Developing an Implicit Association Test to Explore Implicit and Explicit Stereotypes of Empathy in Scientists among University Students in England

This dissertation is submitted for the degree of Doctor of Philosophy

By Yishu Qin

Supervisor: Dr Sara Baker

Faculty of Education

Darwin College

University of Cambridge

September, 2018

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as specified in the text. This dissertation does not exceed the prescribed word limit of the Degree Committee of the Faculty of Education.

Yishu Qin

September, 2018

For my parents, Jufen Chen and Guiyue Qin

Acknowledgements

I would like to first express my gratitude to the University of Cambridge, to the Faculty of Education, to Darwin College for providing resources and opportunities to sustain progress on this thesis. Thank you to my PhD examiners Dr Liz Winter and Dr Michael Reiss for reading and reviewing this thesis.

However, the greatest thanks at the University of Cambridge are for the incredible patience, guidance and unwavering support from my supervisor, Dr Sara Baker. I am extremely grateful to Sara as I could not have received any better advice, ideas, encouragement, commitment and homemade desserts to keep on track.

My grateful thanks go to all my research participants and the university staff who kindly take the time and trouble to advertise and complete my tests and questionnaires. Without over a thousand people contributing to my data collection, there would have been no thesis.

Special thanks to Nicole Lofaro for her help with Project Implicit, to Dr Jie Gao for her help with statistics, to Junlin Yu for the inspiring discussion on gender stereotyping, to Esinam Avornyo for her formatting tips, and to all the other fellow PhD students for sharing their ideas and feedback with me. My gratitude extends to Emma Rixon, whose coordinating work has made my PhD life much easier.

I would also like to thank my friends for always cheering me up and listening to me. Your friendship is treasured – Danni, Mandy, Lei, Xiao, Yuzhi, He Heng, Xia Wu, Yi Zheng, and Zhiyong. Special thanks to Uncle Ji for his precious life advice of wisdom.

Lastly, my deepest gratitude goes to my big family, especially to my dearest mom and dad – Jufen Chen and Guiyue Qin. Words are not enough to thank you for your unconditional love which has made me who I am.

Abstract

Over the past 50 years, strong social norms have developed in Western societies that encourage people to curb the overt expression of prejudice and stereotypes based on social categories such as gender and race. Yet, approaches applying new measurement of stereotypes, focusing on implicit evaluations and beliefs that are not under conscious control, suggest that such suppressed thoughts are not simply eliminated but may leak out in more subtle ways. Recent work has demonstrated that one of the subtle ways is through feature-based stereotyping. Regarding the issue of women's underrepresentation in STEM fields, we argue that instead of referring to gender-based stereotypes, students may look to other characteristics (e.g., empathy) as a feature-based cue to determine whether they fit the image of particular careers.

The main purpose of the PhD project was to develop a new Implicit Association Test (IAT) to explore people's unconscious feature-based stereotyping of empathy in scientists. It was comprised of three stages. In the first stage, the procedure of how to systematically design and computerize the online Science-Empathy IAT (SE-IAT) was presented. In the second stage, psychometric properties of the newly developed SE-IAT were examined and issues with the original SE-IAT were addressed by modifying the test to the Single-Category SE-IAT (SSE-IAT). In the third stage, comparisons between implicit and explicit attitudes were made using the modified SSE-IAT and self-report questionnaires.

Results showed that implicit and explicit stereotypes were distinct constructs. Despite the self-report positive attitudes, participants demonstrated implicit bias against empathy in scientists. Moreover, different gendered patterns, as well as variations by major subject, were also observed for implicit and explicit stereotypes. Finally, explicit stereotype, but not implicit stereotype, was found to relate to scientific career aspirations. These findings indicate that the 'socially clumsy' image of scientists may still be a deeply embedded stereotype amongst University students in England. Interventions such as counter-stereotyping role models are in demand to attract more people to pursue science.

Table of contents

List of Tables

List of Figures

List of Abbreviations

BIT	Balanced Identity Theory
CAS	Career Aspirations in Science
EAST	Extrinsic Affective Simon Task
EQ	Empathising Quotient
GNAT	Go/No-go Association Task
GS-IAT	Gender-Science Implicit Association Test
HESA	Higher Education Statistics Agency
IAT	Implicit Association Test
IPT	Implicit Personality Theory
ISSOS	Interpersonal Subscale of the Stereotypes of Scientists
MODE	Motivation and Opportunity as Determinants model
PAS	Public Attitudes to Science
SCM	Stereotype Content Model
SE-Explicit	Science-Empathy Explicit scale
SE-IAT	Science-Empathy Implicit Association Test
SSE-IAT	Single-category Science-Empathy IAT
SIT	Social Identity Theory
STEM	Science, Technology, Engineering, Mathematics
SQ	Systemising Quotient
UCAS	Universities and Colleges Admissions Service
WISE	Women in Science and Engineering campaign

Chapter 1 Introduction

With the rapid development of science and technology, every country in the world needs more skilled scientists and engineers to keep pace with the times. This demand will not be met unless great efforts are made to recruit and retain women in science, technology, engineering and mathematics (STEM) careers. Although the number of women obtaining undergraduate degrees have increased drastically in “soft” scientific fields such as in Life Science and Medicine, the gender gap remains wide in Physical Science, Mathematics, Computer Science as well as Engineering which have the reputation of requiring a high level of mathematical skill and of being “hard” sciences (UCAS, 2017).

Apparently, there is no single explanation for the lack of gender diversity in math-intensive STEM fields. Early studies often placed emphasis on gender difference in mathematical ability and academic achievement in STEM subjects. However, evidence has shown that men and women have equal intrinsic aptitude for mathematics and sciences (Spelke, 2005) and girls even outperform boys at GCSE, A-level and degree standards in most STEM subjects (UCAS, 2017). Gender differences in subject interests have been shown to be socially constructed, not biologically based (Archer et al., 2012; Leaper, Anderson, & Sanders, 1998; Rudman & Phelan, 2010). Under such circumstances, recent work has switched its attention to social and psychological factors that may constrain women's preference and choices rather than biological sex differences in scientific competence.

Existing evidence has shown that there is a mismatch between women's gender role and the image of scientists (Cheryan, Siy, Vichayapai, Drury, & Kim, 2011). A robust finding is that people tend to depict scientists as “a middle-aged man wearing a white coat and glasses who conducted dangerous experiments alone in a laboratory” (Finson, 2002; Steinke et al., 2007). In contrast, a feminine gender role usually includes traits of being warm, gentle, empathetic, sensitive, soft-spoken, as well as concerned about hair and makeup (Williams & Best, 1990; Worell, 2001). Though we can clearly see a gap between female gender role and the image of a scientist, it remains unknown which specific characteristic of the stereotypical image of scientist steers women away from STEM fields.

One phenomenon that has attracted my attention is that women tend to rate themselves higher in self-report empathy than men (Eisenberg & Lennon, 1983), whereas physical science, computer science as well as maths majors usually rate themselves lower in self-report empathy than those majoring in other subjects (Baron-Cohen, 2002; Manson & Winterbottom, 2012). Empathy, as a key social skill, is regarded associated with arts and humanities careers. It is possible that women do not prefer certain science careers not because they are incapable of the careers, but because they have better perceived empathy than men thus are more reluctant to be seen as "solitary geniuses" who are lacking in empathetic skills.

In this case, it is considered intriguing to investigate how people view empathy in scientists as well as whether stereotypes about empathy in scientists are related to science career aspirations. Though existing research has documented a pervasive stereotypical quirky image of scientists, very few empirical studies have been conducted to investigate contemporary stereotypes about empathy in scientists in particular. There is furthermore a shortage of reliable instruments to capture stereotypes of empathy in scientists.

In order to bridge this research gap, the present PhD project aims to develop a new measurement to assess stereotypes of empathy in scientists. It is important to differentiate between implicit and explicit stereotypes: explicit stereotypes are beliefs and attitudes people know they hold, subject to deliberate (often strategic) control in their expression (Fazio, Jackson, Dunton, & Williams, 1995). In contrast, implicit stereotypes are unconscious beliefs and attitudes that are inaccessible to introspection

(Greenwald & Banaji, 1995). Explicit stereotypes are often measured by self-report questionnaires and participants can alter their answers. Implicit stereotypes are assessed via different reaction time to stimuli that are either paired in a stereotypical way (e,g., empathy and humanities) or not (e.g., empathy and science) and participants have difficulty altering their immediate spontaneous reactions. The present study examines both implicit and explicit stereotypes of empathy in scientists.

Before providing an overview of the thesis structure, it is important to explain why I was interested in conducting such research. Being born in mainland China in 1991, I am the so-called one-child generation when China had a compulsory birth control policy according to which married couples were allowed to have only one child. Despite the proclamation made by Mao Zedong, the founding father of the People's Republic of China, that "women hold up half the sky", Chinese families preferred having sons instead of daughters (Gu & Li, 1995). The belief that girls are inadequate and of lesser value than boys is deeply rooted in our culture and influences people's attitudes and behaviours in various ways. Taking myself as an example, my grandfather was disappointed on the mere fact that I was a girl when I was born. During my middle school years, it was almost a consensus by most teachers that though boys obtained lower grades than girls, they were actually smarter and would surpass girls, especially at STEM subjects, as long as they started to devote more time on study when they entered high school. As a top student and a girl at the same time, I felt threatened by such stereotypes and struggled with the concern that I could have fallen behind boys at later stages. As a result, I gave up Maths and Physics soon after I entered high school.

As China became more and more open to the world, I dreamt to study in the UK which I supposed to be a place where gender equity had been achieved. With such belief in my heart, I started my postgraduate study in Cambridge and became aware that, in spite of the drastic decrease of overt expressions of stereotyping and discrimination, gender equity is still an issue in Western societies and unconscious biases about women as well as many other social groups are still widespread. It seems imperative to me that more empirical research should be conducted to uncover people's underlying biases toward different social groups in order to create awareness for persistent implicit stereotypes. While there are a variety of topics involved in unconscious stereotyping, my personal experience led me to focus on gender stereotypes and stereotypes of scientists. As few measures were developed to capture implicit stereotypes of empathy in scientists to date, I felt it would be meaningful to develop a new implicit instrument to capture unconscious biases about empathy in scientists to lay a solid foundation for further research looking at the potential influence of such stereotypes on career choices. The following paragraph outlines the structure of this thesis as well as the three stages of the present study.

The following Chapter 2 demonstrates the context of the present study and Chapter 3 contains critical reviews of relevant literature. This main research consists of three stages. In Stage one (Chapter 4), a new instrument applying the Implicit Association Test (IAT) paradigm to assess stereotypes of empathy in scientists is designed and developed. In Stage 2 (Chapter 5), the newly-developed IAT is examined and modified to address the discovered psychometric issues. In Stage 3 (Chapter 6), the modified IAT and self-report questionnaires are applied to investigate individual differences in the implicit and explicit stereotypes of empathy in scientists as well as the interplay of these stereotypes in science career aspirations. Results of Stage 2 study are discussed against psychometric criteria for implicit measures and results of Stage 3 study are discussed with reference to existing findings and theories about implicit and explicit stereotyping. Finally, a general discussion of the summary of findings, the contributions and limitations of the present PhD project are presented in Chapter 7.

Chapter 2 Context of the study

The main purpose of this chapter is to delineate the context of the present study and justify why it is important to study stereotypes of scientists. This chapter starts with a brief overview about the remaining gender gap in "hard" science fields (e.g., physics and computer science) in the UK, followed by a review of the potential explanations for women's underrepresentation in these fields. Based on the existing empirical evidence, some of the explanations are considered no longer valid and the potential influence of stereotypes on women's interests in science has received much recent attention. On that basis, justifications for the topic of the present study are demonstrated to provide the foundation for further literature review on related studies.

2.1. The current gender gap in certain STEM fields in the UK

Over the past 50 years, the number of women obtaining undergraduate degrees has significantly increased across a wide range of scientific disciplines in the UK. In the 1960s, all scientific disciplines were male dominated (Ceci & Williams, 2010). Today, more than half of all undergraduate degrees in life science and medicine are awarded to women. As shown in Figure 2.1 below, according to the report from the Higher Education Statistics Agency (HESA, 2018), women occupied 58% in Medicine & Dentistry, 77% in Veterinary Science, 63% in Agriculture, and 63% in Biological sciences among UK undergraduate degree holders in 2016/17. However, despite an upward trend in growth of female undergraduates in the STEM fields, women still remain the minority in other math-intensive subjects, including Physical Science (41%), Mathematical Science (37%), Computer Science (17%), and Engineering & Technology (18%). The recently released UCAS acceptance data also showed that women represent only 25% of the acceptances in Physics, Computer Science and Engineering in 2016 – a figure that has remained unchanged since 2007 (UCAS, 2017).

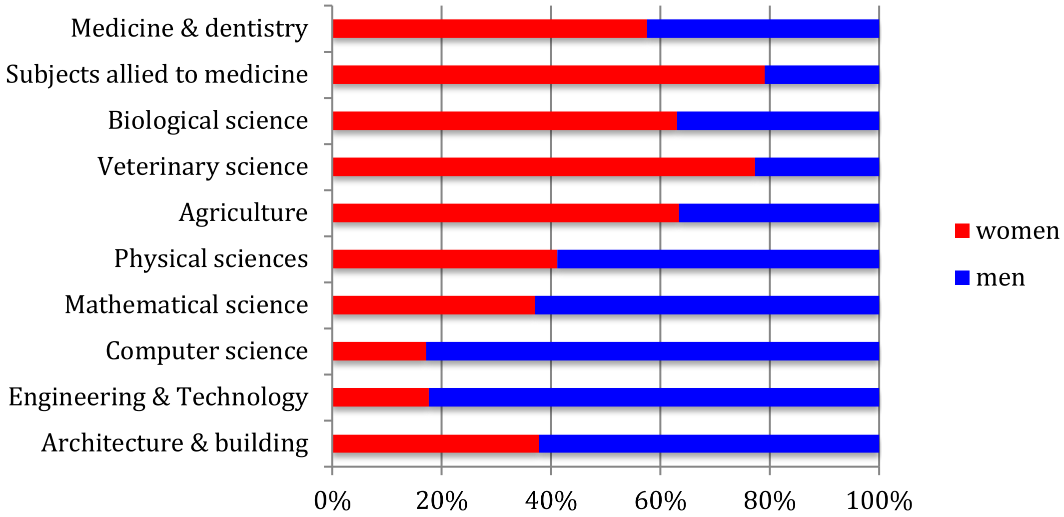


Figure 2.1 Undergraduate science qualifications obtained by students in the UK by gender, 2016/17. Data retrieved from HESA, 2018.

Statistics published by Eurostat in June 2017 show similar trends in subjects studied by women and men in higher education across the European Union countries. In 2015, women occupied only 26% of students studying engineering, manufacturing and construction related subjects and 38% in physical sciences, mathematics and computer science. In the same year, 78% of students studying education and 72% of students studying health related subjects were women.

In addition, there is another issue that has been referred to as the "leaky pipeline" whereby the number of women continues decreasing at the later stage of the pathway in STEM academia and also across STEM industry (Alper, 1993). The wider trend is that some female STEM participation decreases as the seniority of positions increases (Resmini, 2016). For example, according to the annual report provided by the Women in Science and Engineering (WISE) campaign, in 2014, 71% of male engineering graduates who were in employment went into engineering occupation, compared with 58.7% of female graduates in the UK (WISE, 2015). More recent data in 2016 also showed that only 8% of the UK's engineering workforce are women – the lowest number across all the EU countries (Morgan, Kirby, & Stamenkovic, 2016). Women also represent only 17% in IT professionals and 19% in IT Technicians (WISE, 2017).

It is therefore clear that women are still underrepresented in the most mathematically intensive fields – engineering, physics, mathematics, and computer science in higher education and workforce, despite the fact that women have actually outnumbered men in university with 57% of university students being women in the UK nowadays (HESA, 2018). The robust gender difference in some math-intensive STEM fields raises the important question of what its cause is.

2.2. Potential explanations for the gender gap in science

2.2.1. Are women "not cut out for science"?

Why women are underrepresented in certain STEM fields has been under debate for some time. Are there social and psychological factors, or do evolutionary and biological reasons lie behind this underrepresentation? Numerous researchers have proposed various factors to explain this gender gap, such as sex-linked hormones (e.g., Valla & Ceci, 2011), cognitive ability (e.g., Kaiser et al., 2008), stereotype threat (e.g., Good, Woodzicka, & Wingfield, 2010) and social forces (such as the effect of child rearing on careers, Ceci & Williams, 2011), over the past five decades.

Early studies focused mainly on the role of mathematical ability and achievement in STEM fields (e.g., Astin, 1975; Baenninger & Newcombe, 1995; Benbow, Lubinski, Shea, & Eftekhari-Sanjani, 2000; Fischbein, 1990; Lubinski & Benbow, 2006; Maccoby & Jacklin, 1974). This line of studies attempts to examine a particularly contentious hypothesis that gender differences in mathematical ability and achievement account for the lack of women in some STEM fields. However, regarding the intrinsic aptitude for mathematics, Spelke's (2005) meta-analysis of studies on cognitive sex differences revealed no consistent evidence for a male advantage, from birth to maturity, in the cognitive ability for mathematical and scientific thinking. Moreover, currently in the UK, 42% of undergraduate mathematics majors are women, as are 38% of postgraduate mathematics majors, suggesting that whatever these women's mathematical or spatial aptitude, they are compatible with high levels of achievement (WISE, 2015). As such, the claim that women are "not cut out for science" is not valid and should be discarded.

More recent studies suggest that women's underrepresentation in STEM fields has little to do with ability or achievement, but is instead more likely a result from their preferences and choices. It has been reported that women choose at a young age not to pursue careers in certain STEM fields, with few adolescent girls expressing desires to be engineers or physicists, preferring instead to be medical doctors, lawyers, psychologists and artists, which are concerned with "helping people" and "more girly" (Archer & McDonald, 1991; Archer, Osborne, DeWitt, Wong, & Willis, 2013; Rommes, Overbeek, Scholte, Engels, & Kemp, 2007). A survey among 9000 Year 8 (age 13 – 14) students in the UK uncovered that though a higher percentage of girls rated science as their favourite subject, yet girls were still less likely than boys to aspire to science careers (Archer, Osborne, et al., 2013). Similarly, The UCAS data also showed that, though girls outperform boys at GCSE, A-level and degree standards in scientific subjects, they are disproportionately less likely than males to choose certain STEM majors in higher education (UCAS, 2015). Based on these findings, women's underrepresentation in STEM should be attributed more to social and psychological factors that may constrain women's preference and choices rather than biological sex differences in scientific competence.

2.2.2. Potential factors constraining women's aspirations in science

Though women prefer not to enter certain scientific fields, presumably because they don't want to or they don't like these fields, their choices are not actually "free" or not constrained (Cheryan, Master, & Meltzoff, 2015). If women are freely choosing not to pursue sciences, perhaps nothing can or should be done about it – after all, it is their free will. However, a large body of research has demonstrated a variety of social and psychological barriers to women's entry into science and engineering that preclude women from being able to make a truly "free" choice.

Based on the existing literature and empirical evidence, we can distinguish three main factors that may limit women's science subject choices and career aspirations: 1) work-life balance issues; 2) gender roles and values; and 3) stereotypes of scientists.

2.2.2.1. Work-life balance issues

The work-life balance issues are due primarily to factors surrounding family formation and childrearing and have been identified as a critical contributor that leads to the retention difficulty with women working in STEM fields (Glass, Sassler, Levitte, & Michelmore, 2013; Hunt, 2010; Williams & Ceci, 2012). As pointed out by Williams and Ceci (2012), having children in one's career exerts more downward pressure on women than men in all fields, but this challenge is exacerbated in STEM fields due to the fact that the number of women is smaller to begin with. For example, the British case study of female employees in the Cambridge high-tech sector finds that women with children are effectively excluded from informal social networks after hours, limiting their job promotion opportunities (Gray & James, 2007).

In addition, the work-life balance issues not only contribute to attrition from STEM for those who have already become mothers but also to a lack of attraction to STEM for younger women who want to become mothers. A longitudinal study following 12th-grade girls who have an interest in male-dominant STEM careers revealed that most of them did not actually pursue it by the time they were 25 years old because they believed that children and STEM work are incompatible (Frome, Alfeld, Eccles, & Barber, 2006). Female graduate students and postdocs were also found to have lowered their ambitions in STEM academia due to their desire for children (Mason & Goulden, 2004).

2.2.2.2. Gender roles and values

Although many high-qualified women *leave* the STEM industry and academia, a much larger contributor to the gender gap is that girls are much less likely than boys to *choose them in the first place*. Existing evidence shows that gender roles and values may greatly affect course-taking patterns and selection of a major or occupation (e.g., Archer, DeWitt, et al., 2013; Su, Rounds, & Armstrong, 2009; Weisgram, Dinella, & Fulcher, 2011).

From a social psychological perspective, gender roles are defined as society's shared beliefs and expectations about a range of behaviours and attitudes that are considered acceptable, appropriate, or desirable for people based on their socially identified sex (Eagly, 2013). Gender roles are usually centred on the conceptualisations of femininity and masculinity. Although historical trends have seen changes in the feminine gender roles in the past few decades as women begin to take on multiple roles to a greater extent (e.g., female scientists, female CEOs), the essence of femininity remains unchanged (Weisgram et al., 2011). Communal traits are usually cited to describe femininity including being warm, gentle, empathetic, sensitive, soft-spoken as well as concerned about physical appearance. In contrast, men are expected to display more agentic traits – assertive, competitive, dominant and deference to facts over feelings (Diekman, Brown, Johnston, & Clark, 2010; Rudman & Glick, 2001; Witt & Wood, 2010).

Several researchers argue that women are reluctant to enter STEM fields because they believe that STEM careers are in conflict with their gender roles and values (e.g., Diekman et al., 2010; Su et al., 2009; Weisgram et al., 2011). Based on the vocational interest literature, STEM professions are often described as “thing-oriented” that involve manipulating things, machines, objects, tools and animals, whereas non-STEM professions are described as “people-oriented” that involve informing, training, or helping other people (Feist, 2008; Lippa, 2005). People-and-thing oriented individuals can also be termed “empathisers” and “systemisers”, respectively (Baron-Cohen, 2002; Manson & Winterbottom, 2012). Several researchers argue that women are more likely to be empathisers whereas men are more likely to be systemisers. Thus women are more interested in “people-oriented” fields of art and nursing, but men are more interested in “thing-oriented” fields of science and engineering (e.g., Auyeung, Allison, Wheelwright, & Baron-Cohen, 2012; Billington, Baron-Cohen, & Wheelwright, 2007; Manson & Winterbottom, 2012; Su et al., 2009). Studies with university students in America also revealed that female college students and those high in femininity showed stronger altruistic values, whereas men and those high in masculinity were more focused on monetary rewards in future careers (Diekman et al., 2010). The more feminine and less masculine a woman, the more likely she is to avoid male-dominated STEM careers (Weisgram et al., 2011).

