Cultural Studies



What is your major?
Biology



What is your major?
Art History

Appendix C: Instructions for the SST and IAT

Instructions for the CSS IAT were as follows. There were minor differences between the versions of the IAT. Where the two differ, the differences are indicated by underlined text and the wording used in the SCS version is in parentheses.

Instructions for Blocks 3 & 4 were identical, as were the instructions for Blocks 6 & 7. The only difference was that Blocks 3 and 6 included the line, "This is the first of two identical blocks" at the end of the instructions.

	There are seven blocks in this section. In each you will see a series of words in the centre of the screen. You will need to sort these words into various categories which will appear on the left and right hand sides of the screen. The categories may change between blocks, so please read each new instruction screen carefully.
Block 1	For this block you are to classify the words presented into two categories – male and female. The <u>male</u> (female) category label will appear on the left hand side of the screen. To sort words into this category press 'A'. The <u>female</u> (male) category label will appear on the right hand side of the screen. To sort words into this category press 'L'. The category labels will appear onscreen during the block.
Block 2	For this block you are to classify the words presented into two categories – arts and science. The arts category label will appear on the left hand side of the screen. To sort words into this category press 'A'. The science category label will appear on the right hand side of the screen. To sort words into this category press 'L'. The category labels will appear onscreen during the block.
Block 3	In this block two sets of category labels will appear. The word presented will fall into only one of these categories e.g. it will be science OR arts OR male OR female. The category labels of <u>male</u> (female) and arts will appear on the left hand side of the screen. To sort words into either of these categories press 'A'. The category labels of <u>female</u> (male) and science will appear on the right hand side of the screen. To sort words into either of these categories press 'L'. The category labels will appear onscreen during the block. This is the first of two identical blocks.
Block 5	For this block you are to classify the words presented into two categories – female and male. The <u>female</u> (male) category label will appear on the left hand side of the screen. To sort words into this category press 'A'. The <u>male</u> (female) category label will appear on the right hand side of the screen. To sort words into this category press 'L'. The category labels will appear onscreen during the block.
Block 6	In this block two sets of category labels will appear. The word presented will fall into only one of these categories e.g. it will be science OR arts OR male OR female. The category labels of <u>female</u> (male) and arts will appear on the left hand side of the screen. To sort words into either of these categories press 'A'. The category labels of <u>male</u> (female) and science will appear on the right hand side of the screen. To sort words into either of these categories press 'L'. The category labels will appear onscreen during the block.

These instructions were for the SST, and were identical over all four versions.

General	In this task you will be presented with a series of pictures of Canterbury University students, labelled with their subject major. You will need to sort these students into either arts students or science students, depending on their major. In all blocks, if you make a mistake, a cross will appear. You must then choose the correct answer. If you are correct you will see a tick.
Block 1 (practice)	This is a practice block to familiarise you with sorting pictures instead of words. A number of pictures will appear on the screen. If the photo is of a male, press 'A'. If the photo is of a female, press 'L'. The category labels will appear onscreen during the block. Please respond as quickly but accurately as possible.
Block 2 (manipulation)	From now on, you will be sorting the students into categories based on their major – either an arts subject of a science subject. The arts category label will appear on the left hand side of the screen. To sort students into this category press 'A'. The science category label will appear on the right hand side of the screen. To sort subjects into this category press 'L'. The category labels will appear onscreen during the block. Please respond as quickly but accurately as possible.