Women need to overcome more difficulties, such as the somewhat contradictory values of being feminine and being science professionals, than men when pursuing a STEM career. During adolescence, high-performing girls report believing that boys find them unattractive because they defy the traditional female gender roles (Archer, DeWitt, et al., 2013; Kessels, 2005). A longitudinal study with 128 American female science undergraduates uncovered a phenomenon that female science majors who reported to feel less feminine showed lower self-esteem, greater depression, and lower grades than their counterparts who successfully reconciled their gender role and science identity (Settles, Jellison, & Pratt-Hyatt, 2009).

2.2.2.3. Stereotypes of scientists

Another cluster of studies focuses on stereotypes about science majors, occupations and individuals in these fields that can serve as deterrents to entry into STEM fields, especially for women. The psychological literature differentiates between implicit and explicit stereotypes. Implicit stereotypes are assessed via different reaction time to stimuli that are either paired in a stereotypical way (e.g., men and science) or not (e.g., women and science). Implicit measures of stereotypes such as the Implicit Association Test (IAT) have the advantage of being resistant - or at least less susceptible – to social desirability bias (i.e., the tendency of respondents to report an answer in a way they deem to be more socially acceptable than would be their personal answer to project a favourable image of themselves and to avoid negative evaluations; Callegaro, 2008) as participants are either unaware that their stereotypes are being assessed or do not have enough cognitive capacity to alter they immediate spontaneous responses. Explicit stereotypes are assessed via self-report questionnaires and participants can usually discern what is being measured and possibly give politically correct answers (Asendorpf, Base, & Mücke, 2002; Brunel, Tietje, & Greenwald, 2004).

The most robust stereotype about scientists is that they are often perceived as one gender - men. On both implicit and explicit measures of math and science stereotypes, women are less likely than men to associate math and science with themselves in childhood (Cvencek, Greenwald, & Meltzoff, 2016; Cvencek, Meltzoff, & Greenwald, 2011), in adolescence (Kurtz-Costes, Rowley, Harris-Britt, & Woods, 2008; Winter, 2008) as well as in young adulthood (Lane, Goh, & Driver-Linn, 2012). The gender-science measures also reveal that men are more likely to be associated with science whereas women are more likely to be associated with humanities by both women and men implicitly and explicitly (Cai, Luo, Shi, Liu, & Yang, 2016; Carli, Alawa, Lee, Zhao, & Kim, 2016).

Beyond the gender stereotype, other stereotypical traits related to scientists are more documented qualitatively by studies using the "Draw-A-Scientist" method (Chambers, 1983). When asked to draw a picture of a scientist, people most often depict a middle-aged white man, dressing in a lab coat and wearing glasses, doing dangerous experiments alone (Barman, 1997; Finson, 2002; Steinke et al., 2007). Such stereotypical pictures of scientists can emerge as early as Year 2 (Losh, 2010) in elementary school and have been found to last untill adulthood (Finson, 2002). Cross-cultural studies found that images of scientists drawn by students from China matched those from Western culture (Chambers, 1983) and that both Black and Caucasian children in Australia had similar stereotypical images of scientists (Finson, 2002).

These stereotypes of scientists have been found related to academic performance and major choices. Women who implicitly associate math with men were found less likely to engage in math study and have more math anxiety (Nosek & Smyth, 2011). Murphy, Steele and Gross (2007) demonstrated that stereotypes not only affect academic performance, but also the sense of belonging or identification with STEM fields. In order to consider a major or occupation, one must have a sense that one belongs or at least could belong. For example, a series of work by Cheryan and colleagues suggests that the stereotypes of scientists could make women feel they do not belong and dissuade them, but not men, from considering majoring in computer science (Cheryan, 2012; Cheryan et al., 2015; Cheryan, Plaut, Davies, & Steele, 2009).

2.3. Justification of the present study

As noted before, recent work investigating the gender gap in STEM fields has switched its attention from the disparity in competence to other social and psychological factors. Given that the work-life balance consideration is a general issue for women across all fields and no existing data support that STEM fields are more gruelling than other fields (Lubinski & Benbow, 2006), this factor is not included in the present study.

The present study attempted to investigate a relatively novel factor that may perpetuate the underrepresentation of women in science: stereotypes of empathy in scientists. As presented above, it is clear that there is a mismatch between female gender roles and values and the stereotypes of scientists. The unmatched stereotypes can make it difficult for women to imagine themselves as a scientist, as such reduce their interest in science. Reasons to focus on the stereotypes of one particular trait – empathy, are provided as follows.

Over the past 50 years, strong social norms have developed in Western societies that encourage people to curb the overt expression of prejudice and stereotypes based on social categories such as gender and race (Klonis, Plant, & Devine, 2005). Yet, approaches applying new measurement of stereotypes, focusing on implicit evaluations and beliefs that are not under conscious control, suggest that such suppressed thoughts are not simply eliminated but may leak out in more subtle ways. Recent work has demonstrated that one of the subtle ways is through feature-based stereotyping, wherein stereotypic inferences are made on the basis of such features both within social categories and between them (Ko, Muller, Judd, & Stapel, 2008). For example, both Whites and African Americans with more Afrocentric facial features (e.g., full lips and wide noses) are seen as more athletic and aggressive (Blair, Judd, Sadler, & Jenkins, 2002) and both women and men with more feminine voices are judged as more warm, but less competitive (Ko, Judd, & Blair, 2006).

Such findings suggest that the recent active attempts to suppress category-based stereotyping may lead to rebound in another form, with greater feature-based stereotyping following (or even during) the process of avoiding using one's social categories as the basis for judgment. This has been the main assertion that motivated the present research. Regarding the issue of women's underrepresentation in STEM fields, we argue that instead of referring to gender-based stereotypes, students may look to other characteristics as a feature-based cue to determine whether they fit the image of particular careers (Meltzoff, 2013).

An interesting experiment conducted by Cheryan and colleagues (2011) revealed that, regardless of the gender of the scientist representative, female undergraduates who interacted with a stereotypical representative who embodied "geeky" characteristics (i.e., wearing glasses, T-shirt saying "I code therefore I am", and playing video games) for only 2 minutes reported significantly less interest in majoring in science than the control group who interacted with a "normal" scientist representative wearing a solid coloured T-shirt and enjoying hanging out with friends. In this experiment, the traits that individual representative displayed (i.e., person schemas, meaning judgments about the traits that a specific individual possess) seemed to influence the judgment that participants had based on the subject that the representative majored in (i.e., role schemas, meaning norms and expected behaviours of people in particular roles and social types). A schema is defined as a mental structure we use to organise and simplify our knowledge of the world around us. Different schemas act like filters, accentuating and downplaying various elements that we notice. Schemas therefore affect how we interpret things and how we make decisions. Results of Cheryan's experiment showed that, it might not be the gender itself, but certain other features (e.g., appearance or personality traits) of the exemplars that participants noticed in the face-to-face interaction that were coded into the role schemas of scientists. These role schemas were later deemed in contrast to female gender roles and values, which steered female participants away from science. In other words, such evidence indicates that it may be feature-based stereotyping, rather than category-based stereotyping, that functions as the deterrent for women's participation in STEM fields.

According to the Implicit Personality Theory (IPT), we like to see people in as consistent a way as possible and we all have 'implicit' theories about which personality characteristics tend to go together or cluster (Zwbrowitz, 1990). The IPT enables us to infer what people are like when we have very limited information about them. As proposed by Asch (1946), certain central perceived traits that we believe one possesses may determine our overall impression of that person. For example, if we evaluate someone as 'warm', it may produce an overall positive impression, while 'cold' produces an overall negative impression. Stereotype can be thought of as a special kind of IPT that relate to an entire social group.

In contrast to the commonly recognised view of stereotypes as coherently positive or negative, Operario and Fiske (2004) claim that stereotypes can be ambivalent: they comprise both positive and negative attributes about social groups. According to their Stereotype Content Model (SCM), group stereotypes and interpersonal impressions form along two basic dimensions: warmth and competence. Based on such distinction, they argue that, though the two dimensions are conceptually independent, subjectively positive stereotypes on one dimension often are functionally consistent with unflattering stereotypes on the other dimension (Fiske, Cuddy, Glick, & Xu, 2002). Perceived 'competent' groups (e.g., business women, Jews, rich people) can be viewed as 'not nice', while the perceived 'incompetent' groups (e.g., housewives, the mentally retarded, the elderly) can be viewed as 'nice' people. Based on the SCM model, scientists are perceived as 'competent but cold'. While people respect the accomplishments of scientists, they don't like them. Nevertheless, people rate writers and doctors as competent and warm at the same time (Fiske & Dupree, 2014)..

As only a fundamental dimension of social perceptions, we still don't know what makes people think scientists are not warm (yet doctors and writers are warm). Can we break down warmth scientifically? Is there a specific perceived trait of scientists that makes people not like them? In the present study, we propose that empathy, as one key element in interpersonal relationships, may very likely work as the feature-based cue for people to determine their assessment of a certain profession's warmth. As Fiske and Dupree (2014) pointed out, competent professionals whose job involves communicating and caring seem warmer and more trustworthy. Empathy is often considered more related to humanities, but not sciences (Carroll & Chiew, 2006; Manson & Winterbottom, 2012). Furthermore, given that women often have higher perceived empathy than men (Baron-Cohen & Wheelwright, 2004; Rueckert, 2011), it is reasonable to hypothesize that the stereotypes of empathy in scientists may serve as gatekeepers, driving women, who believe that scientists are socially isolated, away from certain male-dominant STEM fields.

To date, very few studies have been conducted to examine how people perceive empathy in scientists, partly because of the lack of measurement. As mentioned earlier, stereotypes can be seen as a special kind of implicit personality theories and such cognitive processes often happen unconsciously. In this case, it is of great importance to develop a new measurement to assess people's unconscious biases, which are difficult to gauge using self-report questionnaires. In the present study, I systematically developed a new implicit association test attempting to investigate stereotypes of empathy in scientists by gender and major subject as well as the interplay of the implicit and explicit stereotypes in science career aspirations among UK university students.

Moreover, it is worth noting that although I treat stereotyping of scientists as a potential negative factor that can steer women away from STEM fields in the present study, categorisation of people is normal and expected from a cognitive point of view. Traditionally, American psychologists studied stereotypes in relation to prejudice and regarded them as false, illogical over-generalizations based on group membership (e.g., Allport, 1954; Lippmann, 1922). Nevertheless, many European psychologists saw stereotyping as a normal mental shortcut (e.g., Operario & Fiske, 2004; Tajfel & Turner, 2001). In addition, given that perceptions about individuals may translate into stereotypes about the social roles that the person engages in, many researchers propose that counter-stereotypical role models can be used as interventions to change traditional stereotypes about certain professions. A counter-stereotypical role model is an individual who engages in a role that is antithetical to traditional stereotypes (e.g., a female scientist, a well-dressed scientist, or a scientist surrounded by people in a party). If the knowledge about an exemplar of a social group forms cognitive schemas, which give rise to stereotypical beliefs and influence behaviour, interventions involving exposure to counter-stereotypical role models should reduce stereotypes and enhance counter-stereotypical aspirations (Wood & Eagly, 2012). For example, Dasgupta and Asgari (2004) presented female students with photos and descriptions of famous women in leadership positions in counter-stereotypical fields such as science, business, law, and politics. Following such exposure to counter-stereotypical women, female participants were quicker to associate women with leadership in an IAT test. The results suggest that exposure to counter-stereotypical role models can act as intervention to reduce stereotypical cognition. Furthermore, the majority of research efforts seeking explanations for women's underrepresentation in STEM fields have been conducted in the United States (Beyer, 2015). However, as presented before, shortages of women in physical science and engineering in higher education as well as in workforce also exist in the United Kingdom. Studies looking at the images of scientists are limited in the UK and mostly conducted with children and adolescents using general questionnaire survey and interview data, such as the ASPIRES project (Archer, DeWitt, et al., 2013). The present study is, to my knowledge, the first attempt to investigate both the implicit and explicit stereotypes of empathy in scientists in higher education in the United Kingdom.

Chapter 3 Literature Review

Given that the present study attempts to develop a new Implicit Association Test to investigate stereotypes of empathy in scientists by gender and major subject, this chapter is to critically review three groups of relevant literature, including 1) studies about empathy, gender and science; 2) studies about stereotyping and images of scientists and 3) studies about Implicit Association Test. The pitfalls and research gaps of the existing studies are pondered upon. Based on the literature review, the chapter ends with introducing the current study, including its research questions, hypotheses and research design.

3.1. Gender, empathy and science

Comparatively little academic work has addressed specifically people's perceptions of empathy in scientists, yet existing studies indicate that scientists could possess a cognitive style that is less empathetic but more systematic than those in non-science fields. The cognitive style could be an important factor to explain the gender difference in certain science fields. In the following sections, theories and empirical evidence about the relations between gender, empathy and science will be reviewed. Issues with existing studies about the relationships between empathy and science are discussed with the purpose to demonstrate why it is important to study stereotypes of empathy in scientists in the present study.

3.1.1. Defining empathy

As mentioned earlier, the present study has placed its focus on the construct of empathy because of its potential effects on individuals' major selection and career aspirations. Before we review the literature about relations between gender, empathy and science, it is important to clarify the definition of empathy.

Ever since the term ‘empathy’ was coined by Titchener (1909) as a translation of the

German word Einfühlung (i.e., an aesthetic term meaning "to project yourself into what you observe"), it has been an important concept in contemporary developmental, social, personality, and clinical psychology. Perhaps because of its wide-ranging application, the notion of empathy has been a broad and somewhat slippery concept (Eisenberg, 2002). Some researchers use the term empathy to refer to the cognitive process of perspective taking (e.g., Einolf, 2012); others take it to mean an affective process (though having some cognitive elements) such as the ability to experience the same feelings of another person (e.g., Moore, 2004); still from the clinicians' perspective, empathy is viewed as a process that serves a communication function in therapy (Hojat, DeSantis, & Gonnella, 2017).

With the recognition of its multidimentional nature, we define empathy as the ability to take another person's perspective, understand the feelings of another and interact within different social situations (Thomson, Wurtzburg, & Centifanti, 2015). This working definition aims to include cognitive, affective and behavioural components encompassed by empathy. Under this view, empathy consists of 1) cognitive empathy, which means the ability to identify and understand another person's point of view; 2) emotional empathy, which means the ability to be sensitive to and vicariously experience the feelings of others; and 3) social skills, which means the ability to give emotionally appropriate reactions in different social situations.

Empathy is without question an important social ability that allows us to interact effectively with each other in the social world. It has been taken as a desirable characteristic that positively correlates with interpersonal relationships and mental health. For example, participants who reported higher empathy also reported more prosocial behaviours (Davis et al., 2001), were less aggressive (Capage & Watson, 2001), and had more supportive peer relationships in both adolescence and adulthood (Davis et al., 2001; Eisenberg, Morris, McDaniel, & Spinrad, 2009). Empathy deficits are related to social distress and anxiety and may even lead to offensive behaviours (Geer, Estupinan, & Manguno-Mire, 2000).

3.1.2. Empathy and the gender gap in science

From the bio-cognitive view, some researchers argue that men and women have different cognitive styles that lead them to varied academic majors and professions (e.g., Baron-Cohen, Knickmeyer, & Belmonte, 2005; Billington et al., 2007; Manson & Winterbottom, 2012). The attempt to categorise individuals by different cognitive styles is in the shape of empathising - systemising theory (Baron-Cohen, 2002). According to Baron-Cohen and his colleagues, empathising is defined as the ability to read emotions and thoughts in others and to give appropriate responses (Baron-Cohen & Wheelwright, 2004). Systemising is defined as the ability to analyse and construct systems (Baron-Cohen, Richler, Bisarya, Gurunathan, & Wheelwright, 2003). Individuals can possess different cognitive styles due to their levels of empathising and systemising skills.

Two important hypotheses have emerged from the cognitive style theory to explain the underrepresentation of women in STEM fields. First is that women tend to be empathisers while men tend to be systemisers (e.g., Baron-Cohen, Knickmeyer, & Belmonte, 2005). Second is that systemisers tend to enter the STEM fields while empathisers are more likely to choose humanities and social sciences (e.g., Billington et al., 2007; Manson & Winterbottom, 2012; Wheelwright et al., 2006). Nevertheless, existing research has provided controversial evidence for both hypotheses.

In the following sections, empirical studies examining gender differences in empathy from the biological and developmental perspective using different kinds of measurements are critically reviewed. Existing research on the relationship between empathy and academic major selection is also carefully pondered upon.

3.1.2.1. Do women have better empathy than men?

Studies using Empathising Quotient (EQ) questionnaire and Systemising Quotient (SQ) questionnaire measuring empathising and systemising skills consistently report a female advantage in empathy and a male advantage in systemising (Baron-Cohen et al., 2003; Focquaert, Steven, Wolford, Colden, & Gazzaniga, 2007; Lawson, Baron-Cohen, & Wheelwright, 2004; Manson & Winterbottom, 2012; Wheelwright et al., 2006). These findings have been cited as evidence for the claims proposed by Baron-Cohen (2002) that men are predisposed to learn about objects and their mechanical relations, whereas women are predisposed to learn about people and their emotional interactions. However, close inspections of studies on sex-linked hormones, cognitive development in human infants, children and adults, studies using different kinds of measurements have suggested a rather nuanced picture about gender differences in empathy.

3.1.2.1.1. Biological sex difference in empathy

Firstly, research on the biological basis of the E-S cognitive style suggests that gender differences in empathy may be related to sex-linked genes (Chakrabarti et al., 2009) and hormones (Baron-Cohen et al., 2005). It is argued that the more prenatal testosterone (i.e., a sex steroid hormone that men produce twice as much as women in the womb) one has been exposed to, the more likely one may have a low empathising but high systemising cognitive profile (Chapman et al., 2006). Prenatal testosterone exposure is usually assessed by measuring the ratio of the length of the second (ring) finger and fourth (ring) finger (i.e., 2D:4D digit ratio). Individuals with a low empathizing, but high systemizing, cognitive profile have been found more likely to have a lower 2D:4D (more masculine) finger ratio than people with the opposite cognitive profile (Manning et al., 2010; Von Horn et al., 2010; Wakabayashi & Nakazawa, 2010).

However, Hönekopp's (2012) meta-analysis of the correlations between 2D:4D finger ratio and E-S cognitive style suggests that the influence of prenatal testosterone on cognitive style is usually significant among individuals with autistic spectrum conditions, but not with typically developing individuals. Moreover, when looking at the typically developing population in particular, a significant correlation of 2D:4D finger ratio was usually found with systemising ability (Manning et al., 2010) but not with empathy or emotion recognition performance (Teatero & Netley, 2013; Voracek & Dressler, 2006). All these findings suggest that prenatal testosterone may promote an extreme low empathy but high systemising cognitive style only in people with a specific cognitive profile (e.g., autistic spectrum conditions), but may not substantially contribute to the gender difference in empathy among typically developing individuals.

3.1.2.1.2. Developmental gender differences in empathy

Although sex hormones may not directly relate to empathy, research on the development of social cognition indeed reveals a female advantage on empathy to some extent. A great number of studies focusing on the development of theory of mind (ToM), the ability to reason about others' mental states (Baron-Cohen, Leslie, & Frith, 1985), suggest that gender differences in social cognition appear early in life. A study of one-day born infants revealed that female infants preferred to look longer at the active person, whereas male infants looked longer at a similar sized inanimate object (Connellan, Baron-Cohen, Wheelwright, Batki, & Ahluwalia, 2000). However, it is worth mentioning that though more girls showed preference for human faces over objects than boys (36% versus 17% of the sample) in their study, the largest group of newborn girls tested in this study actually showed no preference (47%). Therefore, it is possible that the claimed gender difference might not be that evident and this experiment needs to be independently replicated. Furthermore, Fine (2010) points out that parents may not have been totally blind to the baby's sex before it is born and there might have already been gendered 'congratulations' cards and gifts (e.g., pink for girls and blue for boys) around the bed. As such, there could be social influence on newborn babies and we should not simply attribute the gender differences shown in them to biology.

For preschoolers, evidence indicated that girls had better knowledge of the distinction between real and apparent emotion than their male counterparts (Banerjee, 1997) and grasped the concept of false belief (i.e., the understanding that others can have beliefs that are in contrast to the reality) earlier than boys (Charman, Ruffman, & Clements, 2002). Though all typically developing children would pass the false-belief task sooner or later around age 4 (Bloom & German, 2000), girls still outperformed boys in judging the intentions and emotions of characters in stories and movies during middle childhood (Calero, Salles, Semelman, & Sigman, 2013; Devine & Hughes, 2013).

During adolescence, girls usually score higher than boys in self-report empathy (Allemand, Steiger, & Fend, 2015; Van der Graaff et al., 2014) but this female advantage does not exist in studies using neurophysiological measures of empathic arousal (Michalska, Kinzler, & Decety, 2013). Longitudinal studies of adolescent empathy found that girls' empathic concern remained stable and higher than boys during adolescence, whereas boys showed a decrease in empathic concern from early to middle adolescence with a rebound to the initial level thereafter (Mestre, Samper, Frias, & Tur, 2009). A potential explanation for the gender gap in social cognition development could be that girls enjoy a more emotion-oriented growing environment than boys. A meta-analysis of studies examining gender effects on parental language indicated that mothers tend to talk more and to use more supportive speech with daughters than with sons during toddlerhood (Leaper et al., 1998) Moreover, older siblings are also found to mention feeling states more frequently to girls than boys (Brown, Donelan-McCall, & Dunn, 1996).

3.1.2.1.3. Gender differences by different measures of empathy

Instead of a rather consistent female advantage found in childhood, research of adults has produced mixed findings about gender contrasts in empathy assessed by various measurements. The most robust evidence of a female advantage in empathy is observed in studies that employ self-report measures (Eisenberg & Lennon, 1983; Michalska et al., 2013). For instance, women were found scoring significantly higher than men on the Empathy Quotient (EQ) questionnaire, in which participants were required to rate themselves on statements such as: “I tend to get emotionally involved with a friend's problems” (Baron-Cohen & Wheelwright, 2004; Nettle, 2007; Wakabayashi et al., 2007; Wakabayashi, Sasaki, & Ogawa, 2012). Studies relying on other self-report questionnaires such as the Interpersonal Reactivity Index (Davis, 1983) also revealed a female advantage in empathy (Berg et al., 2015; Harton & Lyons, 2003; Rueckert, 2011).

Beyond self-reported empathy, behavioural tasks provide mixed evidence for gender differences in empathy. Researchers utilising emotion reading tests found that women outperformed men in telling different emotions from static pictures of eyes (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Chapman et al., 2006; Voracek & Dressler, 2006). Nonetheless, no such female advantage was observed in tasks requiring participants to infer the specific content of protagonists' thoughts and feelings in stories and movies (Hall, 1978; Ickes, Gesn, & Graham, 2000; Russell, Tchanturia, Rahman, & Schmidt, 2007). Moreover, when measures that rely on physiological and neuroimaging data are considered, reports of gender contrasts on empathy become even less clear (Michalska et al., 2013). For instance, Eisenberg and Lennon's (1983) meta-analysis of studies of empathy demonstrated that changes in autonomic nervous system activity such as electrodermal activity, heart rate, blood pressure, as well as gestural measures, showed no clear evidence for gender differences in either children or adults, while self-report questionnaires and (to a lesser extent) emotion identification tasks did reveal a female advantage. In addition, a meta-analysis of 65 neuroimaging studies of responses to emotional stimuli found no greater activation to viewing emotional materials for women than men (Wager, Phan, Liberzon, & Taylor, 2003).

To sum up, existing studies suggest a ubiquitous female advantage in self-report empathy but not in the actual empathetic ability, meaning that women are more likely than men to perceive themselves as empathetic, yet they may not actually have better empathetic abilities than men. From the biological perspective, gender-linked hormones have been found only contributing significantly to people with special cognitive profiles but not in a typically developing population. Moreover, from the developmental perspective, girls have displayed a consistent advantage in empathy during childhood and adolescence as compared to boys, but such developmental advantage does not lead to a prominent advantage in adulthood. Though boys have some fluctuations in empathy development during adolescence, men have shown comparative emotional arousal and empathetic skills with women during adulthood. Despite the controversial evidence for the female advantage in empathetic ability, women indeed showed more empathetic concerns and consistently reported higher perceived empathy about themselves than men.

3.1.2.2. Are empathy and science incompatible?

According to the second hypothesis of the empathising-systemising theory, people who study sciences have been described as lacking in empathy but outperformed in systemising when compared to people who study humanities (Billington et al., 2007; Focquaert et al., 2007; Manson & Winterbottom, 2012). This hypothesis is firstly examined by a large-scale study among 1761 students from Cambridge, revealing that physical science majors scored significantly lower than those from biological science, social science and humanities fields in the EQ questionnaire but higher in the Systemising Quotient (SQ) questionnaire (Wheelwright et al., 2006). Systemising is defined as the drive to analyse and construct systems by understanding the rules that govern the system (Baron-Cohen, 2002). Participants are asked to rate themselves on statements such as “when I learn a language, I become intrigued by its grammatical rules” (Wheelwright et al., 2006). The differences in empathising-systemising cognitive styles by major subject were later replicated by Manson and Winterbottom (2012) with another 321 Cambridge students. In this study, they also claimed that cognitive style is a better predictor than gender for major selection using logistic regression to compare the predictive power between gender and cognitive style to major choices (Manson &

Winterbottom, 2012). Similar consequences were also found in cross-cultural studies among students from Belgium (Focquaert et al., 2007) and Malaysia (Zeyer et al., 2013).

These studies have made crucial headway in exploring the link between empathy and major selection, but they are still prone to pitfalls and several research gaps exist. First of all, most existing studies are conducted under the framework of the empathising-systemising cognitive style, but it is important to note that empathising has been found largely independent from systemising, therefore there is *not* necessarily a trade-off between empathising and systemising (Carroll & Chiew, 2006; Lawson et al., 2004). That is to say, even though scientists may be categorised under the label "systemiser" with a relatively higher systemising skill than empathising skill, it does not automatically mean that his or her empathy is absolutely low. Researchers must be aware that it is not the discrepancy between empathising and systemising that determines whether one chooses science majors or not. Given that women and men are compatible with high levels of mathematical ability and achievement, it would be intriguing to focus on the unique role of empathy in major selection.

Secondly, female and male scientists may show different levels of empathy. It has been proposed that women and man may employ different strategies when making major subject decisions (Ceci, Ginther, Kahn, & Williams, 2014). An interesting study conducted by Valla et al. (2010) among 144 undergraduate students revealed that being in a scientific field of study was associated with poor empathising skills in men, but not in women. Moreover, they also found that systemising ability was coupled with sacrifices in empathising abilities in men but not in women and high systemising is associated with science major selection for women but not for men. Based on these findings, Valla et al. (2010) then argue that men's choice of science field of study could more be associated with weakness in empathy regardless of their systemising ability, but women's choice of science could more be associated with their strength in systemising. Among women and men with comparably outstanding mathematical aptitude and achievement, women are more likely to have outstanding empathetic skills (at least perceived empathy) at the same time (Thomson et al., 2015; Valla et al., 2010). Therefore, empathy and science should definitely not simply be seen as incompatible constructs. Scientists do not necessarily have a higher-systemising-lower-empathising cognitive style, yet they can have both good mathematical ability and interpersonal skills at the same time.

Moreover, variations within the science field should also not be ignored. Though all STEM majors require students to systematically study the natural world through observation and experiment, life science, which requires students to take ethics into consideration, can be more 'people-oriented' than physical science. Performance in subjects like medicine requires high levels of empathy in order to be successful (Hirsch, 2007; Wright, McKendree, Morgan, Allgar, & Brown, 2014). Biology students may have plenty of opportunities to cooperate with their colleagues in labs whereas students majoring in computer science often instead focus on highly independent coding jobs on their own (Diekman, Clark, Johnston, Brown, & Steinberg, 2011). Supporting this claim, Thomson et al.'s (2015) survey of empathy among 404 undergraduates showed equivalent self-report empathy among students in life science and social science. Khorashad et al. (2015) even reported an advantage of medical students over students from all other majors in emotion recognition in Persia.

The final caveat springs from the applied measurements for empathy and calls for extra attention to the interpretations of results. As we saw earlier, the distinct female advantage in empathy captured by self-reports was not that evident in ability tasks and people can over-report or under-report certain behaviours due to their personal bias or social pressure (Zerbe & Paulhus, 1987). Despite a wealth of self-report data supporting the claim that scientists have lower empathy than people from other fields, studies utilising ability tasks revealed mixed findings. For instance, Billington et al. (2007) found a significant advantage in emotion identification task for students studying humanities, but the effect is only marginal in the study carried out by Carroll and Chiew (2006) and did not even exist in a more recent study carried out by Khorashad et al.

(2015). It is possible that science majors underestimate their empathetic ability due to the impact of certain stereotypes about scientists. Related research about stereotyping and image of scientists will be discussed later.

To sum up, existing studies revealed mixed evidence for the claim that scientists are lacking in empathy. Although some studies found that scientists scored lower than other majors in self-report empathy, there is no evidence for a trade-off between scientific ability and empathetic ability in general. Students majoring in life science and medicine even showed some advantage in empathy over other humanities majors. The relations between empathy and science major selection were gender-dependent, for which only certain men may choose to study science due to poor empathetic skills but women in science could have both good interpersonal skills and scientific competence. After all, scientists have shown comparable empathetic ability with people from other fields, but they underestimate their empathy in self-reports. I suspect that scientists' inaccurate perception of their empathy may stem from the stereotypical image of scientists being socially awkward. Therefore, the present study proposed to look at people's stereotypes about empathy in scientists and the potential variations in such stereotypes by gender and major subject.

3.2. Implicit and explicit stereotypes of scientists

As discussed above, existing evidence of a consistent disadvantage in self-report empathy, but not their empathetic abilities, among science majors indicates that stereotypes about scientists may be the reason why scientists underestimate their own empathy. Moreover, many researchers have argued that the existence of stereotypes about scientists may also influence the appeal of STEM majors and careers to certain individuals (e.g., Cheryan et al., 2015; Schmader, Johns, & Forbes, 2008).

In the following sections, related theories and empirical studies about implicit and explicit stereotypes as well as images of scientists will be critically reviewed. Readers will be equipped with all the necessary information about how implicit and explicit stereotypes differ and are related, the existing empirical evidence of the content of implicit and explicit stereotypes of scientists, as well as factors related to individual differences in stereotypes of scientists, setting up a stage to further discuss the limitations of existing studies and demonstrate the research gap that the present study aimed to address.

3.2.1. Conceptualising implicit and explicit stereotypes

3.2.1.1. Defining stereotypes and related constructs

Before we approach the definitions of implicit and explicit stereotypes, it is important to first and foremost define *stereotype* and its related constructs *prejudice* and *discrimination*. Though laypeople may use these terms synonymously, psychologists draw distinctions between each concept. *Stereotypes* are usually defined as associations and attributions of specific characteristics to groups; *prejudice* as attitudes and feelings reflecting an overall evaluation of a group, and *discrimination* as biased behaviours toward, and treatment of, individuals due to group membership (Dovidio, Hewstone, Glick, & Esses, 2010).

Stereotypes are cognitive schemas used by social perceivers to process information about others, and it has the function of simplifying a complex environment (Hilton & von Hippel, 1996). Prejudice is conceptualized as an attitude that has both a cognitive component (e.g., beliefs about a target group) and an affective component (e.g., like or dislike; positive or negative) that creates or maintains unequal status and role differences between groups (Eagly & Diekman, 2005). When it comes to discrimination, it implies more than simply distinguishing among social objects, but refers also to actively negative behaviour or, more subtly less positive responses, that create, maintain, or reinforce advantages for certain groups and their members over other groups and their members (Mummendey, Otten, Berger, & Kessler, 2000).

While discriminations can occur toward a group or its members, stereotypes and prejudices are intrapsychical phenomena that occur within an individual. That is, stereotypes and prejudices may vary not only in their transparency to others but also in the level of awareness to the person who possesses them (Dovidio et al., 2010). Specifically, people can implicitly hold cognitive representations of certain beliefs without explicitly endorsing such stereotypes, without feelings of prejudice, and without awareness that such stereotypes could affect one's judgment and behaviour.

3.2.1.2. Defining implicit and explicit stereotypes

The idea of *implicit/explicit* stereotypes is based on the dual-process model of social cognition which claims there are two modes of thought: one is the *controlled* processing denotes mental functioning that is conscious, deliberative, effortful, and/or voluntary in nature, whereas the *automatic* processing is unconscious, unintentional, effortless, and/or involuntary (Bargh, Chen, & Burrows, 1996; Chaiken & Trope, 1999; E. R. Smith & DeCoster, 2000). Explicit stereotypes are beliefs and attitudes people know they hold, subject to deliberate (often strategic) control in their expression (Fazio et al., 1995). In contrast, implicit stereotypes involve a lack of awareness and unintentional activation (Greenwald & Banaji, 1995).

Ever since Greenwald and Banaji (1995) coined the term *implicit social cognition* , researchers have been using it in different ways. Some researchers use the term implicit merely to refer to measurement procedures that gather data without asking for verbal responses (e.g., Fazio & Olson, 2003). Other researchers use the term to describe assessed constructs that do not require conscious introspection (e.g., Banaji, 2001). To overcome the confusion arising from differences in linguistic conventions and semantics that can disrupt research progress, De Houwer, Teige-Mocigemba, Spruyt, and Moors (2009) have suggested a taxonomy to outline the differences between the terms direct/indirect and explicit/implicit. They suggest using the terms *direct* and *indirect* to describe features of measurement procedures whereas the terms *implicit* and *explicit* for features of the psychological attributes captured by different measurement procedures. In the context of the present study, we use the term *implicit* as the synonym of *unconscious* to refer to the automatic and uncontrolled mental processes and *explicit* as the synonym of *conscious* to refer to the deliberate and controlled mental processes.

Accordingly, explicit stereotypes are usually assessed directly by self-report measures whereas implicit stereotypes are "inaccessible to introspection" therefore must be measured indirectly (Greenwald & Banaji, 1995). A widely used direct measurement of explicit attitudes is the Semantic-Differential Method. The Semantic-Differential procedure involves presenting individuals with opposing adjective pairs and asking them to locate the object on a rating scale anchored by the opposing adjectives (Osgood, 1952). For example, the object may be *dog*, and adjectives might include *smart-foolish*, *pleasant-unpleasant*. When applying the Semantic-Differential scales, participants are directly asked to report their attitudes toward certain objects, thus are able to consciously reflect on their responses, therefore such method is widely considered a direct way of capturing explicit attitudes. However, this type of instrument cannot capture implicit attitudes.

In order for a measure to be called implicit, it needs to evoke the to-be-measured attribute to automatically cause the measurement outcome in the absence of substantial cognitive resources, substantial time, and awareness of certain goals (Moors & De Houwer, 2006). As such, implicit measurement procedures often rely on computerized stimuli and recording of reaction time latency. The most widely used implicit measurement is the Implicit Association Test, which assesses participants' reaction time to compatible or incompatible pairs of concepts (see later Section 3.3 for a full review).

A number of cognitive models help us to understand the distinction and relationship between the implicit and explicit cognition. The following sections will cover the most dominant models and theories for the implicit and explicit stereotypes, as they are crucial for understanding and interpreting the measured outcomes in the present study.

3.2.1.3. Theoretical models for the implicit and explicit stereotypes

There are many different cognitive theories drawing distinctions between aspects of cognitive processes. These postulate alternative explanations through which implicit and explicit stereotypes may have an impact on behaviours. These theories can be summarized into three theoretical models: 1) *double dissociation*, that implicit and explicit attitudes are completely independent of each other, thus implicit attitudes predict spontaneous behavior whereas explicit attitudes predict deliberative behaviour (Dovidio, Kawakami, & Gaertner, 2002) 2) *additive*, that the implicit and explicit measures reflect a single construct with different procedures, therefore attitudes inferred by the two types of measures explain different portion of variance in the behaviour (Fazio & Olson, 2003); and 3) *interactive*, that the implicit and explicit attitudes are distinct but related, as such they interact in influencing behaviour (Strack & Deutsch, 2004).

3.2.1.3.1. <u>Double dissociation model</u>

At one extreme, some researchers argue that the implicit and explicit stereotypes are completely independent of each other and may give rise to different kinds of behaviours. Based on this dual-processing model, implicit and explicit attitudes are viewed as two separate constructs that coexist in memory (Wilson, Lindsey, & Schooler, 2000). Implicit stereotypes are automatic and unconscious biases, whereas explicit stereotypes are deliberative and conscious evaluations. From this perspective, the low correlations (average implicit-explicit measurement correlation of 0.19) between implicit and explicit measurements of stereotypes are taken as evidence for the existence of two independent cognitive constructs, but not the discriminant validity between two different measurements of the same cognitive system (Chaiken & Trope, 1999; Wilson et al., 2000).

Given that implicit and explicit attitudes are regarded as independent constructs, they are assumed to direct different behaviours and demonstrate a *double-dissociation* pattern. Implicit attitudes are assumed to influence spontaneous actions and nonverbal behaviours that are uncontrollable or with no attempts to control them. In contrast, explicit attitudes are assumed to influence more deliberative behaviours and expressive responses that are under conscious control (Dovidio et al., 2002). This double-dissociation pattern has been confirmed in a few studies examining racial attitudes (e.g., Ashburn-Nardo, Voils, & Monteith, 2001; Green et al., 2007; McConnell & Leibold, 2001). For example, implicit negative attitudes toward Blacks were found to successfully predicted subtle less positive nonverbal behaviours (e.g., less speaking time, less smiling, less eye contact) during an interaction with a Black experimenter as compared to an interaction with a White experimenter. However, it was the explicit self-report racial attitudes, but not the implicit attitudes, that related to ratings about legitimacy of anti-Black community laws (McConnell & Leibold, 2001).

3.2.1.3.2. Additive model

At the opposite extreme, implicit and explicit attitudes are viewed as expressions of a single construct, only under different conditions. From this perspective, mental representations are similar to icebergs, with explicit stereotypes residing above the surface of conscious control and the implicit stereotypes residing below it (Fazio et al., 1995). According to the Motivation and Opportunity as Determinants model (MODE Model; Fazio & Towles-Schwen, 1999), the expression of an implicit attitude or an explicit one depends on two moderating factors: 1) whether individuals are motivated enough to consciously reflect on their beliefs and attitudes and 2) whether they are able to engage in such reflection.

The MODE model suggests that when people have the opportunity (e.g., sufficient time and cognitive ability) and motivation (e.g. concerns with social desirability) to evaluate the consequences of various actions, explicit attitudes primarily influence behaviours as people reflect on their relevant attitudes. Nevertheless, implicit attitudes take control in situations when opportunity is not permitted (e.g., because of time pressure) or the motivation is absent (e.g., because the task is not important). According to this model, implicit and explicit attitudes can be best understood as implicit and explicit measures of the same attitude: one relying on self-reports, which are assumed to allow participants to evaluate and manipulate their responses; and the other relying on reaction time tasks, which are assumed to limit the opportunity for participants to introspect their responses.

In this case, Fazio et al. (1995) also claims that the implicit attitudes that are automatically activated early in the judgment process (upstream) can manipulate selective attention and information search, and as such influence explicit judgment and behaviour (downstream). Consequently, automatically activated stereotypes can have an impact on both subtle (e.g., nonverbal reactions) as well as controlled behaviours (e.g., self-reported preferences). That is to say, if we follow the assumption of a single system with a single representation and two different measurements, the general predictive model should be an *additive* pattern, where both implicit and explicit measures of the same attitudes can cast a distinctive influence on behaviour.

3.2.1.3.3. <u>Interactive model</u>

According to the more recent models such as the Associative Propositional Evaluation Model (Gawronski & Bodenhausen, 2006) and the Reflective Impulsive Model (Strack & Deutsch, 2004) as well as the work of Smith and DeCoster (2000), implicit and explicit cognition are best comprehended in terms of associative and propositional mental processes that interact with each other. Associative processes are defined as the activation of mental associations between concepts in memory, which is assumed to be affective in nature and often has spontaneous effects on behaviour. Propositional processes are defined as the validation of information implied by the activated association, which is considered to affect controlled behaviours when one desires to evaluate the truth or falsity of certain beliefs (Gawronski & Bodenhausen, 2006). Implicit stereotypes reflect the automatic (or impulsive) *associative activation* processes in memory that are fast to form, whereas explicit stereotypes reflect *propositional validation* processes that require reasoning and logical thought (Gawronski & Bodenhausen, 2014).

From this perspective, the associative and propositional processes are viewed functionally distinct, but also assumed to mutually interact, such that the associative processes may influence the propositional processes, and vice versa. The "bottom-up" effects of associative on propositional processes work by mentally activating associated concepts that provide basic information for propositional evaluations. The "top-down" effects of propositional on associative processes work because the propositional reasoning can activate new information in the course of validating activated information. Specifically, if people are motivated to believe a particular propositional judgment, they may selectively search for related information confirming the validity of such proposition. Therefore, biased retrieval as well as activation of confirmatory information can produce mutual interactions between implicit and explicit mental processes (Galdi, Arcuri, & Gawronski, 2008; Peters & Gawronski, 2011).

Based on the interactive theoretical model, implicit and explicit cognitions should work together in predicting behaviours. Evidence collected to support this model is focused on cases where a negative implicit attitude conflicts with a positive attitude. For instance, research about people's defensive and secure self-esteem found that people with secure self-esteem (i.e., the congruence between high implicit self-esteem and high explicit self-esteem) were less narcissistic, showed less in-group bias, and engaged less in dissonance reduction compared to participants with defensive self-esteem (e.g., low implicit self-esteem but high explicit self-esteem; Jordan, Spencer, Zanna, Hoshino-Browne, & Correll, 2003).

To sum up, researchers have articulated three different theoretical models about implicit and explicit stereotypes and their relations with behaviours. The three models correspond to the situation when a) implicit stereotypes predict spontaneous behaviour and explicit stereotypes predict deliberative behaviour and not vice versa (*double dissociation* pattern), b) implicit and explicit measures of stereotypes contribute distinct impact on behaviours (*additive* pattern), and c) implicit and explicit attitudes interact synergistically to predict behaviour (*interactive* pattern). Empirical evidence from the present study examining the relationship between implicit and explicit stereotypes of empathy in scientists and science career aspirations will be interpreted and discussed with reference to these models.

3.2.1.4. Variations in implicit and explicit stereotypes

Given that the present study proposed to examine individual differences in implicit and explicit stereotypes of empathy by gender and major subject, it is also of great importance to review related theories and findings of factors that may affect variations in stereotypes.

3.2.1.4.1. <u>Are women less stereotypical than men?</u>

Gender differences are ubiquitous in the social psychology literature and there is a wide-held belief that women and men differ in prejudice and stereotyping. For example, it is almost cliché in Western cultures to state that women are more emotionally sensitive than men. Some research has found support for this stereotype, such that women actually do smile more than men (LaFrance, Hecht, & Paluck, 2003) and show more empathetic concerns than men (Koenig & Eagly, 2005; Rueckert, 2011). Given that empathy can encourage people to adopt perspectives of others and lead to prejudice reduction (Todd, Bodenhausen, Richeson, & Galinsky, 2011; Vescio, Sechrist, & Paolucci, 2003), an argument could be made that women are nicer and more supportive than men, thus are less prone to stereotypes and prejudices than men.

At first glance, there appears to be consistent evidence of the female advantage in producing less stereotypical beliefs. For example, women have reported more favourable attitudes than men on social issues such as against desegregated neighbourhood (Hughes & Tuch, 2003) and acceptance of homosexuals (Ratcliff, Lassiter, Markman, & Snyder, 2006). In comparisons, surveys have found that men reported higher levels of xenophobia, endorsement of White superiority, and racism compared to women (Ekehammar, Akrami, & Araya, 2003; Sidanius, Pratto, van Laar, & Levin, 1999).

However, inconsistent findings were revealed when researchers used implicit measures of stereotypes. Ekehammar et al. (2003) found that although women expressed lower explicit racial prejudice than men, they actually held higher implicit racial prejudice. Similarly, women showed less explicit prejudice than did men about female authority, but their implicit attitudes were similarly negative (Rudman & Kilianski, 2000). In contrast, women demonstrated significantly lower implicit prejudice than men about homosexuals and people with disabilities (Dozo, 2015).

These findings suggest that gender differences in stereotypes vary according to whether it is implicitly or explicitly measured, and according to the substantive topic in question. In the present study, we will compare both the implicit and explicit stereotypes of empathy in scientists between women and men.

3.2.1.4.2. Do people show more positive views about ingroup members explicitly but not implicitly?

Moreover, a closer examination of existing literature reveals that people's intergroup attitudes can also vary based on whether it is their self-report explicit attitudes or implicitly measured unconscious bias. Commonsense and many psychological theories (e.g., Social Identity Theory; Tajfel & Turner, 2001; Self-Categorisation Theory; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987) suggest that positive affiliation with one's own social group is a basic and fundamental fact of human existence. Cognitively, people have been found to remember more detailed as well as more positive information for ingroup than for outgroup members (Hogg & Hains, 1996; Howard & Rothbart, 1980). As a result, people help ingroup members more than outgroup members (Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997) and report relatively more positive beliefs about ingroups (Howard & Rothbart, 1980). The strong tendency to respond more positively to people from our ingroups than we do to people from outgroups is known as ingroup favouritism.

However, with the development of new instruments that allow researchers to measure implicit attitudes and beliefs, we found many conditions under which the ingroup favouritism principle does not operate as expected and an opposite outgroup favouritism emerges. Nosek, Banaji, and Greenwald's (2002) study of a large sample of White and Black participants' implicit and explicit racial attitudes via the Internet (Ns > 17,000) revealed that though African Americans explicitly expressed much stronger ingroup favouritism in self-reports, they exhibited no implicit ingroup favouritism. Furthermore, when playing a video game stimulating police chase, African American and White American participants were found equally likely to harbour implicit stereotypes associating Black with criminality and responded faster to shoot at the Black armed fictitious characters compared to the White (Correll, Park, Judd, & Wittenbrink, 2007). In some other studies, African Americans were even found to favour White Americans over their ingroup implicitly (but not explicitly; Ashburn-Nardo, Knowles, & Monteith, 2003; Dasgupta, 2004).

To explain why people show implicit outgroup favouritism, Jost and Banaji (1994) proposed the system justification theory, which maintains that all people, including those in the inferior or stigmatized groups, have a tendency to legitimize existing social norms and hierarchies even at the expense of personal and group interest. Stereotypes and prejudices about their own group are a way to satisfy that tendency to rationalize the group's social position (Jost, Banaji, & Nosek, 2004). Under such circumstances, the present study will examine "whether science majors have less stereotypical views about empathy in scientists than humanities majors?" using both implicit and explicit measures.

3.2.1.4.3. <u>Will cognitive dissonance influence implicit and explicit stereotypes of identified groups differently?</u>

The desire to maintain consistency between social cognitions including attitudes, stereotypes, self-evaluation, and self-identity has been recognized by many psychologists as an important human motive (Festinger, 1962; Greenwald et al., 2002; Heider, 1958). That is to say, cognitive dissonance involving conflicting self-evaluation, self-identity, stereotypes and attitudes can produce a feeling of discomfort leading to an alteration in one of the attitudes or beliefs. Based on the cognitive consistency principle, change in any one of the three sets of associations – group identity (e.g., self = science), self-evaluation (e.g., self = empathy) or stereotype (e.g., science ≠ empathy) would induce balancing change in at least one of the others.

Cvencek, Greenwald and Meltzoff (2012) proposed the Balanced Identity Theory (BIT) for consistency in implicit social cognition. In line with the aforementioned Social Identity Theory (SIT; Tajfel & Turner, 2001) for explicit social cognition, BIT also assumes a close relationship between group membership and self-concept. BIT predicts that people who identify strongly with a group should display more positive attitudes toward ingroups (Cvencek et al., 2012). However, the most substantial difference between SIT and BIT relies on how the SIT and BIT treat the role of self-evaluation in the tendency to favour one's own group relative to other groups.

SIT treats self-evaluation as a motivational force that leads people to use group identities to generate positive self-regard either by viewing their ingroups more positively or viewing outgroups more negatively (Rubin & Hewstone, 1998). According to SIT, people with low self-esteem in a specific trait or social state should be more motivated to distinguish between ingroups and outgroups and relate relatively stronger positive attributes to their ingroups in order to establish a positive social identity for themselves and hence promote their personal self-esteem (Rubin & Hewstone, 1998). Nevertheless, BIT takes *associations* as its conceptual building blocks and treats self-evaluations as an associative connection of self to positive attribute, and the balance-congruity principle calls for the link between the self-associated group and positive attribute to be strengthened by the link of self to positive attribute (Cvencek et al., 2012; Greenwald et al., 2002). In other words, the BIT emphasizes the transmit of the positive attribution of self-evaluation to self-identified groups. In contrast with SIT's expectation that group members with low self-esteem should display stronger positive views about ingroups, BIT predicts the reverse – that those who have high self-esteem in a specific trait should develop stronger positive views about that trait for their ingroups.

SIT was developed before the field widely recognised the distinction between implicit and explicit measures and has been tested mostly with self-report explicit measures (Rubin & Hewstone, 1998). Mixed evidence has been found for the SIT predictions. For instance, using the minimal group experiments (i.e., a method to investigate the minimal conditions required for favouring one's own group relative to other groups based on arbitrary and virtually meaningless distinctions between groups; Tajfel, Billig, Bundy, & Flament, 1971), Wagner, Lampen and Syllwasschy (1986) found that individuals with lower self-esteem showed relatively stronger intergroup discrimination, but some other researchers found the opposite pattern that individuals with higher self-esteem exhibited higher prejudice than did lower self-esteem groups (Crocker, Thompson, McGraw, & Ingerman, 1987). Nevertheless, tests of BIT have been carried out with both implicit and explicit measures and results of a meta-analysis of 18 studies of gender and racial stereotypes revealed that the relationships predicted by BIT are evident more strongly and consistently when assessed with implicit measures than when assessed with parallel explicit self-report measures (Cvencek et al., 2012).

In the present study, following the SIT account for explicit stereotypes, men majoring in science who have relatively lower self-evaluated empathy than their female counterparts are hypothesised to be more motivated to show weaker explicit stereotypes about empathy in scientists in order to promote their personal self-esteem in empathy from their social identity. In contrast, according to the BIT account for implicit stereotypes, women majoring in science (self = science) who have relatively higher self-evaluated empathy (self [women] = empathy) than their male counterparts are hypothesised to transmit this positive self-evaluation of empathy to their identified science academic group (science = empathy), leading to a comparatively weaker implicit bias about empathy in scientists than men majoring in science. The two hypotheses will be examined using both implicit and explicit measures of stereotypes of empathy in scientists in the present study.

To sum up, taken these findings and theories together, variations in stereotypes vary according to whether they are measured implicitly or explicitly. Therefore, it is essential to use both implicit and explicit measures to gain a fuller picture about stereotypes of empathy in scientists in the present study. Existing research on implicit and explicit stereotypes of scientists is scrutinized in the following section, setting up a stage to further discuss the limitations of existing studies and demonstrate the research gaps that the present study aimed to address.

3.2.2. Existing research on stereotypes of scientists

In this section, empirical studies utilising different kinds of measures to assess people's perceptions about scientists are critically reviewed. As mentioned earlier, it is of great importance to distinguish between implicitly-measured and explicitly-measured stereotypes, therefore concentrations are placed on examining the existing measurements used to assess stereotypes of scientists. Related findings about variations in stereotypes of scientists are also scrutinised.

3.2.2.1. Early research

Empirical studies about stereotypes of scientists can be traced back to the study conducted by Mead and Métraux (1957) asking 35,000 high school students to write an essay describing their image of a scientist. Analysis of these essays revealed that the typical high school student in this study perceived scientists as being an elder or middle-aged man wearing a white coat and glasses who conducted dangerous experiments alone in a laboratory. Later in the 1960s, Beardslee & O'Dowd (1961) generated long lists of observed stereotypes and categorised these stereotypes as either "positive" (e.g., intelligent, highly trained, devoted) or "negative" (e.g., brainy, dull, work alone) after conducting unstructured interviews with 1,200 college students. In the 1970s, the stereotypical image reported by Mead & Métraux (1957) persisted and this persistence was confirmed by several other empirical studies using questionnaires to survey public attitudes to science in America (Etzioni & Nunn, 1974; Ward, 1977).

In the 1980s, researchers started to develop questionnaires for stereotypes of scientists in particular. Krajkovich and Smith (1982) developed the Image of Science and Scientists Scale to measure students' perceptions of science and scientists. This scale includes both items about perceptions of scientists' personality traits (e.g., "when I think about a scientist, I think of a person who is intelligent") as well as items about attitudes toward science as a job (e.g., "a scientist's work is dangerous"). Participants are asked to rate these items from a 6-point Likert scale ranging from "strongly agree" to “strongly disagree”. Though it was proven to have strong reliability with a coefficient alpha of .86 (Krajkovich & Smith, 1982), researchers criticized that the scale has considerable semantic confusion among the terms “scientific attitude” “attitude toward science” and “image of science and scientists” (Koballa, Krajkovich, & Smith, 1983; Wyer, 2003).

3.2.2.2. Recent research on stereotypes of scientists using explicit measures

3.2.2.2.1. UK government’s public attitudes to science survey

To understand the UK public’s attitudes toward science, scientists and science policy, the UK government has funded a series of questionnaire surveys – known as the Public Attitudes to Science (PAS) series since 2000. The PAS surveys address a broad range of topics including public attitudes toward the current state of scientific literacy, current scientific issues and scientific engagement. According to the PAS 2014 report, the UK public’s overall perceptions of scientists and engineering were “overwhelmingly positive” with 90% of participants indicating that scientists “make valuable contribution to society” and 83% of participants agreeing that scientists “make life better for average person” (Ipsos MORI, 2014, p. 44). However, when asked to rate scientists and engineers on their personal traits, 44% said they are poor at communication and 50% considered them to be secretive (Ipsos MORI, 2014, p. 47). These findings suggest that the perception of scientists as less-social people may still be a deeply embedded stereotype in the UK society.

Although the PAS surveys have long been used in the UK as a source of authoritative knowledge about science-related attitudes, there are doubts about the robustness of its methods and claims. First of all, the sampling of the PAS 2014 study is problematic. This study consisted of a main survey of 1749 UK adults aged 16+ and a booster survey of 315 young people aging from 16 to 24 years. Clearly, the sample sizes for both surveys are limited and the young participants recruited for the booster survey are not representative of the general UK population. Secondly, there is a lack of procedural transparency about data processing in the report. The underreporting of test statistics such as the *p*-values makes the claims made in the PAS report less credible (Smith & Jensen, 2016). Higher quality research with rigorous methodology is still needed to investigate the objectives of the PAS survey including how people perceive scientists.

3.2.2.2.2. <u>ASPIRES project</u>

To understand how young people's aspirations in science develop in the UK, a five-year study called ASPIRES combining quantitative online surveys of over 19,000 students aging 10 to 14 years and longitudinal interviews with a selective subsample of 83 students and 65 of their parents were conducted from 2009 to 2013 (Archer, Osborne, et al., 2013). The ASPIRES project covers many topics including aspirations in science; attitudes towards school science; self-concept in science; images of scientists; participation in science-related activities outside of school; parental expectations; parental school involvement; parental attitudes towards science; and peer attitudes towards school and towards school science (Archer, Osborne, et al., 2013).

Regarding students' perceptions about scientists, the ASPIRES survey data showed that only 21% of Year 9 students agreed that scientists are "geeks" and even fewer (14%) believed that scientists are 'odd'. Only 25% agreed that scientists spend most of their time working by themselves. Such findings suggest that young people's explicit views of scientists have become less stereotypical in recent years. However, the interview data showed that science aspirations and attitudes are patterned by gender. Girls were found to describe careers in science as 'not girly' and those who define themselves as highly feminine were particularly unlikely to aspire to a scientific career (Archer, DeWitt, et al., 2013). Medicine was a particularly popular scientific aspiration among girls as it was concerned with 'helping people' (Archer, Osborne, et al., 2013). Girls who aspire to science by Year 9 were a minority among their peers and described themselves as being odd for studying science, while at the same time being 'girly' and enjoying socialising with others (Archer, Osborne, et al., 2013). Nevertheless, boys who aspire to science were more evenly divided between those who described themselves as 'cool/sporty' and those who did not. These findings suggest that girls who aspire to science careers may require considerable resilience to maintain their aspirations in science.

All in all, in contrast to the above-mentioned PAS survey results, young adolescents from the ASPIRES project reported less stereotypical views about the isolated and geeky image of scientists. It is possible that with the increasing awareness of the harms of stereotyping, adolescents may be reluctant to report their bias due to social desirability concerns. However, ASPIRES found a widespread explicit stereotypical view among young people in the UK that science careers are 'not girly'. It remains unknown which specific feminine trait is considered unrelated to science careers.

3.2.2.2.3. ROSE project

The Relevance of Science Education (ROSE) project is a cross-cultural research project investigating young people's attitudes toward science and technology. It comprised surveys among 15-year-old students from 40 countries from all continents (Sjøberg & Schreiner, 2012). The ROSE project used a long questionnaire containing 108 items addressing students' science-related out-of-school experiences, interests in learning science, views and attitudes toward school science, views of scientists and career aspirations.

According to the ROSE report, young people from all the countries agreed that science is beneficial to society. However, students from more developed countries (e.g., UK, USA, Japan) showed weaker interest in learning science than those from developing country (e.g., Nigeria, India, Ghana). In developed countries, young people showed very low interest in “becoming a scientist”, in particular very few girls would like to get a job in STEM fields. Nevertheless, young people from developing countries believed that science can “make their lives healthier, easier and more comfortable" and rated scientists very high on an empathetic notion “helping the poor”. However, children from developed countries, especially girls, were very unlikely to associate science and scientists with such positive notions. In most developed countries, girls reported more interest in getting a job that can help other people than boys, and less than 20% of girls think of scientists as "helping the poor" (Sjøberg & Schreiner, 2012).

3.2.2.2.4. Stereotypes of Scientists survey

Instead of surveying a wide range of topics relating attitudes toward science, scientists and scientific experiences as the above-mentioned project, Wyer, Schneider, Nassar-McMillan and Oliver-Hoyo (2010) developed the Stereotypes of Scientists scale concentrating on assessing people’s perceptions about scientists’ professional competence as well as their interpersonal skills. In this scale, participants are required to rate items such as “scientists are intelligent” or “scientists make friends with people from other departments”.

This scale has so far been used only in their own study with 1106 college students in the US (Wyer et al., 2010). They found that American college students nowadays explicitly reported positive attitudes toward both scientists’ academic competence and interpersonal relations. However, given that self-report questionnaires are prone to the disadvantage of allowing participants to censor and control answers, this scale may not be able to capture participants’ unconscious attitudes toward scientists.

To sum up, recent research on attitudes toward scientists using explicit self-report questionnaires found mixed results about stereotypes of empathy in scientists. The PAS 2014 survey suggested that around half of the adults in the UK believed scientists are not good at communication. Nevertheless, very few adolescents from the ASPIRES project reported stereotypical views about scientists' social skills. However, adolescents recruited in the ROSE project still reported little faith in scientists to help other people. Furthermore, it is worth noting that questionnaires used in these projects often cover a wide range of topics and only a few items tapped into perceptions about scientists, not to mention stereotypes about empathy in scientists in particular. Research focusing on the image of scientists has mostly been conducted in the US. Last but not least, self-report measures are limited to introspective responses which may be affected by social desirability concerns. Therefore, the present study addressed these research gaps by investigating perceptions of empathy in scientists in the UK using both implicit and explicit measures.

3.2.2.3. Recent research on stereotypes of scientists using implicit measures

As presented before, in the late 1990s, researchers began to differentiate between implicit and explicit stereotypes, thus the Draw-a-Scientist Test, as well as the Implicit Association Test, have been used in recent decades as implicit measures to capture people's unconscious beliefs about scientists.

3.2.2.3.1. <u>Draw-a-scientist studies</u>

The Draw-a-Scientist Test (DAST), in which no introspective verbal responses are required, only a drawing of a person is made and evaluated with reference to a selection of stereotypes, was developed by Chambers (1983) and used mostly to measure children's perceptions of scientists. This instrument can be regarded as an implicit measurement given that test takers are unlikely to be aware of the purpose of the study. The DAST can be advantageous in developmental research because children may learn stereotypes before reporting them explicitly (Galdi, Cadinu, & Tomasetto, 2014).

As presented earlier in the context of study chapter, the stereotypical image of a scientist that people often draw is an older or middle-aged man who wears a white coat and glasses doing experiments alone in a laboratory (Finson, 2002). Evidence from the DAST studies revealed that such stereotypical image of scientists can be formed very early in life, by the time children reach the second grade in elementary school (Losh, Wilke, & Pop, 2008; Newton & Newton, 1998). Moreover, Finson's (2002) meta-analysis of studies utilizing DAST reveals a remarkable consistency in the image of scientists drawn by people across nations, across age groups, across education levels. The latest meta-analysis of DAST studies conducted by Miller, Nolla, Eagly and Uttal (2018) found that children depicted female scientists more often in later decades, but they still associate science with men as they grow older.

Despite being successfully utilised in many studies for a long time, DAST has been criticized for its limited application to older age-groups and being prone to the experimenter bias in terms of coding (Finson, 2002). Moreover, the DAST also entails a prerequisite of drawing skills and boys' drawings have been found more abstract than girls', which often provided little information for researchers to extract important traits such as the drawing's gender (Thomas et al., 2006).

3.2.2.3.2. Gender-Science IAT studies

Another popular implicit measurement that has been used to gather unconscious stereotypes about scientists is the Gender-Science Implicit Association Test (GS-IAT; Nosek, Banaji, & Greenwald, 2002b). The GS-IAT captures people's automatic associations between men and science by assessing different reaction times to stimuli that are either paired in a stereotypical way (i.e., men and science) or not (i.e., women and science). Research using the GS-IAT has shown that people's implicit association between men and science emerges early in childhood (Cvencek et al., 2011), is developmentally stable (Nosek et al., 2002b), ubiquitous across different cultures (Nosek et al., 2009), but also substantially variable across different individuals (Smyth

& Nosek, 2015).

Existing studies have found that the implicit male-science stereotype predicts judgment and behaviour that contribute to the gender gap in science-related activities and occupations. These studies have found significant relationships of implicit male-science stereotypes with participants' math engagement (Nosek & Smyth, 2011), performance and achievement, intentions to choose scientific major subjects (Lane et al., 2012; Nosek et al., 2009; Zitelny, Shalom, & Bar-Anan, 2017), and careers (Cundiff, Vescio, Loken, & Lo, 2013). These relations were usually moderated by gender. Among women, stronger implicit male-science stereotypes were usually correlated with worse math performance and achievement, as well as weaker identification with math and science.

Nevertheless, among men, the implicit male-science stereotypes sometimes had no correlations with relevant outcomes, and on other studies, stronger implicit male-science stereotypes could, in turn, be correlated with better performance, achievements, and the stronger identification with math and science (Zitelny et al., 2017). In their meta-analysis examining seventeen studies using both implicit and explicit measures for gender-science stereotypes, Zitelny et al., (2017) reported that fifteen of those studies found that the implicit male-science stereotype had a stronger relationship with an outcome variable than the explicit stereotype. It was speculated that implicit stereotypes may "shape choices by subtly constraining preferences without the individual's awareness or conscious exertion of choice" (Nosek et al., 2002b, p. 50) and "sincere and conscious beliefs that men and women are equally well suited for STEM fields do not preclude internalization of the beliefs at a less conscious level" (Lane et al., 2012, p. 222).

Moreover, Smyth and Nosek (2015) conducted an online survey with 176,000 college-educated participants examining relations between gender ratios and male-science stereotyping. They argued that gender-science stereotypes should change as conditions in local environments change, including gender ratios. Higher female proportion of a

STEM major should be correlated with weaker science-male stereotyping. Evidence from their study only supported this hypothesis for the explicit stereotype, but not the implicit stereotype. Participants from high-female disciplines including biological and health sciences showed weaker explicit male-science stereotype than those from low-female disciplines such as computer sciences, physics and engineering. However, implicit male-science stereotype did not correspond with disciplines' gender ratios, but was correlated with scientific intensity, positively for men and negatively for women. That is to say, women who majored in subjects perceived as more science-intensive (e.g., physics) showed weaker implicit male-science stereotype than did men in the same disciplines. Furthermore, particularly among women, those who identified with more science-intensive subjects (e.g., physics) showed weaker male-science implicit stereotypes than women majoring in less science-intensive subjects (e.g., biology).
Existing research using the Gender-Science IAT has found evidence for relations between implicit gender-science stereotypes and scientific academic achievement and career aspirations. On the basis of these findings, we propose to examine whether stereotypes of empathy in scientists also correlate with participants' science career aspirations. Moreover, given that variations in implicit and explicit male-science stereotypes have shown different relationships with gender ratios of the discipline and scientific intensities, the present study will also discuss the individual differences in the stereotypes of empathy in scientists differently by whether they are measured implicitly or explicitly.

To sum up, previous studies looking at stereotypes of scientists have three main limitations. Firstly, most self-report questionnaires used in projects investigating attitudes toward science in the UK covers a broad range of topics (e.g., attitudes toward school science) and only a few items tapped into perceptions about scientists in particular. Secondly, research focusing on the image of scientists has mostly been conducted among children and adolescents in the US using DAST. The DAST could only provide a general picture of a scientist and was unable to systematically examine a specific trait in the scientist. It remains unknown how young adults in the UK perceive scientists nowadays and a new instrument is in need to measure stereotypes of empathy in scientists in particular. Last but not least, most existing studies only used the IAT to measure implicit gender stereotype of scientists. Though we have found a robust link between men and science, further investigation is required to figure out which specific stereotypical trait of scientists prevents people from associating science with women. As discussed earlier in Section 3.1, women have been found to rate themselves high in empathy but scientists usually rate themselves low in empathy (Baron-Cohen, 2002; Manson & Winterbottom, 2012). This contradictory evaluations of empathy among women and among scientists may serve as a gatekeeper for women to pursue science. Research could be fruitful to examine whether stereotypes of empathy in scientists are related to gender, major subject selection as well as career aspirations in science. Given that there is a lack of reliable measurement to capture people's perceptions of empathy in scientists, the first goal of the present study is thereby to develop a measurement that taps into stereotypes of empathy in scientists using the Implicit Association Test (IAT) paradigm. The following sections present how IATs differ from self-report questionnaires as well as existing studies about psychometric properties of the IAT, providing background information for further development and testing of the new IAT for stereotypes of empathy in scientists in the present study.

3.3. Implicit association test (IAT)

Unlike self-report questionnaires that ask participants directly how they think or feel about the given research topic, IAT utilizes alternative response parameters, that is the reaction time, to indirectly infer an individual's underlying beliefs. The mechanism of the IAT is to compare the strengths of associations among concepts without introspection on the part of the participant. The test assesses the strength of associations between target categories (e.g., *Science* versus *Humanities*) and attribute categories (e.g., *women* versus *men*), which are arranged on two separate dimensions. For example, when using the IAT to measure gender-science stereotyping, participants are asked to categorise words representing *Humanities* versus *Sciences* (e.g., physics, chemistry, history, etc.,), and pictures representing *women* and *men*, as rapidly as possible, using designated response keys. On certain trials, the categories *Science* and *women* would be paired on the same response key (while *Humanities* and *men* on the other), whereas on other trials *Science* and *men* would be paired together (and *Humanities* with *women*).

Categorisation response times would then be compared across these two conditions to infer the nature of gender stereotypes of scientists held by participants. For example, if one responded faster to respond when items representing *Science* paired with images of *women*, it would be assumed that one held a counter-stereotype implicit belief about the gender of scientists. In contrast, if one was faster to respond when items representing *Science* paired with images of *men*, it would be assumed that one held stereotype-congruent implicit views about the gender of scientists. The rationale behind the IAT procedure is that individuals should be faster to respond when strongly associated constructs are paired on the same response key, than when they are separated across response keys. Therefore, The IAT is argued to reflect automatic cognitive associations held by respondent and may have the potential to overcome some of the limitations of self-report measures (Greenwald, McGhee, & Schwartz, 1998)

3.3.1. What can IAT tell us that differ from self-report questionnaires?

3.3.1.1. Unconscious versus conscious representations

A major difference between the IAT and self-report measures is that IATs avoid introspection for the assessment of mental representations. Thus, IATs have the potential to tap unconscious representations that are inaccessible to introspection (Asendorpf et al., 2002; Bosson, Swann Jr., & Pennebaker, 2000; Greenwald & Banaji, 1995). One empirical finding that is often interpreted as evidence supporting the unconscious nature of the implicit representations measured by IATs is that IAT scores often diverge from those obtained on corresponding self-report questionnaires (Lane, Banaji, Nosek, & Greenwald, 2007; Nosek, Greenwald, & Banaji, 2007).

However, it is worth mentioning that low correlations between the two kinds of measurements can also be the result of many other factors other than the lack of introspective access (Gawronski, LeBel, & Peters, 2007). For example, some researchers argue that the low correlations between IATs and self-reports may be partly due to the lack of *conceptual correspondence* (i.e., the extent to which the attitudes are measured at the same level of abstractness and with the same degree of specificity; Ajzen & Fishbein, 1977) as well as *structural fit* (i.e., the degree of methodological similarity between different tests, Payne, Burkley, & Stokes, 2008) between implicit and explicit measures (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005).

That is to say, in most explicit self-reports, participants are required to make evaluations on a series of propositional statements. In contrast, the IATs avoid all the propositions involved in the self-reports, and instead only assess relative attitudes toward two different target concepts based on the comparisons between reaction time to register different responses. For example, many standard self-report scales for racial prejudice (e.g., Modern Racism Scale, McConahay, 1986) assess participants' political opinions on issues such as affirmative action or the discrimination of ethnic minority members. In contrast, racism IATs typically assess participants' responses to members of ethnic minority groups (e.g., faces of Black and White individuals). Thus, even though general evaluations of ethnic minority members may be systematically related to people's political opinions (Gawronski, Peters, Brochu, & Strack, 2008), the low correlations between these self-report racial prejudice scales and the IAT measures are very likely resulted from their conceptual and structural distinctnesses, not necessarily because one is tapping unconscious representations (Gawronski et al., 2007; Hofmann et al., 2005). Therefore, cautions are stressed on the content and structure of the applied self-report measures in the present study when interpreting their correlations with the IAT.

3.3.1.2. Uncontrolled versus controlled representations

Another primary difference between the IAT and self-report explicit measures resides in their differing susceptibility to self-presentation and social desirability. Specifically, it is assumed that performance on IATs is uncontrollable by participants as they are either unaware that their perceptions are being assessed or do not have enough cognitive capacity to alter their immediate spontaneous responses (Asendorpf et al., 2002; Fazio et al., 1995). Therefore, IAT scores are resistant - or at least less susceptible – to such motivational distortions. Building upon this assumption, a group of researchers claimed that implicit measures may provide a "bona fide pipeline" to people's true attitudes and beliefs (Dunham, Baron, & Banaji, 2006; Rutland, Cameron, Milne, & McGeorge, 2005; Teachman, Gapinski, Brownell, Rawlins, & Jeyaram, 2003).

Although implicit measures such as the IAT tap into mental associations that an individual may not want to report, or may not be able to report, a number of articles have criticized the misuse of the implicit measures as lie-detectors or revealing more "true" or "real" mental representations than self-reports (Arkes & Tetlock, 2004; Gawronski et al., 2007; Karpinski & Hilton, 2001). It has been argued that the IAT and self-report results can differ for multiple reasons: one can be unaware of the implicit bias and report a unique explicit responses; or one is aware of the implicit bias, but genuinely reject them as not consistent with his or her beliefs so chooses to report an alternative response; or one is aware of the implicit bias but chooses to report an alternative response due to self-presentation concerns and social desirability. Only the third possibility would fit the assumption that implicit measures detect a deliberate concealing of endorsed explicit attitudes and beliefs (Nosek, Greenwald, et al., 2007). Moreover, evidence has shown that participants were often surprised by their IAT results when revealing undesirable implicit bias and reported feelings of guilt especially for implicit racial bias (Ashburn-Nardo et al., 2001; Monteith, Ashburn-Nardo, Voils, & Czopp, 2002).

Therefore, though IATs are resistant to deliberate attempts to control responses, it is inappropriate to classify IAT and self-reports as having distinguished degrees of access to reality or truth (Lane et al., 2007). Having implicit bias does not make an individual a racist, sexist or any other accusation. When interpreting findings of the present study, it is important to bear in mind not to regard explicit attitudes as "fake" or "unreal" but treat the implicit bias as the "real" or "true" attitude.

3.3.1.3. Old versus new representations

Another common assumption in research using implicit measures is that the IATs assess highly stable old representations that have their roots in long-term socialization experiences, whereas standard self-report measures tap newly acquired attitudes, at least as long as people are motivated and able to retrieve their new attitudes from memory (Conner & Barrett, 2005; Dovidio, Kawakami, & Gaertner, 2002; Petty, Tormala, Briñol, & Jarvis, 2006; Rydell & McConnell, 2006; Sinclair, Dunn, & Lowery, 2005). Some researchers have proposed that recently acquired attitudes often do not overwrite old attitudes, but instead coexist with old, ostensibly stable, implicit ones (Petty et al., 2006; Wilson et al., 2000). The proposed matching of measurement procedures with "old" versus "new" attitudes implies that implicitly assessed evaluations should exhibit a higher level of robustness against attempts to change attitudes than do self-reported evaluations (Gawronski et al., 2007).

However, it is important to note that existing empirical studies revealed mixed evidence for this assumption. Some researchers found that experimental attempts to change attitudes affect only explicitly self-reported but not implicitly assessed attitudes (Gawronski & Strack, 2004). But several other studies found corresponding changes in both explicitly and implicitly measured attitudes (Gawronski, Walther, & Blank, 2005; Olson & Fazio, 2001; Richeson & Nussbaum, 2004). There are also some studies revealing that conditioning manipulation affected implicit but not explicit attitudes (Dasgupta & Greenwald, 2001; Olson & Fazio, 2006).

Moreover, the IATs have also demonstrated a relatively high level of context sensitivity (e.g., Blair, 2002; Gawronski & Bodenhausen, 2006). For example, Wittenbrink, Judd, and Park (2001) found that implicitly measured attitudes toward Black people differ as a function of the background context in which these individuals are presented (e.g., at a barbeque or against a graffiti wall). Taken together, these findings suggest that the proposed equation of implicit measures versus self-report measures with old versus new representations (respectively) is empirically unfounded. When interpreting the findings of the present study, it is important to bear in mind not to regard implicit attitudes as more stable and long-lasting than the explicit ones.

3.3.1.4. Activation of associations versus validation of judgments

Following a more recent conceptualization, implicit measures are regarded as providing a proxy for *activation* of associations in memory (i.e., associative processes), whereas self-reports reflect the propositional outcome of *validation* processes (i.e., propositional process; (Gawronski & Bodenhausen, 2014). Both processes reflect different aspects of mental representations, and the activation of associations can occur independently of whether an individual considers these associations as accurate or not (Gawronski et al., 2007). That is to say, the validation process is concerned with assessing the truth or falsity of the activated information, but the rejection of a given proposition as invalid does not necessarily lead to a deactivation of associations the proposition is based on. In this case, rejections of propositions can affect judgments assessed with self-report measures but not necessarily the activation of associations assessed with implicit measures (Gawronski & Bodenhausen, 2014; Gawronski et al., 2007). In other words, having automatically activated stereotypical associations of concepts does not mean that one will acknowledge those activations when making reflective judgments.

To sum up, when interpreting the findings of the present study, it is important to bear in mind that both the implicitly and explicitly measured stereotypes are "real" to some extent and reflect different kinds of cognitive processes. Both the automatic *associative* and validated *propositional* attitudes are real attitudes and they can coexist at the same time. The IATs are more resistant to deliberate attempts to control responses due to social desirability bias than the self-report measures, but they are still sensitive to manipulations of contextual factors and do not reveal higher-level stability than self-report measures. Therefore, it is of great importance to carefully control construct-unrelated factors which may affect the reliability and validity of the measurement when developing an IAT. The following sections review studies of the psychometric properties of IAT measures.

3.3.2. Psychometric properties of the IAT

Because the IAT represents a procedural paradigm for measuring implicit cognition rather than a single measure of a specific construct, there is no single incarnation of the IAT to be validated. Given that the IAT can be adapted to measure different constructs (e.g., racial attitudes, gender attitudes, food preference, etc.,), two IATs may have little in common other than the basic structure of the task. Although the IAT paradigm produces many reliable and valid tasks, this does not mean that any single IAT is necessarily a good measure of the target construct. Given that the present study aims to develop a new IAT to assess individual differences in stereotypes of empathy in scientists, it is especially important to evaluate whether the newly developed IAT can meet relevant psychometric criteria. The following sections review the most common psychometric properties for the IAT in general.

3.3.2.1. Reliability

The concept of reliability is very much tied up with the repeatability or reproducibility of any assessment. For any measure to be useful, it needs to have consistency so that it produces more or less the same result for a person each time it is used (Coaley, 2010). High reliability means that a measure gives similar results on different occasions.

3.3.2.1.1. Internal consistency

As a measurement based on response latencies measured in milliseconds, error variance can be easily introduced into studies using IATs - a sneeze, a car horn, or even an eyeblink can add unwanted variance in response latency (Lane et al., 2007). Indeed, evidence has shown that the internal consistency of measures based on reaction time is generally lower than that of those based on self-reports (Buchner & Wippich, 2000). Even so, IATs have gained exponential popularity partly due to their superior reliability over other latency-based implicit measures such as the Go/No-Go Association Task (averaged split-half reliability r = .20; Nosek & Banaji, 2001) or the priming method (e.g., split-half reliability r = .06, Bosson, Swann & Pennebaker, 2000). The split-half method used to test internal consistency is to create two parallel tests from the items within one test, then to compute a composite score for each subtest and correlate the two composite scores (Furr & Bacharach, 2008). According to the rule of thumb, an estimate of 0.6 to 0.7 indicates acceptable reliability, and 0.8 or higher indicates good reliability. The split-half reliability for existing IAT measures tends to range from .70 to .90 (Schmukle & Egloff, 2005).

The split-half approach is based on the perspective that two halves within a test represent parallel subtests, and the reliability of the complete test is based on the associations between the two subtests. From the "item-level" perspective, Chronbach's alpha takes the logic of internal consistency a step farther by conceiving of each item as a subtest. Consequently, the associations among the items can be used to estimate the reliability of the complete test (Furr & Bacharach, 2008). In one study examining the internal consistency of a number of implicit measures, the IAT showed satisfactory internal consistency (Chronbach's alpha = .78), which is relatively rare for other implicit measures such as the evaluative priming task (Cunningham, Preacher, & Banaji, 2001). A meta-analysis of 50 studies using IATs has also reported an acceptable averaged internal consistency of the IAT with Chronbach's alpha = .79 (Hofmann et al., 2005). Based on these findings, the newly-developed IAT in the present study is expected to show satisfactory internal consistency around the rule of thumb of .70.

3.3.2.1.2. Test-retest reliability

On the other hand, the test-retest reliability (i.e., the consistency of measurement at different time points; Furr & Bacharach, 2008) of the IAT has been found less satisfactory, ranging from .25 to .69 with a mean estimate of about .50 (Lane et al., 2007). That is to say, when we use the same measure in the same population at two different time points, the measure is expected to produce similar results which should demonstrate strong positive correlations with each other. For example, a measurement producing the same data output at every time would, therefore, show a perfect test-retest reliability of r = 1. The satisfactory test-retest value for the self-report questionnaire is usually above .70.

Though the IAT shows somewhat low test-retest reliability as compared to self-report measures, it is important to note that other implicit measures also demonstrate relatively weak test-retest reliability compared to explicit measures and it is not uncommon for implicit measures to show test-retest reliability below .50 (Bosson et al., 2000). The IAT even showed superior averaged test-retest reliability (.69) over other implicit measures, which ranged from -.05 (e.g., Stroop task) to .63 (e.g., Initials birthday preference task) and averaged .03 (Bosson et al., 2000). Given that the IAT usually showed low test-retest reliability as well as the limited time scope of a PhD project, the test-retest reliability of the newly developed IAT is not examined in the present study.

3.3.2.2. Validity

Construct validity concerns the evidence which shows that the test really is a measure of what it claims to measure (Coaley, 2010). In statistical terms, construct validity represents the extent to which a measure's variance is linked with the variance of its underlying construct (Barrett, Phillips, & Alexander, 1981). There is no single correlation for this and validation requires review of a range of correlations and whether they match what would be expected. For example, the simplest method is to examine the correlation of scores with other accepted measures of the same thing (Coaley, 2010). Therefore, construct validity depends on all the evidence gathered to show that a test does relate to its construct. The following sections review a series of validity criteria as well as construct-unrelated variance that may violate the validity of the IAT.

3.3.2.2.1. Known-groups validity

A valid measurement needs to be able to reliably differentiate between members of different groups, based on prior knowledge or predictions about them. It is important for an IAT to successfully discriminate between groups in order to be regarded as a measure of personal attitudes rather than shared stereotypes embedded in the culture that one lives in (Nosek, Greenwald, et al., 2007). For example, consistent with the prevailing gender role expectations, women indeed were found to implicitly associate self more with arts as compared to math than men did (Nosek et al., 2002b). Such finding can be interpreted as evidence that the IAT is able to predict known group differences.

Moreover, existing evidence has shown that the IAT is sensitive to more subtle differences in the societal evaluation of different groups despite the ambiguity of ingroup preference. Using the IAT method, African Americans showed reduced ingroup preference as compared to European Americans (Ashburn-Nardo et al., 2003), and over-weight and poor people even showed an outgroup preference instead (Rudman, Feinberg, & Fairchild, 2002). Successful discrimination between group members even extends to groups that are defined by behaviours rather than demographics. For example, smokers showed more positive implicit attitudes toward and stronger identity with smoking than non-smokers (Houwer, Custers, & Clercq, 2006). In the present study, it is hypothesized that the IAT can differentiate between science majors and humanities majors by showing different implicit attitudes toward empathy in scientists.

3.3.2.2.2. Relationship with explicit measures

The convergent-discriminant validity is one type of subordinate validity which is based on the idea that a measure needs to correlate with others of the same thing but not with those measuring other constructs (Campbell & Fiske, 1959). One of the important criteria to evaluate the convergent-discriminant validity of IAT is to examine its correlations with explicit measures, in other words, self-report measures (Greenwald, Nosek, & Banaji, 2003; Nosek, Greenwald, et al., 2007). The relationship between implicit and explicit attitudes has received a great deal of attention that has produced mixed evidence for the original proposed question: "Do implicit and explicit attitudes relate to one another?"

Some of the initial research efforts with the IAT emphasized the distinctiveness of the implicit and explicit cognitions in finding weak to weak relations between IAT and self-report measures (Greenwald et al., 1998). However, with the accumulation of research on the relationship between implicit and explicit measures, recent studies have produced mixed evidence regarding implicit-explicit correlations. As reported by Lane et al. (2007), across 17 IATs that were available at public websites, correlations between implicit and explicit measures range from r = .13 to r = .75 (median r = .22). Laboratory studies have also shown similar variability, with some studies revealing slight or moderate (but generally positive) correlations between IAT and self-reports of the same construct (Bosson et al., 2000; Egloff & Schmukle, 2002), and other studies showing strong and robust correlations between IAT and self-reports (e.g., Cunningham et al., 2001; Jellison, McConnell, & Gabriel, 2004; McConnell & Leibold, 2001). A meta-analysis of IAT studies found that across 126 studies, implicit-explicit correspondence ranged from r = -.25 to r = .60, with an average implicit-explicit correlation ofcit .19 (Hofmann et al., 2005).

Even when IATs and explicit measures do correlate, evidence shows that implicit and explicit attitudes are still distinct constructs (Wilson et al., 2000). Using structural equation modelling, across 57 different pairs of attitude objects, Nosek and Smyth (2007) found that implicit and explicit attitudes were better fit by a model in which they loaded onto two separate factors, rather than a single, latent factor even when implicit and explicit attitudes were highly correlated with one another.

Given that the extent to which implicit and explicit attitudes are correlated varied widely across studies, the more appropriate question for future research to answer should be "Under what conditions, and for what kind of people, are implicit and explicit measures related?" (Olson & Fazio, 2004). The hunt for the convergent-discriminant validity of the IAT may also be more successful when large samples and advanced statistical techniques, such as meta-analysis or latent variable modeling, are used. In the present study, the relationship between implicit and explicit measures of the stereotypes of empathy in scientists are still examined. Following the idea that implicit and explicit stereotypes are distinct constructs, we still hypothesize the implicitly measured stereotypes of empathy in scientists to show weak correlation with explicitly measured self-report stereotypes in the present study. However, it is important to bear in mind that the simple correlations between one IAT and self-report measures should not be interpreted as robust evidence for the convergent-discriminant validity of the IAT.

3.3.2.2.3. Predictive validity

Given that a long-standing interest of psychologists is to understand how attitudes predict behaviour, it is not surprising much research effort has been put on examining whether attitudes captured by IATs are related to meaningful behaviours. The ability to successfully predict behaviours is also an important aspect of psychometric properties of a measurement. As presented before, researchers have drawn three types of theoretical models of the implicit and explicit stereotypes (i.e., double-dissociation, addictive and interactive; see Section 3.2.1.3) and each model entails a distinct hypothesis about the predictive validity of the IAT.

According to the double-dissociation model, IATs and self-report measures should predict spontaneous and controlled behaviours respectively. Indeed, there is evidence showing that IATs are capable of successfully predicting less controlled behaviours when participants are under a high cognitive load (Hofmann, Friese, & Strack, 2009) or under the influence of alcohol (Hofmann & Friese, 2008). Furthermore, Greenwald, Poehlman, Uhlmann and Banaji (2009) meta-analysed 184 independent samples and found that both implicit and explicit stereotypes were related to a range of behaviours but implicit bias measured by IATs were superior to explicit bias measured by self-reports in predicting physiological responses and non-verbal discriminant behaviours, whereas explicit measures were superior to IATs in predicting more deliberate behaviours such as political candidate choices and brand preferences.

As suggested by the additive model, the IAT and the explicit measures should be seen as different measures of the same attitudes. From this point of view, attitudes can be compared to icebergs, with explicit attitudes residing above the surface of conscious control and implicit attitudes residing below it (Karpinski & Hilton, 2001). If we follow this assumption of a single cognitive system but two different measures, both implicit and explicit measures of the same attitudes should provide a distinctive prediction of behaviours. However, no empirical evidence to date has been found supporting this hypothesis (Perugini, 2005).

According to the interactive model, implicit and explicit processes should work together in a multiplicative way in order to influence behaviours, possibly when individuals feel ambivalent towards something. For example, a study by Frost, Ko, and James (2007) revealed that individuals who were most likely to be passive aggressive when interacting with others showed high implicit aggression when completing a conditional reasoning task but scored low in self-report aggression questionnaire. Similarly, Jordan and his colleagues (2003) found that individuals displayed the highest level of narcissistic behaviours scored low in implicit self-esteem measured by an IAT but high in explicit self-esteem measured by a self-report questionnaire.

To sum up, existing evidence has shown that IATs can predict a range of meaningful behaviours. It is possible to articulate three predictive models that reflect the relations between implicit and explicit attitudes and behaviours. Empirical studies have found evidence for both the double-dissociation model and interactive model, but not the additive model. The present study will also examine the predictive power of the newly developed IAT for stereotypes of empathy in scientists by looking at its relationship with career aspirations in science.

3.3.2.2.4. Construct-unrelated variance

Several studies have revealed a number of construct-unrelated variables that may have influence on the IAT validity include: order of the compatible and incompatible tasks; cognitive fluency; and prior experience with the IAT.

Firstly, the most commonly observed extraneous factor is the order of the compatible and incompatible task. In the IAT, the compatible task is the task when participants are required to pair items in a stereotypical way, whereas the incompatible task refers to the task when participants are required to pair items in a counter-stereotypical way. Regardless of the content of the tasks, the performance of the preceding pairing task tends to interfere with the performance of the subsequent pairing task. IAT effects are found slightly biased toward indicating that the associations drawn upon in the first-performed task are stronger than those drawn upon the later-performed task (Back, Schmukle, & Egloff, 2005; Klauer, 2005). This extraneous effect will be controlled by counterbalancing the task order as well as adding more trials for practice in the present study (See later Section 4.3.4).

The second extraneous influence is the individual difference in average response latency, or so-called cognitive fluency (i.e., the ease with which information is processed; Mierke & Klauer, 2003). Participants who react generally more slowly tend to show larger IAT effects (representing stronger implicit bias) than those who react more quickly (McFarland & Crouch, 2002). This extraneous effect can be reduced by applying an advanced scoring algorithm (Greenwald et al., 2003; See later Section 5.2.5).

Moreover, existing evidence also suggests that effect magnitudes with the IAT tend to decline for participants who have prior experience taking an IAT (Greenwald & Nosek, 2001). The advanced scoring algorithm has also been proposed as an effective way to reduce the influence of this factor (Greenwald et al., 2003). The present study will apply this algorithm and test its ability to control the extraneous order variance (see later Section 5.3.6) and prior experience variance (see later Section 5.3.7).

3.4. Research questions and overview of the present study

The Literature Review chapter has summarised 1) existing research on relations between gender, empathy and science, 2) theories about implicit and explicit stereotyping as well as empirical research on stereotypes of scientistsand, 3) literature regarding the Implicit Association Test and its psychometric properties. It has been found that existing research on stereotypes of scientists has been mostly focused upon the gender stereotype of scientists, stereotypes of empathy in scientists are underinvestigated. Considering the lack of measurement to assess stereotypes of empathy in scientists, the first goal of the present study is to apply the IAT paradigm to develop a new IAT to capture implicit perceptions about empathy in scientists.

Using the newly developed IAT as well as self-report questionnaires, the present study aims to investigate three main research questions as follows:

1. What are the implicit and explicit stereotypes of empathy in scientists among UK university students nowadays?

2. What are the individual differences in the implicit and explicit stereotypes of empathy in scientists by gender and major subject?

3. How do the implicit and explicit stereotypes of empathy relate to students' career aspirations in science?

An overview of the present PhD project is summarised in Table 3.1.

Table 3.1 Research stages of the present PhD project

Stages	Research aims	Participants
Stage one	Designing and developing an IAT for stereotypes of empathy in scientists - Select appropriate categories and stimulus items for the Science-Empathy IAT (SE-IAT) - Design the procedures of the SE-IAT - Create a website to implement the SE-IAT online	32 participants for item selection and a focus group of 6 participants to test the website
Stage two	Testing and modifying the newly-developed IAT - Examine the validity and reliability of the SE-IAT - Modify the SE-IAT to the Single-category SE-IAT (SSE-IAT) accordingly	485 participants from four Russell Group universities
Stage three	Applying the SSE-IAT and self-report questionnaires to investigate the implicit and explicit stereotypes of empathy in scientists - Apply the SSE-IAT and self-report questionnaires to explore contemporary implicit and explicit stereotypes of empathy in scientists - Examine individual differences in the implicit and explicit stereotypes by gender and major subject - Examine the interplay of the implicit and explicit stereotypes in science career aspirations	1448 participants from eight Russell Group universities

Chapter 4 Stage One: designing and developing an online Implicit Association Test to measure stereotypes of empathy in scientists

4.1. Introduction

In Stage one, the two main research goals are 1) to systematically apply the IAT paradigm to develop a new Science-Empathy IAT (SE-IAT) to assess the implicit bias toward empathy in scientists and 2) to create a website to administer the new SE-IAT online. To date, very few studies have reported the specific steps of designing and developing an IAT. However, many problems may occur in the development process leading to additional construct-unrelated variance to the IAT effects. Therefore, it is considered of great importance to inform readers of how the SE-IAT was built and the specific methods used during the process to control for the potential confounds. In this chapter, three major steps to design and computerize the SE-IAT are presented in detail. In Step I, appropriate categories and stimulus items were firstly selected according to ratings of a focus group. In Step II, elements of the SE-IAT procedure, such as the order of the blocks and the number of trials, were designed in a mindful way to avoid the potential confounds as much as possible. In Step III, by comparing the pros and cons of different IAT software packages, the SE-IAT website was finally implemented online using the Project Implicit virtual lab service. The newly developed website was piloted with another focus group to test its face validity and updates were made accordingly.

4.2. Step I: Selecting Appropriate Categories and Items

4.2.1. Category label selection

The critical materials of an IAT are four categories defined by category labels (e.g., *Flower* and *Insect* as target categories; *Pleasant* and *Unpleasant* as attribute categories) and the stimulus items that serve as exemplars for those categories (e.g., pictures of flowers and insects as a stimulus items for target categories; positive and negative adjectives as a stimulus item for attribute categories). Two initial choices in developing an IAT arise in determining how to represent the chosen categories. Both chosen category labels, and the specific representative items, determine the construal of the concept (Lane et al., 2007). Existing evidence has shown that category membership, rather than the valence of individual exemplars, is most important in determining IAT effects (De Houwer, 2001; Mitchell, Nosek, & Banaji, 2003). For example, with the same stimulus items, when categorization is based on occupation, participants were found to prefer (well-liked) Black athletes to (disliked) White politicians, but when categorization is based on race, participants preferred White politicians to Black athletes instead (Mitchell et al., 2003). Such findings imply that the nature and construal of the categories play a large role in determining IAT effects and researchers must be precise in defining the constructs of interest and should carefully choose the appropriate category labels to represent the constructs (Lane et al., 2007).

Given the IAT structure requiring contrasting categories, choosing a companion category label to the category of interest is inevitable. The counterpart category labels should be sensible, mutually exclusive categories that are ideally from the same domain, such as choosing *family* as the counterpart target category for *career* (Lane et al., 2007).

In the present study, *Liberal arts* was selected as the counterpart target category to *Science* in SE-IAT. The phrases *STEM* and *non-STEM* were not adopted in order to keep instructions clear and easy to understand for participants (Cohen, Manion, & Morrison, 2003). Compared to 'science', an everyday vocabulary, *STEM* is more complex as an abbreviation for *Science, Technology, Engineering and Mathematics*. It cannot be guaranteed that every participant could quickly comprehend the word 'STEM' during the test. The incomprehensive jargon may add extra cognitive processing burden for participants and slow down their responses, resulting in construct-unrelated variance to the IAT effects. As such, it is not recommended to use *STEM* and *Non-STEM* as target category labels. Moreover, the category labels of *Science* and *Liberal* *arts* have been successfully used in the well-established Gender-Science IAT (Nosek et al., 2009). It is safer to use the existing labels that have been successfully validated in previous studies than to adopt untested new labels.

Regarding attribute categories, *rationality* was selected as the counterpart label to *Empathy*. *Rationality* has been used as the contrasting category to *Emotionality* in the Gender-Reasoning IAT (Smeding, 2012). Though *emotionality* is not completely identical to *empathy*, they are closely related concepts. *Rationality* serves well as a sensible, mutually exclusive concept to both *emotionality* and *empathy*, thereby it was borrowed from the Gender-Reasoning IAT to be used in the SE-IAT. It is worth noting that *apathy* or *indifference* may also work well as the counterpart to empathy. However, it is not preferred for two reasons. First, *rationality* has been validated in the Gender-Reasoning IAT while *apathy* and *indifference* have never been used in any existing IAT, therefore *rationality* enjoys a psychometric advantage over the other two options. More importantly, both *apathy* and *indifference* are prefixed words that simply mean "not empathy". It has been pointed out by Lane et al. (2007) that the use of negations of words or phrases such as "unintelligent" requires additional time for participants to correctly differentiate the contrasting concepts, therefore causing undesired construct-unrelated variance to the IAT effects. In this case, it is not recommended to simply use the negations of words as counterparts when selecting contrasting category labels.

4.2.2. Stimulus item selection

Furthermore, the items representing categories also matter to the IAT effects and so they need to be selected carefully as well. Researchers found varied IAT effects when they held the category labels constant but changed the items used to represent these categories (Govan & Williams, 2004; Mitchell et al., 2003; Nosek, Greenwald, & Banaji, 2005; Steffens & Plewe, 2001). For example, when pictures of disliked Blacks and liked Whites presented the categories *Black* and *White*, participants exhibited strong and significant preference for Whites over Blacks. However, the typically seen preference for Whites was significantly diminished when liked Blacks and disliked Whites represented the categories (Govan & Williams, 2004; Mitchell et al., 2003; Steffens & Plewe, 2001). Such findings indicate that selected stimulus items should represent the category in the same way. Consistency among stimulus items can prevent drastic changes in the elicited implicit attitudes, thus ensuring the reliability of the IAT. As such, in order to produce the most valid measure of corresponding implicit cognition, stimulus items that best represent the construal of the construct of interest without the confounding valence should be carefully generated first.

In addition, when using IAT as an individual difference measure, it is also important to make sure that participants can readily identify each item as denoting the appropriate category without extensive deliberation. By conducting a pilot study to select the items that can be quickly sorted into the corresponding category, we tried to maximize the variance of interest: that due to individual differences in the cognition but not task-specific variability (Lane et al., 2007). Ambiguity about an item's appropriate categorization may slow reaction time, adding construct-unrelated variance to the IAT effects (Nosek, Greenwald, et al., 2007). As such, items should be categorised solely on the basis of their membership in the corresponding category and not be confounded with any of the other categories (Steffens & Plewe, 2001).

According to the above-mentioned criteria for stimulus items, a two-step procedure was followed in the present study to choose the most appropriate stimulus items for the SE-IAT. In the first step, an item pool was generated on the basis of the words used in relation to the Gender-Science IAT, Gender-Reasoning IAT as well as the Empathy Quotient (EQ) questionnaire from earlier studies. In the second step, a focus group was organized to rate items on their representativeness of categories. Only items that could best represent one category without any ambiguous double memberships were kept as the final items. The two-step item selection procedure is explained in detail below.

4.2.2.1. Item generation

When generating the stimulus item pool for the SE-IAT, different approaches were employed for target categories (i.e., *science* and *liberal arts*) and attribute categories (i.e., *rationality* and *empathy*). The items for *science* and *liberal arts* were directly adapted from the Gender-Science IAT. The only change is that in the Gender-Science IAT, stimulus items are subjects (e.g., physics) but in the initial SE-IAT version created for the present study, stimulus items were adapted to professions (e.g., physicist). This change was made due to the purpose of examining stereotypes of individuals in STEM instead of specific subjects. See Table 4.3 in later Section 4.2.2.2.4 for the selected items representing *science* and *liberal arts*.

The attribute stimulus items for *rationality* were generated from the Gender-Reasoning IAT. Some synonyms of the original stimulus items from the GR-IAT were also included. As for the items representing *empathy*, keywords, including adjectives and nouns, were derived from the Empathy Quotient (EQ) questionnaire (Baron-Cohen & Wheelwright, 2004). For example, "caring" was extracted from the question "I really enjoy caring for others" in EQ as a candidate item for *empathy*. Finally, the item pool consisted of 30 items, 11 nouns, 18 adjectives and 1 two-word phrase. Table 4.1 presents the attribute item pool for SE-IAT.

Table 4.1 Attribute item pool for the preliminary SE-IAT

Rationality				
Analytical	Calculative	Consistent	Coherence	Deduction
Insight	Induction	Logical	Ordered	Organised
Rational	Reasoning	Systematic	Sanity	Standardised
Empathy				
Affection	Appreciation	Caring	Concern	Considerate
Comprehensive	Ethical	Empathetic	Emotion	Feeling
Intuitive	Soul	Sensitive	Understanding	Perspective-taking

4.2.2.2. Item selection

In order to select the attribute stimulus items that can best represent *rationality* and *empathy* without ambiguous category membership, a small group of university students was organized to rate the generated items about their representativeness of *rationality* and *empathy*.

4.2.2.2.1. Participants and procedure

Participants were recruited by oral invitation. The study was introduced as a short survey about vocabulary categorisation to random students from two colleges at the University of Cambridge. Participants were offered chocolate bars for their participation in the study.

The survey was carried out in the common rooms of the recruited participants' colleges using the traditional paper-and-pencil administration. Each participant was given a questionnaire comprising the written instructions, item categorization questions and personal information questions. The researcher was present at the survey session to guarantee that participants understood the instructions and completed the questionnaire independently. It took each participant less than 5 minutes to complete the questionnaire.

A total of 32 participants took part in the study, of which 2 had to be removed from further analysis for not having finished the questionnaire. The final sample of 30 participants consisted of 16 (53.4%) women and 14 (46.6%) men. The mean age was 25.9 years (SD = 5.2), ranging from 19 to 39 years. Regarding their major subjects, 20 (66.7%) participants were majoring in liberal arts and 7 (23.3%) participants were majoring in science, the remaining 3 participants did not report the major subject. Though information about their nationality and English levels was not collected, around 80% of the participants were Caucasians and 20% were Asians. Given that all overseas students enrolled in the University of Cambridge are required to demonstrate competence in the English language at a very high level by taking a standard English language test (e.g., The language requirements are 7.5 out of 9 for the International English Language Testing System (IELTS) and 110 out of 120 for the Test of English as a Foreign Language (TOEFL), all participants were deemed to speak proficient English so were eligible to evaluate the items.

4.2.2.2.2. Materials

A vocabulary categorisation scale was developed by the researcher to assess each item's representativeness of *rationality* or *empathy*. Following the item rating method introduced by Fleischhauer, Strobel, Enge and Strobel (2013), participants were asked to rate how much they associate each item with *rationality* or *empathy* on a 7-point Likert scale with the attribute anchors 'strongly rationality' versus 'strongly empathy' (See Appendix I for the full attribute item selection scale). Items associated with *rationality* were scored 1 to 3, the smaller the score the more representative was the item of *rationality*. On the contrary, items associated with *empathy* were scored 5 to 7, the bigger the score the more representative was the item of *empathy*. Items in the middle that scored close to 4 were the ambiguous ones that participants found could be associated with both *empathy* and *rationality*.

4.2.2.2.3. Results

Table 4.2 presents the mean categorisation ratings of the attribute items. Items with mean scores smaller than 4 are listed on the left as more associated with *rationality* in an ascending order, and items with mean scores bigger than 4 are listed on the right as more associated with empathy in a descending order. Items presented in the top rows were those with clear category membership rated closest to the two extreme scores (1 and 7) while those in the bottom rows were those with ambiguous category membership that rated close to the middle point (4).

Table 4.2 Mean ratings for attribute items associated with rationality or empathy

Rationality	Mean	Empathy	Mean
Rational	1.42	Empathetic	6.30
Logical	1.69	Emotion	6.18
Reasoning	1.79	Feeling	6.18
Organised	1.81	Affection	6.06
Deduction	2.00	Sensitive	5.97
Systematic	2.09	Caring	5.87
Coherence	2.51	Considerate	5.70
Consistent	2.52	Perspective-taking	5.27
Induction	2.75	Concern	5.25
Ordered	2.75	Appreciation	5.24
Analytical	2.85	Soul	5.06
Calculative	2.85	Understanding	4.72
Standardised	3.71	Intuitive	4.42
Insight	3.72	Ethical	4.18
Sanity	3.90	Comprehensive	3.85

4.2.2.2.4. Discussion

There are three rules in selecting the appropriate stimulus items for *empathy* and *rationalit*y. First and foremost, the selected items should be representative of only the corresponding category (Nosek, Greenwald, et al., 2007). Based on the results of the present study, items with mean ratings ranging from 2.75 to 5.25 were considered to have ambiguous double memberships of *rationality* and *empathy*, therefore were dropped from the final selection.

The second rule is to avoid items that can be categorised on the basis of irrelevant stimulus features. Specifically, selected items should have similar word length with different initials. For example, if all *rationality* items were less than 5 letters and the *empathy* items were more than 10 letters, participants could sort the items based on evaluative word length rather than the word meanings. Similarly, if all *rationality* items were words with the same initial "c", participants could sort these items based on similarities in initials rather than the category membership of the words. Such confounding valences of the items were taken into consideration and carefully avoided. For a good overview of the influence of valence variety and similar letter length on IAT effects, see Teige-Mocigemba, Klauer, & Sherman (2010).

The third rule is to ensure that the category membership of selected attribute items is clear and will not be confounded with target categories (Nosek, Greenwald, et al., 2007). For example, using "subject-related" *rationality* and *empathy* items such as "calculative" and "appreciation" could introduce confusion about whether to categorise the items on the basis of their membership to academic discipline or the attribute evaluation. Therefore, items with confounding membership to *science* and *liberal arts* were also dropped.

Furthermore, regarding the number of items in each category, evidence has shown that varying the number of items in both the target and attribute categories had no significant influence on the IAT effect magnitude (Greenwald et al., 1998; Nosek et al., 2005). Greenwald et al. (1998) reported no difference of IAT effect magnitude between stimulus sets of 25 items and of 5 items. As long as the number of stimulus items for each category is more than 2, the overall magnitude of implicit biases was consistent, and smaller number of items did not impair the reliability of the task, nor did it increase the influences of potential confounding factors. Such findings suggest that the magnitude and reliability of IAT effects were relatively unaffected by the number of stimulus items per category, except that effects were somewhat weaker when only a single stimulus item per category was used (Nosek et al., 2005). The common amount of stimulus items per category in the existing IATs is between 5 to 8. As such, we decided to choose 8 items (per category) that meet the three rules mentioned before. Table 4.3 presents the final selected items for the preliminary SE-IAT (1st version).

Table 4.3 Selected items for the preliminary SE-IAT (1st version)

Science	Liberal arts	Rationality	Empathy
Scientist	Artist	Consistent	Affection
Chemist	Linguist	Coherence	Considerate
Physicist	Philosopher	Deduction	Caring
Mathematician	Historian	Logical	Emotion
Engineer	Educator	Organised	Empathetic
Computer scientist	Anthropologist	Rational	Feeling
Astronomer	Sociologist	Reasoning	Perspective-taking
Biologist	Musician	Systematic	Sensitive

Note: Science and Liberal arts are target categories, and Rationality and Empathy are attribute categories

4.3. Step II: SE-IAT Procedural design

After having selected the materials for the SE-IAT, the next step was to design the procedure of how to administer the test. IAT has been used with procedural variations, usually without any intention to collect data to discriminate alternate versions. It is important to consider that measurement procedures are tools and different types of research require different kinds of tools (Gawronski & De Houwer, 2013). There is no procedure as the "best" one, only the most suitable paradigm for particular research aims. The present study followed the most widely used standard IAT procedure recommended by the founder of the test, Nosek et al.(2007), which has shown good psychometric properties in many existing studies. Potential construct-unrelated variables that need extra vigilance are also tackled with various means during the procedure. Different aspects with the procedure design, including the IAT structure, the number of trials, the inter-trial intervals and the order of compatible and incompatible tasks are explained in detail in the following sections.

4.3.1. Seven-block design

First of all, the SE-IAT adopts the standard 7-block structure design. There are three main categorization tasks in the SE-IAT: single-category classification (Block 1, 2, and 5), compatible configuration of double categorization (Block 3 and 4), and incompatible configuration of double categorization (Block 6 and 7).

The SE-IAT starts by training participants to press the left response key ("E" on keyboard) when a *Science* item appears on the screen and the right response key ("I" on keyboard) when a *Liberal arts* item appears on the screen. In the following Block 2, participants are trained to press left for *Rationality* items and right for *Empathy* items. The next Blocks 3 and 4 combine both discrimination tasks, making the so-called compatible combined blocks in which items representing *Science* and *Rationality* share the same right response key, whereas those representing *Liberal Arts* and *Empathy* share the left response key. The following Block 5 is again a single discrimination task and switches the positions of target categories, such that *Liberal Arts* items are assigned to the left and *Science* items are assigned to the right this time. The final Blocks 6 and 7 again combines the attribute and the previously reversed target discrimination, making the so-called incompatible combined blocks in which the *Liberal arts* and *Rationality* now share the same right response key and the *Science* and *Empathy* items share the left key. It is worth noting that in both combined blocks, the first set (Block 3 and 6) is for practice and the second set is the actual testing set (Block 4 and 7). Table 4.4 presents a schematic overview of the 7 blocks of the SE-IAT.

Table 4.4 Seven blocks in the SE-IAT

			Response Key Assignment	
Block	N trials	Task	Left key (E)	Right key (I)
1	20	Target discrimination	Science	Liberal Arts
2	20	Attribute discrimination	Rationality	Empathy
3	20	Initial combined task	Science, Rationality	Liberal Arts, Empathy
4	40	Initial combined task	Science, Rationality	Liberal Arts, Empathy
5	40	Reversed target discrimination	Liberal Arts	Science
6	20	Reversed combined task	Liberal Arts, Rationality	Science, Empathy
7	40	Reversed combined task	Liberal Arts, Rationality	Science, Empathy

Note. Nosek, Greenwald, and Banaji (2005) recommend increasing the number of trials in the fifth block to 40, in order to combat compatibility-order effect. As such, the SE-IAT has adopted this change and involved 40 instead of 20 trials in this particular block.

For participants who possess a stereotype-congruent belief that empathy is more related with liberal arts than with science, the incompatible combined blocks should be more difficult than the compatible combined blocks. On the contrary, the incompatible combined blocks should be easier for participants who possess a counter-stereotype belief that empathy is more associated with science than with liberal arts. Magnitude of associations is indexed both by the speed of responding (faster responding indicating stronger associations) and the frequency of errors (fewer errors indicating stronger associations).

4.3.2. Number of trials

A trial is defined as the time from the onset of a single stimulus item to the correct categorization of that item (Nosek, Greenwald, et al., 2007). For each trial, participants rapidly classify one item into the corresponding category. Regarding the number of trials in each block, evidence indicates that including 20 trials in Block 1, 2, 3 and 6 (that are blocks for practice) and 40 trials in Block 5 (that is the Block of reversed target discrimination) and critical Block 4 and 7 (that are Blocks of combined tasks for test) yields good psychometric properties (Lane et al., 2007; Nosek, Greenwald, et al., 2007). There is no evidence for any benefit of using more trials. A total number of 200 trials are included in the SE-IAT.

As can be seen from Table 4.4, the number of trials for practice in Block 5 (40 trials) of the subsequent combined blocks doubles the amount of trials for practice in Block 3 (20 trials) of the precedent combined blocks. This is designed with the aim of reducing the undesirable order effect of combined blocks. Considering that the precedent combined task usually interferes with the performance of the subsequent combined task, Nosek et al., (2005) suggested to double the number of trials for practice in the subsequent combined task to provide more time for participants to get prepared for the upcoming new categorization task in the subsequent combined blocks, as such to reduce the order effect.

4.3.3. Intertrial intervals

The interval between occurrence of one trial and the following trial – the intertrial interval – is set as 150 milliseconds in the SE-IAT. Though Greenwald et al., (1998) found no appreciable effect of using longer intertrial interval (750ms), a relatively short interval allows shorter test time, which is important to reduce the fatigue effect (Cohen et al., 2003). In this way, the procedure of the seven blocks can be administered within five minutes (Nosek, Greenwald, et al., 2007).

4.3.4. Order of combined blocks

As mentioned before, one of the common construct-unrelated effects observed on the IAT is the tendency for the precedent combined task to interfere with performance in the subsequent combined task (Klauer, 2005; Lane et al., 2007; Nosek et al., 2005). Specifically, participants who complete the compatible combined blocks first and the incompatible combined blocks later usually showed larger IAT effects than those who complete the combined blocks in a reversed order. Nosek et al. (2005) reported that the well-documented order effects on the IAT could not always be eliminated by extra practice trials in the subsequent combined blocks. Therefore, the order of the two critical combined tasks is also counterbalanced across participants with the purpose of controlling the order effect.

To conclude, the SE-IAT adopted the standard 7-block structure design with two single-categorisation blocks, two compatible combined blocks, one reversed target categorization block, and two incompatible combined blocks. There are 200 trials in total and the intertrial interval is set as 150 milliseconds. The order of the compatible and incompatible combined blocks is counterbalanced for the purpose of controlling the undesired order effect.

4.4. Step III: Creating and testing the SE-IAT website

4.4.1. Web versus Lab-based research

As a computer-based timed task, SE-IAT was administered using a customized website so that all the data for the current PhD project could be collected online. There are three reasons for the decision of utilising web-based technology. Firstly, a web-based survey can guarantee anonymity and minimize the respondent's feelings of jeopardy by avoiding the observation of human administrators (Coutts & Jann, 2011). Given that stereotype is a sensitive topic, self-administrated online surveys can increase responses and yield more honest, candid answers (Coutts & Jann, 2011). Secondly, a web-based study can easily incorporate complex skip patterns to randomise study items automatically. In the present study, not only the order of IAT test and the questionnaires but also the order of the items within each test and questionnaire were required to be counterbalanced. Web-based technology allows relatively easy implementation of counterbalanced items online. Thirdly, a web-based study enables fast and large data collection (Bethlehem & Biffignandi, 2011). Compared to the traditional lab-based study requiring experimenters to administer studies to participants at the same time in the same place, a web-based study is more flexible allowing multiple participants to run studies from their home computers at the same time at different locations. Though some data cleaning is necessary, online survey data can be ready to analyse very soon after the field work is carried out (Couper, Traugott, & Lamias, 2001).

Nevertheless, disadvantages of a web-based study were not ignored and remedies to these disadvantages were involved. One major concern regarding web-based studies is whether data collected online is comparable to data gathered in lab-based settings. Online testing makes it impossible to have a standardised testing environment for all participants. Whether it is at home, at work, at a café or wherever, participants are potentially in environments full of distractions, which may threaten the validity of the study (Couper et al., 2001). To reduce such environmental bias, a statement requiring participants to complete the present study on their own in a quiet environment with stable Internet access was included in the introduction webpage (see Appendix IV) at the beginning of the study. Concerns have also been raised that in online studies participants may easily succumb to careless responding, experiencing a lack of focus, or even deceive in their answers as they are not monitored by the experimenters (Kraut et al., 2004). Given the present study includes a timed IAT task, data from participants who respond too fast or make too many mistakes can be identified as unqualified and removed from further analysis. In addition, though studies have shown that multiple submissions are rare in web-based research, it can occur that the same participant may attempt to repeat the study whether deliberately or not (Stieger & Reips, 2010). Multiple submissions are controlled for in the present study by recording the user's IP addresses and devices.

Aside from the participants' internal factors that may affect the reaction time in a web-based study, various external factors may cause unsystematic errors as well. Concerns are raised that the reaction times are very likely to vary depending on the different computer hardware and software that the participants use. For example, keyboards have been found to have different sampling rates and monitors have been found to have various refresh rates. Furthermore, different web browsers also contribute to the uncertainty in precision timing. Similarly, timing may also be influenced by the operating system running multiple processes at the same time thus slowing its processing time down (Cek, 2015). Researchers suggest that increasing the sample size is the easiest way to reduce the nonsystematic noise in terms of timing, which might affect the sensitivity to small effects of reaction time (Barnhoorn, Haasnoot, Bocanegra, & van Steenbergen, 2015).

Despite all the mentioned pros and cons of web-based study, studies have shown that traditional laboratory findings can be replicated in online samples, suggesting that the differences in timings online are relatively trivial. A recent study by Houben and Wiers (2008) comparing IAT effects obtained via the Internet and those obtained using Inquisit (standard reaction time software used in the laboratory) revealed no difference between the two conditions. Such results indicate that the online IAT versions are as sensitive to individual differences as are the IAT versions in the standard laboratory settings, which is in line with the initial findings by Nosek, Banaji and Greenwald (2002a) showing that data regarding attitudes and stereotypes in a laboratory did not differ significantly from data collected via a website. Considering the above-mentioned advantages of web-based research, the present PhD project adopted the web-based approach and implemented all the study materials online in one website. Instruments were included in the website for participants to self-administer the study on their own personal computers.

4.4.2. Creating the SE-IAT website

4.4.2.1. Brief overview of major existing software packages for timed test

There are a number of reaction time packages, both commercial as well as open-source, that are available for computerizing IAT nowadays. Either option comes with a certain trade-off. Open-source software packages usually require researchers to have knowledge of certain programming language(s) in order to create a customized test. Whilst most commercial software packages offer a user-friendly interface with drag-and-drop features, they do not require any programming skills. But those commercial software packages are often not offered online like most open-source packages. Table 4.5 summarised the advantages and disadvantages of current available reaction time software packages.

Table 4.5 A brief overview of reaction time software packages (Cek, 2015)

Softwares	Online?	Requiring Programming skills?	Advantages	Disadvantages
Commercial				
DirectRT	No	Yes	Simple syntax to follow to customize test	Can only be used offline
E-Prime	No	No	No programming skill needed	Can only be used offline; need to become acquainted with a multi-window visual user interface first
Inquist for Web	Yes	Yes	Can be used with third party questionnaire software such as Survey Monkey	Requiring relatively complex syntax to program tests
Superlab	No	No	No programming skill needed; questionnaires can be implemented at the same time	Can only be used offline
Open-source				
FreeIAT	No	No	Avoids Internet reaction time issues	Can only be used offline; available only with Windows system
WebIAT	Yes	Yes	Tested on most systems and browsers	Requiring researchers to host a Web server
WEXTOR	Yes	No	Fulfills all ethical requirements for web-based studies	Hasn't been tested for reaction time studies
OpenSesame	No	Yes	Highly flexible and adaptable	Not online; requiring relatively complex

Softwares	Online?	Requiring Programming skills?	Advantages	Disadvantages
				programming
jsPsych	Yes	Yes	Fulfills all ethical requirements for web-based studies	Requiring researchers to host a Web server; requiring extensive programming skills
Tatool	Yes	Yes	Avoids Internet reaction time issues	Requiring complex programming knowledge; incompatible with third party questionnaire software
WebExp	Yes	Yes	Fulfills all ethical requirements for web-based studies	Requiring extensive programming skills; incompatible with third party questionnaire software
JATOS	Yes	No	Very simple user-friendly graphical interface	Hasn't been tested for reaction time studies; incompatible with third party questionnaire software
ORTEngine	Yes	No	Very simple user-friendly graphical interface	Less flexible; requiring complex programming knowledge to be used with third party questionnaire software
ScriptingRT	Yes	Yes	Validated for reaction time studies	Requiring extensive programming knowledge; incompatible with third party questionnaire software

Note. This table is adapted from Cek (2015).

4.4.2.2. Using the Project Implicit virtual lab to implement SE-IAT online

There are three requirements when selecting the appropriate software to create the SE-IAT website. First, due to the time constraints of the present PhD project, it is impractical to expect the researcher to put much effort in learning a new programming language from scratch, thus software packages requiring programming skills including *Inquisit for Web*, *OpenSesame*, *Webexp*, *ORTEngine*, and *ScriptRT* were ruled out. Second, the SE-IAT needs to be administered with self-report questionnaire, reaction time software packages that cannot be used with third party questionnaire software including *Tatool* and *JATOS* were not selected. Third, software packages that can only be used offline including *DirectRT*, *E-Prime*, *Superlab* and *FreeIAT* were dropped as well. The remaining open-source software packages *WebIAT* and *WEXTOR* fulfill all the three requirements, but neither was selected because *WebIAT* requires researchers to host a personal web server and *WEXTOR* had not been tested with any reaction time task before.

Under such circumstances, the task of creating a website to administer SE-IAT and self-report questionnaires online was outsourced to the technical service team from Project Implicit (PI), a virtual lab managed by IAT experts from Harvard University. The homepage link to PI is *implicit.harvard.edu.*

The Project Implicit Virtual Lab was selected because 1) both implicit measures and questionnaires could be easily implemented in one web browser; 2) the virtual lab possesses a mature and stable implicit measure infrastructure that has been used and tested successfully by many other researchers; 3) anonymity and security in transmission of individual responses are secured and 4) large data storage is provided.

Before implementing the study, the researcher was required to sign a contract with the PI team to set up the number of studies and the time length for data collection. In the current project, the researcher was allowed to conduct two rounds of data collection.

Before each round of data collection, the researcher would be given a test link to examine and modify all the implemented research materials. The website would be valid for only one year.

The first test link to the website was sent to the researcher on 14th June, 2016.
The test link is as below:

https://app-prod-03.implicit.harvard.edu/implicit/Launch?study=/user//nlofaro/contract.qin.empathy/manager.expt.xml&refresh=true

Figure 4.1 is a screenshot for the incompatible pairing block of the online SE-IAT. When conducting the SE-IAT task, the category labels appear on the top left and right of the computer screen to remind participants of the response key mapping rules. The target and attribute labels are in different colours for the purpose of enhancing the distinctiveness of nominal features of given stimulus. Specifically, the *Liberal arts* and *Science* labels and items are in green while the *Rationality* and *Empathy* labels and items are in blue. The stimulus items for sorting are put in the middle of the screen. When a stimulus item is correctly categorised, a new item will immediately appear. But when the stimulus item is incorrectly categorised, a red "X" below the stimulus item will appear and the participants are obliged to fix the error before proceeding.

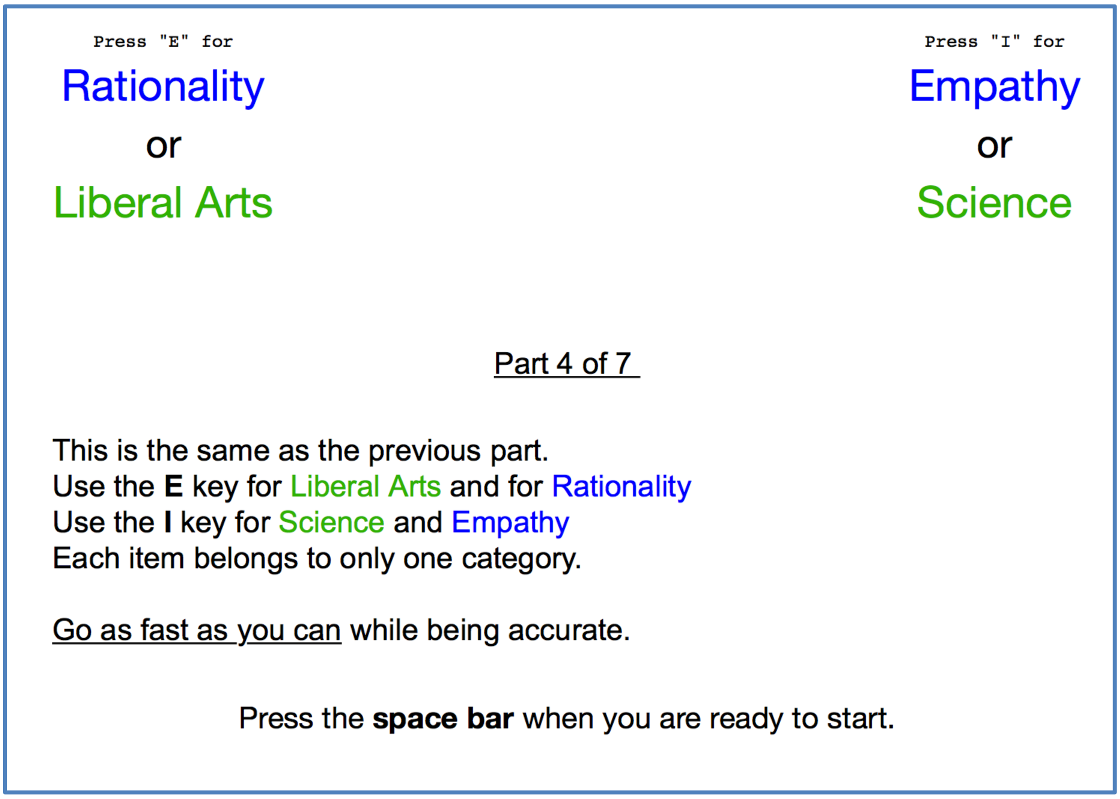


Figure 4.1 A screenshot for the SE-IAT incompatible pairing block

4.4.3. Testing and modifying the SE-IAT website

Before using the newly developed website to collect data, it is important to test the practicality and face validity of the implemented study materials online. A focus group was organized to go through all the measures in the website using their own computers to see if they can successfully complete the study online.

4.4.3.1. Participants

A focus group of 6 participants was organized to test the link. According to Morgan (1997), the optimum size of a focus group should be 6 to 10. Moreover, Krueger and Casey (2009) have also urged caution of the use of homogeneous group - a group of people who have a common background, position or experience – because homogeneous groups have their own well-established dynamics that can influence contributions. Considering these requirements, the gender and academic majors of the participants in the focus group are balanced. Moreover, the focus group must be able to represent participants in the main study. Given that I proposed to investigate stereotypes of empathy in scientists among UK university students, the final participants in my focus group were 6 students from the University of Cambridge. The information about age, gender, major subjects and nationality of the 6 participants is shown in Table 4.6.

Table 4.6 Demographics of the focus group for testing the initial IAT website

No.	Gender	Age	Subject	Nationality
1	Female	32	Pathology	Chinese
2	Female	26	Education	British
3	Female	27	Psychology	British
4	Male	23	Economics	Chinese
5	Male	21	Mathematics	British
6	Male	21	Medicine	British

4.4.3.2. Procedure

The focus group was required to go through all the study materials implemented in the website using their own personal computers. The main purpose is for them to review the practicality and face validity of the online SE-IAT. Face validity is the degree to which a measure appears to be related to the specific construct it is supposed to measure, in the judgment of non-experts such as test takers (Furr & Bacharach, 2008). Before participants start the test, they were informed of the research aim and were encouraged to give detailed advice on how to improve the test website in terms of revising the content of SE-IAT as well as making the whole procedure easier to self-administer from the perspective of test takers. Participants were asked to give answers to the question "*Did you have any difficulty completing the study? If so, what would you suggest the researcher to do to improve the study?*". This open question was intended for stimulating discussion among participants to bring up any important issues that had not been identified in the design phase. As a result, several problems were pointed out and modifications were made accordingly. Any item that the focus group suggested to be unrelated to the supposed construct were dropped or revised. Instructions that were found unclear were rephrased. The detailed updates are presented in the following paragraphs.

4.4.3.3. Results and updates

The newly developed online SE-IAT was successfully run with all participants' personal computers with different operating systems including Microsoft Windows and Mac OS. Reaction time was recorded accurately in milliseconds. However, there are six issues observed and the website was updated accordingly.

4.4.3.3.1. <u>Items for Science and Liberal arts were changed from occupations to subjects</u>

After discussion of the focus group, items representing target categories were changed back from occupations to subjects for two reasons. First, the test takers reported that they found it easier for them to associate subjects than occupations with target categories of *Science* and *Liberal arts*. One of the test takers reported: "*I would prefer subjects to occupations as the stimulus items. I think the subjects are related to the label Science and Liberal arts more directly. Instead, occupations are more likely to be related to the label Scientists and Artists.*" Moreover, some occupations were found too ambiguous for test takers to sort into just one category. For example, one test taker who studies mathematics reported: "*I found Philosopher could be sorted into both Science and Liberal arts because several mathematicians were also known as philosophers, such as Bertrand Russell and Rene Descartes.*" On the contrary, he said "*But I would have no problem sorting Philosophy into the category Liberal arts given that philosophy is generally known as a subject in liberal arts.*" In this case, the stimulus items for *Science* and *Liberal arts* were changed from occupations to subjects, which are again in line with the items in the well-established Science-Gender IAT and Science-Rationality IAT.

4.4.3.3.2. Ambiguous items with double membership were identified and dropped

In addition, discussion of the focus group also revealed certain ambiguous items that could be sorted into more than one category. The item *anthropologist* and *educator* were reported to be confusing because some anthropology research also draws upon physical and biological sciences and some educators can be specialized in science education. Therefore, these two items were replaced with items of little confusion such as *Literature* and *Fine arts* to represent the target category *Liberal arts*. Test takers reported no confusion with stimulus items for attribute categories of *empathy* and *rationality*.

4.4.3.3.3. Wording of the attribute items for Empathy and Rationality was polished up

Regarding wording and phrasing of the test, English native speakers in the focus group also gave three suggestions. First, all items for attribute categories of *Empathy* and *Rationality* were unified into adjectives. For example, the word *emotion* was adjusted to *emotional* to be identical with all the other adjective items representing the category *Empathy*. Second, the compound word "*perspective-taking"* was replaced by its synonym "*thoughtful"* for it was the only two-word phrase, making it different from all the other one-word items that might lead to construct-unrelated variance. Third, one of the test takers pointed out that the adjective "*affective*" in the Empathy category was easily mistaken for "*effective*", which could be sorted into the opposite *Rationality* category. To avoid a confounding effect, the word "*affective*" was replaced by a more frequently-used word "*affectionate*". Table 4.7 displays the updated items for the SE-IAT.

Table 4.7 Updated items for the preliminary SE-IAT (2nd version)

Science	Liberal arts	Rationality	Empathy
Chemistry	Fine arts	Consistent	Affectionate
Physics	Linguistics	Coherent	Considerate
Mathematics	Philosophy	Deductive	Caring
Engineering	History	Logical	Emotional
Computer science	Literature	Organised	Empathetic
Astronomy	Sociology	Rational	Feeling
Biology	Politics	Reasoning	Thoughtful
Geology	Music	Systematic	Sensitive

4.4.3.3.4. Device requirement was added in the instructions

Some participants tried to click on the test link with their tablets and found it unable to be run on tablets or smartphones. Therefore, a requirement specifying the valid devices for the test was added in the instructions. Participants are asked to use desktop or laptop computers to complete the test. Smartphones or tablets without a physical keyboard were warned to be invalid for the test.

4.4.3.3.5. A test progress indicator was added for participants to track their progress

When completing 200 trials in the SE-IAT, participants reported feeling bored and impatient during the process for they had no idea how long it would last. A test progress indicator (e.g.,"*Block/Question 1 of 7*"), which shows the current phase of the study, was added for participants to track their progress during the test in order to reduce the fatigue effect and improve concentration on the study.

4.4.3.3.6. A question for academic discipline identification was added

A question asking the participants' self-identification with *science* or *liberal arts* was added considering the situation that some participants may major in integrated subjects. It has been considered not only inappropriate but also impossible for the researcher to judge the academic discipline that the participants affiliated to. Therefore, it is essential for participants themselves to report their identified academic discipline.

4.5. Limitations and suggestions for future research

The current approach of designing and developing the SE-IAT website has several limitations. Firstly, in neither pilot study were the participants truly representative of the target population in the main studies. When selecting items for SE-IAT, participants were mainly graduate students from two colleges at Cambridge University. Similarly, when testing the website, participants were limited to 6 students from Cambridge University. Though the gender and major ratio were balanced in each pilot study, the limited number of participants may restrict the generalisability of their ratings and opinions on the test.

Secondly, though the face validity of the SE-IAT was tested with a focus group, the specific SE-IAT scores were not calculated. It remains unknown whether the newly developed SE-IAT can successfully capture implicit bias toward empathy in scientists. Moreover, psychometric properties of the SE-IAT (e.g., split-half reliability) were not examined. Future research is required to validate the SE-IAT with participants of a larger sample size by checking the SE-IAT effect and other indices of its psychometric properties.

Thirdly, only one version of the SE-IAT was tested and updated in the current study. However, systematic variations of task features (e.g., different items, number of trials, intertrial intervals) may alter performances (Baron & Banaji, 2006; Fleischhauer et al., 2013; Lane et al., 2007). For example, Fleischhauer et al., (2013) tested and examined the psychometric properties of four preliminary IAT versions with different category and stimulus item labels in their preliminary study to finalize a new IAT measuring cognitive motivations. All the four preliminary IAT versions yielded different results when examining their construct-unrelated variances. Future studies should take a more rigorous approach when develop new IATs by examining alternative preliminary IAT versions with different task features.

Finally, the SE-IAT website was established using a paid implicit test development service. The commercial online IAT lab is costly and inflexible because researchers are not allowed to modify the IAT directly. However, more flexible open-source software packages for timed tests usually require programming expertise (Cek, 2015). Considering the rapid growth of the need to use IAT in various research fields, future efforts are called for to develop user-friendly open-source IAT online platforms for researchers with limited programming skills to customize IAT websites for various research purposes.

4.6. Conclusion

The current preliminary study designed and created a preliminary SE-IAT website for the purpose of making up the lack of implicit measures for stereotypes of empathy in scientists. With this newly developed implicit measure, we can have a better understanding of the up-to-date stereotypes of empathy in scientists by examining the underlying bias toward empathy in scientists that people are unable to report in questionnaire.

This chapter reports the systematic procedure of developing an IAT. As a relatively novel implicit method which effect can be contaminated by many construct-unrelated variables, many problems may occur during the process of developing an IAT, obliging researchers to show scientific responsibility in using the IAT paradigm to customize their own implicit tests. Therefore, it is deemed important for researcher to provide a holistic approach of designing an IAT and highlights several important remedies for the respective contaminants. The three-step approach to develop the SE-IAT website has been presented as a good example for future researchers to follow the steps to create their own IAT.

In line with previous findings that the IAT effect is influenced by variations in valence of category labels and items (Bluemke & Friese, 2008; De Houwer, 2001; Fleischhauer et al., 2013; Gast & Rothermund, 2010), the current study has also found evidence that both definitions as well as morphological features of items can affect the face validity of an IAT. Methods such as using item ratings to examine membership of categories as well as using focus group to re-evaluate representativeness of items have been applied in the current study to reduce such construct-unrelated variance. After the SE-IAT was designed and developed in the current stage, the new test needs to be validated by examining its psychometric properties in the next stage.

Chapter 5 Stage Two: Testing and modifying the IAT to measure stereotypes of empathy in scientists

5.1. Introduction

The present PhD project aims to investigate individual differences in the implicit and explicit stereotypes of empathy in scientists. As no implicit measure on stereotypes of empathy in scientists has been created to date, a preliminary Science-Empathy IAT was systematically designed and created in previous Stage 1 (see Chapter 4). As mentioned in the literature review, controversial evidence has been found for certain psychometric properties (e.g., explicit-implicit correlations) of different kinds of IATs, therefore heated debate has never stopped on whether IAT can pass as a sound tool to capture variations in implicit social cognition. Under such circumstances, it is of paramount importance to evaluate whether the newly developed IAT meets relevant psychometric criteria before using it to assess individual differences.

In this chapter, the preliminary SE-IAT was tested against certain psychometric criteria. Given the limited scope of the present study, not all psychometric criteria, such as the convergent validity with other implicit measures, could be assessed. But the following critical criteria were applied and examined in order to validate the appropriateness of IAT to address relatively stable individual differences in the stereotypes of empathy in scientists. Firstly, the measure should be able to replicate an IAT effect, which means that there should be significant differences in response latencies between the incompatible and compatible blocks. That is to say, participants are expected to respond slower as well as to make more mistakes when completing the incompatible combined blocks as compared to the compatible combined blocks of the IAT. Meanwhile, the measure should demonstrate acceptable reliability and validity. As for the reliability, the split-half reliability, which is the most commonly examined reliability for IAT, was drawn on. As for the validity, various indicators were scrutinized to evaluate the IAT including the face validity (i.e., the extent to which the wording of items matches the concept assessed); Known-groups validity (i.e., the extent to which the measure is able to differentiate between members of groups based on prior knowledge or predictions about them); Implicit-explicit correlations (i.e., the extent to which the IAT is similar or different from related self-reports); and construct-unrelated variance that may threaten the validity (e.g., order effect and effect of prior experience with IAT).

5.2. Method

5.2.1. Sampling and participants

All participants were recruited through their departments or colleges from four Russell Group universities located in or close to London. The Russell Group universities were selected because these universities are all comprehensive universities providing a wide range of academic courses. Given that one of the main purposes of the present study is to investigate variations in stereotypes between different academic majors, it is essential to recruit participants in both STEM and non-STEM fields from each university. However, when examining the subjects available in universities with lower rankings, it has been found that these universities usually do not provide math-intensive subjects such as physics or mathematics, therefore only Russell Group universities that provide both STEM and non-STEM courses were targeted.

The opt-in sampling method was adopted to recruit participants. An advertisement email containing a brief introduction to the project and the link to the study website (see Appendix III for the study advertisement) was sent out to all the students in the target departments or colleges by their department administrators or college student representatives from 9th August to 3rd December, 2016 (see Appendix II for the participant recruitment email).

Before the demographic information of the present sample is presented, it is necessary to clarify three things. Firstly, concerning the potential replicated responses, information about the IP addresses, the computer systems, the web browsers, and the time when the study had been done for each participant was routinely recorded (see Appendix VIII for an example of the session record). Participants were informed about this action in the consent form and the collected information was kept strictly confidential and only available to the researcher. A thorough examination of the recorded IP addresses revealed no replication of the responses in the present study.

Secondly, the academic fields of the participants are decided by their own identifications rather than categorised by the researcher. Participants were asked to choose *"the academic field you mostly identified with"* from four options of *science*, *liberal arts*, *others*, and *don't know* and were also asked to provide information about their specific majors. Nevertheless, it was found that participants who were in the same majors could be identified with different academic fields. Such situation may happen due to the fact that students in the same majors can have various concerns about the topic or take different methodological approaches to do research, therefore may be identified with either science or liberal arts. For example, students in Human Geography are more likely to be identified with Liberal arts while others in Physical Geography are more likely to be identified with Science. Also, it was not uncommon that postgraduate students could have studied more than one subject from different academic fields. Under such circumstances, it is considered more appropriate to let participants identify their own academic fields than to use their reported majors to classify participants into different academic fields.

Thirdly, it is necessary to clarify that not all, but most, participants in the present study are native UK citizens. Although the present study initially planned to target only UK domestic students, it was deemed rather troublesome by the gatekeepers of the participating universities to identify them to send the invitation email. As a result, the invitation email containing the study link was sent to both domestic and international students altogether for the convenience of the gatekeepers. Although it was stated in the email that native English speakers were preferred, it turned out that quite a few non-native students still clicked on the study link and decided to take part in the study.

Therefore, questions regarding the nationality and English level of the participants were added. All non-native participants showed sufficient English language ability to complete the study (International English Language Testing System [IELTS] scores above 6.5). The detailed demographic information is presented as below.

Among 1098 clicks on the study link, 580 students continued to participate in the study. However, 50 participants who did not finish the IAT test were removed from the final sample. Moreover, according to guidelines created by Greenwald, Nosek, and Banaji (2003), 10 participants who made too many errors in the IAT test, meaning that they sorted the items into the wrong categories too many times (error rates > 20%) as well as 35 participants who did the IAT test too fast (reaction time in each trial less than 300 milliseconds > 10%) were also dropped from the final sample. There was no trial in the IAT test greater than 10,000 ms in the present study.

Consequently, the final sample is comprised of 485 participants. The female to male ratio is 60.0% to 38.4% (291 females to 186 males, missing data 8). The science to liberal arts ratio is 58.4% to 34.8% (283 identified with science to 169 identified with liberal arts, the rest 25 identified with other academic field occupying 5.2% and missing data 6). The respective female to male ratio of participants identified with science and liberal arts is presented in Table 5.1. As can be seen from the table, the female to male ratio with participants identified with liberal arts in the sample is imbalanced with only 29.0% participants being male. The gender ratio is more balanced for participants identified with science.

Table 5.1 Female to male ratio in Science and Liberal arts of Stage 2 final sample

Gender	Science (n = 282)		Liberal arts (n = 169)	
	N	Percentage	N	Percentage
Female	152	53.3%	120	71.0%
Male	130	46.7%	49	29.0%

Note: There are 25 identified with other academic field, and 9 missing values

The majority of participants (86.2%) were in the age range of 18 to 25 years old (See Table 5.2 for details). Also, in terms of education level, three quarters of participants were pursuing Bachelor degrees whereas 24.3% were postgraduates in the final sample (See Table 5.3 for details).

Table 5.2 Age range of the Stage 2 final sample

Age range	N	Percentage
< 18	4	0.9%
18 - 20	279	59.2%
21 - 25	127	27.0%
26 - 35	52	11.0%
36 - 45	4	0.8%
> 45	5	1.1%

Note: There are 14 missing values.

Table 5.3 Education level of the Stage 2 final sample

Education level	N	Percentage
Bachelor's degree	351	74.7%
Master's degree	63	13.4%
Doctoral degree	51	10.9%
Postdoc	5	1.0%

Note. There are 7 participants identified with other education levels, and 2 missing values.

In terms of ethnicity, participants were asked to directly fill in their ethnicity in the blanks of the question. According to the classification of ethnicity by the Higher Education Statistics Agency (HESA, 2018), participants were classified into four ethnic groups. The majority of participants (74.2%) were White (See Table 5.4 for details). The proportions of different ethnic groups of Stage 2 study are also in consistent with the national data (HESA, 2018). Finally, regarding the English level, most participants (77.4%) were native English speakers as shown in Table 5.5.

Table 5.4 Ethnic group of the Stage 2 final sample

Ethnic group	N	Percentage
White	329	74.3%
Asian/Asian British	47	10.6%
Black/African/Caribbean/Black British	5	1.1%
Mixed/Multiple/Other	62	14%

Note. There are 42 missing values

Table 5.5 English level of the Stage 2 final sample

English level	N	Percentage
Native English speaker	367	77.4%
Non-native speaker	107	22.6%

Note. There are 11 missing values

5.2.2. Procedure

As explained before, the online survey approach was adopted to collect data throughout my PhD project (see Section 4.4.1 for the reasoning of taking online survey approach). Starting from 17th May, 2016, participant recruitment emails (see Appendix II) were firstly sent to the gatekeepers of the colleges and departments from the target universities. After gaining approval from the gatekeepers, the recruitment emails, which contains the link to the study website, were sent to all the students from their institutions. According to the record, the first response was gained on 9th August and the last on 3rd December, 2016.

After clicking on the study link, an introduction webpage providing instructions about the study was presented first before participants proceeded to the actual study. Participants were informed that the present study is to look at individual differences in attitudes toward or beliefs about scientists and they would be required to answer some questions and a timed task in which they need to sort words into categories as quickly as possible. Only once they have read the instructions and agreed to take part in the study were the participants able to move on to the next step by clicking the "continue" button at the right bottom of the page (see Appendix IV for the introduction webpage).

The order of the IAT and the self-report questionnaire is counterbalanced. Little evidence has shown that the order in which self-report and IAT measures are completed would affect IAT performance as well as self-report (Hofmann et al., 2005). Thus, the present study followed the procedural guideline composed by Nosek, Greenwald, and Banaji (2007) to counterbalance the order of IAT and self-report measures in the absence of reasons for using just a single order.

The order of the compatible and incompatible blocks in the IAT is also counterbalanced, given that the precedent combined task usually interferes with the performance of the subsequent combined task (Klauer, 2005; Lane et al., 2007; Nosek, Greenwald, et al., 2007). Demographic information was gained after participants completed both the IAT and the self-report questionnaire.

Participants were not informed of their own results. A debriefing was provided to explain what an implicit attitude is and how the IAT works. Cautions were also stressed that implicit bias captured by the IAT may not be consciously endorsed in order to prevent participants from misinterpreting the test and overworrying about their performance (See Appendix VII for the full debriefing). At last, participants were also asked to provide their email addresses if they would like to take part in the lottery draw. At the end of the study, 10 participants who participated in the lottery draw were randomly selected and given Amazon vouchers as incentives. It usually took participants 10 to 15 minutes to complete the whole procedure. The link to the present study website is as below (See Appendix V for screenshots of the SE-IAT procedures):

https://app-prod-03.implicit.harvard.edu/implicit/Launch?study=/user//nlofaro/contract.qin.empathy/manager.expt.xml&refresh=true

5.2.3. Ethical issues

5.2.3.1. Ethical Approval

Guidance for conducting ethical research within the British Psychological Society Code of Ethics and Conduct (BPS, 2009) and the Revised Ethical Code for Educational Research (BERA, 2018) was carefully followed. Study documentation including a research proposal, consent form and ethics and risk assessment form, was submitted to the Ethics Committee of the Faculty of Education, University of Cambridge.

5.2.3.2. Valid consent

In the recruitment letter (See Appendix II), a brief explanation of the present project was provided. Once they clicked on the test link included in the recruitment letter, participants were first led to the introduction webpage (See Appendix IV). This webpage detailed the voluntary nature of the research, confidentiality within the research process and the right to withdraw at any minute during the process without any question. All participants were deemed capable of giving valid consent to participate in the study. After they agreed to take part in the study, participants then proceeded to the test by clicking the "continue" button on the bottom of the webpage.

5.2.3.3. Confidentiality

Data were collected and analysed in accordance with the Data Protection Act (2018). Online data were protected by SSL encryption. Participants did not provide their names and each response was given a reference number automatically. IP addresses were routinely recorded but were accessible only to the researcher. The raw study data were securely stored in a password protected account on the Project Implicit website, which only the researcher and the Project Implicit team had access to. The Project Implicit team was not allowed to use or distribute the data without permission of the researcher for any other purpose. Data were saved and analysed using the researcher's password-protected personal laptop.

5.2.3.4. Debriefing

There was no anticipated negative impact of the research on participants. Inevitably, the research may draw the attention of the participants to both conscious and unconscious bias about scientists. But it is considered beneficial rather than detrimental. However, in order to prevent participants from misinterpreting their performance, they were not provided with their own scores. Participants were debriefed in the end of the study about the nature of implicit attitudes and the difference between these and consciously endorsed attitudes (See Appendix VII). A link to the Frequently Asked Questions webpage was provided. Contact information of the researcher was provided for the participants to contact the researcher if they have any issue relating to the study.

5.2.4. Measures

In the present study, both the IAT and the explicit self-report measures are applied. As introduced in the Literature Review chapter, IAT is designed to assess relatively automatic mental associations that are difficult to gauge with explicit self-reports. One important indicator for the success in assessing unconscious attitude representations is the finding that implicit measures often show rather low correlations with explicit measures (Fazio & Olson, 2003; Gawronski et al., 2007; Hofmann et al., 2005). Given the main purpose of the present study is to validate an IAT for implicit stereotype of empathy in scientists, it is essential to include related explicit self-reports to check if there is any distinction between the outcomes of implicit and explicit measures.

In terms of selecting and adapting the appropriate self-reports for stereotypes of empathy in scientists, three aspects were taken into consideration. To begin with, the selected self-report measures should focus on examining similar concepts that the SE-IAT claims to measure, that is, in this study, perceptions about empathy in scientists, not other characteristics such as the stereotype about appearance or working environment of the scientists. Evidence has shown that the *conceptual correspondence* (i.e., the extent to which the attitudes are measured at the same level of abstractness and with the same degree of specificity; Ajzen & Fishbein, 1977) between implicit and explicit measures have significant influence on their correlations (Hofmann et al., 2005). For example, many standard self-report scales for racial prejudice (e.g., Modern Racism Scale, McConahay, 1986) assess participants' political opinions on issues such as affirmative action or the discrimination of ethnic minority members. In contrast, racism IATs typically assess participants' responses to members of ethnic minority groups (e.g., faces of Black and White individuals). Thus, even though general evaluations of ethnic minority members may be systematically related to people's political opinions (Gawronski et al., 2008), the low correlations between these self-report racial prejudice scales and the IAT measures are very likely to be resulted from their conceptual distinctness, not necessarily because one is tapping unconscious representations (Gawronski et al., 2007; Hofmann et al., 2005). Therefore, cautions are stressed on the content of the selected self-report measures to be consistent with what the SE-IAT intends to address in the present study.

In addition, the *structural fit* (i.e., the degree of methodological similarity between different tests, Payne, Burkley, & Stokes, 2008) between the explicit measures and the

IAT also deserves extra attention. By the *structure* of a measure, researchers mean the parts that make it up and how they work together to measure attitudes (Payne et al., 2008). In most explicit self-reports, participants are required to make evaluations on a series of propositional statements. In contrast, the IATs avoid all the propositions involved in the self-reports, and instead only assess relative attitudes toward two different target concepts based on the comparisons between reaction time to register different responses. As concluded by Payne et al. (2008), the structures of IATs and the self-reports differ in three key ways, including the stimuli presented (e.g., propositions vs. simple words or pictures), the level of abstractness of the judgments to be made (e.g., broad social opinions vs. concrete classifications), and the metric in which responses are measured (e.g., numerical scales vs. response latencies). Moreover, IATs assess relative rather than absolute attitudes. That is to say, given that IATs consist of two groups of contrasting concepts (e.g., target concepts: liberal arts and sciences; attribute concepts: empathy and rationality), instead of measuring the absolute evaluation of the target concept (e.g., whether scientists have good empathy), they actually can only measure comparative evaluations of the target concepts (e.g., whether empathy is more correlated with liberal arts as compared to science). Accordingly, correlations with the IAT may be higher for self-reports assessing corresponding relative attitudes as compared to those measuring absolute propositional opinions (Gawronski et al., 2007; Hofmann et al., 2005). In this case, it is deemed necessary to incorporate both relative and absolute self-reports in the present study for the purpose of clarifying the correlations between implicit and explicit measures, which have usually been used as an index for the convergent or discriminant validity of an IAT (Lane et al., 2007; Nosek, Greenwald, et al., 2007).

Last but not least, the selected self-report measures should have demonstrated consistently acceptable reliability and validity across studies. Though very few self-reports have been developed to target stereotypes of empathy in scientists in particular to date, it is sensible to select empathy-related items from existing self-report measures for general stereotypes of scientists to be adapted to the present study. As for the relative self-report, a paradigm that has been widely used to develop self-report questions with similar structure to the corresponding IAT is applied (e.g., Fleischhauer,et al., 2013; Greenwald et al., 2003). The following sections provide readers with more detailed information about how each measure was selected and adapted. The items of each measure are presented in the appendices for reference. Table 5.6 illustrates the implicit and explicit measures applied in the present study.

Table 5.6 Explicit and implicit measures applied in Stage 2

Explicit measures	**Implicit measure**
• Relative self-report: Science - Empathy Explicit Scale (SE-Explicit) • Absolute self-report: Interpersonal Subscale of the Stereotype of Scientists (ISSOS) questionnaire	• Science Empathy - Implicit Association Test (SE-IAT)

5.2.4.1. Interpersonal Subscale of the Stereotype of Scientists questionnaire (ISSOS)

As discussed in the literature review, stereotypes about scientists have been measured in different ways in different context. The criteria for selecting the absolute explicit measure in this study are that the measure should: 1) be suitable to capture stereotypes among adults in the UK context; 2) assess up-to-date stereotypes of scientists and 3) target corresponding constructs with the SE-IAT, that is the stereotype about empathy in scientists. Existing explicit measures for stereotypes of scientists were examined carefully against the three criteria: The Draw-a-Scientist Test (DAST) (Chambers, 1983) has often been used among children to provide qualitative information about the image of scientists; the Women in Science Scale (WiSS) (Erb & Smith, 1984) targets only perceptions about female scientists; and The Image of Scientists Scale (ISSS)

(Krajkovich & Smith, 1982) has often been used among middle school students and to measure general negative and positive images of scientists, so that all the three measures were deemed either age-inappropriate for college students or measuring non-corresponding content with the SE-IAT, therefore were dropped from the present study. Finally, the Interpersonal Subscale of the Stereotype of Scientists (SOS) questionnaire (Schneider, 2010) is selected. Table 5.7 summarises the existing explicit measures for stereotypes of scientists.

Table 5.7 Summary of existing explicit measures for stereotypes of scientists

Name	Time	Participants	Content
DAST	1983	Children	Qualitative information about the image of scientist (e.g., gender, glasses, lab coat, facial hair, products of research)
WiSS	1984	Adolescents	Attitudes about 1) Women's scientific ability 2) Sexism
ISSS	1982	Adolescents	Attitudes about 1) Positive image of scientists 2) Negative image of scientists 3) Science avocation
SOS	2010	College students	Attitudes about scientists' 1) Professional competence 2) Interpersonal competence

Note: DAST, Draw-a-Scientist Test; WiSS, Women in Science Scale; ISSS, Image of Science and Scientists Scale; SOS, Stereotypes of Scientists Scale; K-12, publicly supported school grades from kindergarten to the 12th grade.

The SOS questionnaire was first developed in 2010 to explore the contemporary images of scientists among college students in the US (Schneider, 2010; Wyer et al., 2010). The questionnaire was designed to identify the characteristics and content of contemporary college students' image of scientists, both what they "do" in their daily work life and who they "are" as people. The Confirmatory Factor Analysis (SEM) on SOS conducted by Schneider (2010) with 1061 US college students revealed two subscales: one indicates attitudes toward scientists' professional competence (e.g., whether scientists are perceived as technically competent and logical), and the other indicates attitudes toward scientists' interpersonal competence (e.g., whether scientists are perceived as cooperative and able to maintain good friendships). For the US sample, consistently high reliability of the two subscales were reported (Cronbach α = .81 for the professional competence subscale and Cronbach α =.77 for the Interpersonal competence subscale).

Considering the three criteria for the explicit measures, the Interpersonal Subscale of the Stereotypes of Scientists (ISSOS) is selected as the psychometrically sound absolute explicit measure suitable for assessing up-to-date stereotypes about empathy in scientists among adults in the UK. Despite the fact that empathy is not an identical concept with social competence, numerous studies have found systematically strong positive correlations between empathy and social competence, suggesting an important role for empathy in facilitating social interactions (Allemand et al., 2015; Eisenberg, 2002). For example, adolescents who reported higher empathy also reported more pro-social behaviours, were less aggressive, and had more supportive peer relationships (Eisenberg et al., 2009). Moreover, adults who reported higher empathy were more willing to volunteer (Davis et al., 2001), more grateful to others (Mccullough, Emmons, & Tsang, 2002), donated more to charity (Wilhelm & Bekkers, 2010), and were more likely to forgive others (McCullough, Worthington, & Rachal, 1997). Supported by these findings, measuring explicit attitudes toward social competence of scientists is considered an alternative way to assess explicit stereotypes of empathy in scientists in the present study.

The ISSOS is comprised of 9 items tapping attitudes toward various aspects of scientists' social life. Participants are required to rate on a 7-point Likert scale from *"strongly disagree"* to *"strongly agree"* on statements describing either a positive or a negative view about scientists' interpersonal relationships, such as *"Scientists have fun with colleagues at work"* or *"scientists are out of touch with what is happening in the world"* (See Appendix VI for the full scale). The internal consistency of the ISSOS in the present study is acceptable (Cronbach's α = .72).

5.2.4.2. Science-Empathy Explicit Scale (SE-explicit)

Given that no adequate self-report questionnaire was available to measure the similar relative attitudes toward associations between *Science-Liberal arts* and *Empathy-Rationality* in the SE-IAT, a new scale was constructed using one of the most popular ways to measure relative explicit attitudes in the IAT literature: the combination of Feeling Thermometer ratings and Likert ratings (e.g., Blair, Judd, Havranek, & Steiner, 2010; Fleischhauer et al., 2013; Greenwald & Farnham, 2000; Karpinski & Hilton, 2001; Payne et al., 2008). In the Feeling Thermometer, participants express their feelings toward a given person, group or issue in terms of degree, with their attitudes corresponding to temperature. Ratings below the midpoint of the thermometer are considered as a "cold" attitude, which indicates a negative view of a person, group, or issue; conversely, ratings above the midpoint are considered as a "hot" attitude, which indicates a positive view of a person, group, or issue. The midpoint of the feeling thermometer is reserved to indicate a neutral attitude towards a person, group, or issue (Nelson, 2011). Feeling Thermometer is a very popular survey instrument in the stereotype field for it enables researchers to collect information about the direction, as well as intensity, of participants' attitudes toward a specific person, group or issue.

In the present study, the Feeling Thermometer method is used to capture whether participants hold stereotype-congruent or counter-stereotype views about the associations between *Science-Liberal arts* and *Empathy-Rationality* as well as how strong their attitudes are. Specifically, the Feeling Thermometer item was as follows:

Which statement best describes your view?

- I strongly associate *liberal arts* with *empathy* and *science* with *rationality*.
- I moderately associate *liberal arts* with *empathy* and *science* with *rationality*.
- I slightly associate *liberal arts* with *empathy* and *science* with *rationality*.
- I associate *empathy* and *rationality* with *science* and *liberal arts* equally.
- I slightly associate *science* with *empathy* and *liberal arts* with *rationality*.
- I moderately associate *science* with *empathy* and *liberal arts* with *rationality*.
- I strongly associate *science* with *empathy* and *liberal arts* with *rationality*.

Participants choosing the first three options were regarded to hold stereotype-congruent attitudes; conversely, participants choosing the last three options were regarded to hold counter-stereotype attitudes; and those who chose the middle option were regarded to hold a neutral attitude.

Using the Likert ratings, participants can express how much they agree or disagree with certain statements. In the present study, participants were asked to rate on a 7-point Likert scale from "strongly disagree" to "strongly agree" for two SE-IAT content-corresponding statements, including "Scientists are less empathetic than humanity majors" and "Scientists are more rational than humanity majors".

Combining the Feeling Thermometer item and the Likert items, the SE-explicit is comprised of 3 items in total. Though the 3-item scale using different kinds of questions may not have a good internal consistency, the SE-explicit is irreplaceable in the present study for its corresponding relative structure to the SE-IAT. As explained earlier, the distinctness in structures may partly account for the differences between the outcomes of implicit and explicit measures (Hofmann et al., 2005; Payne et al., 2008). In order to control for the impact of the lack of fit in structure, SE-explicit is designed to resemble the SE-IAT on purpose. Nevertheless, given that SE-IAT incorporates two groups of relative concepts, it is inevitable to create the so-called "double-barreled" question (i.e., a question asks about more than one construct; Olson, 2011) in the SE-explicit, which is usually regarded as an informal fallacy in a survey. Specifically speaking, in the Thermometer Feeling item, the statement "I associate *liberal arts* with *empathy* and *science* with *rationality*" asks two attitude targets (e.g., the attitude towards *empathy* in different majors as well attitude towards *rationality* in different majors) as one construct (e.g., Do you prefer liberal arts + *empathy* and *science* + *rationality* or *liberal arts* + *rationality* and *science* + *empathy*?). Although it is clear that such "double-barreled" question could violate the reliability and validity of the SE-explicit, considering the purpose of encompassing the relative explicit measure is to deal with the structural fit effect, the structural resemblance to the SE-IAT has taken priority over the psychometric soundness of the SE-explicit questionnaire itself. Given this, despite its poor internal consistency, the SE-explicit questionnaire is still treated as an appropriate relative explicit measure in the present study.

5.2.4.3. Implicit measure: Science – Empathy IAT (SE-IAT)

In previous chapter, the design of the updated version of the SE-IAT was described. It was created to capture participants' underlying beliefs about the associations between *science-liberal arts* and *empathy-rationality*. This newly developed SE-IAT is applied in the present study. See Section 4.3 for the detailed procedure and Table 4.7 for the updated items in the SE-IAT.

As discussed in the literature review, IAT has been the most widely used implicit measure in various fields due to its superior reliability over other implicit measures such as the evaluative priming procedure (e.g., Cunningham, et al., 2001), the Go/No-Go Association Task (GNAT; Nosek & Banaji, 2001) or the dot probe task (Schmukle, 2005). The internal consistencies of the IAT have been reported to be satisfactory (internal reliabilities averaged .79 across 50 studies in a meta-analysis conducted by Hofmann et al., 2005) ranging from .70 to .90 (Nosek, Greenwald, et al., 2007). In terms of its discriminant validity, a number of studies looking at various domains have reported somewhat low correlations between the IAT and explicit measures in a range from $r = -.25$ to $r = .60$ (Hofmann et al., 2005) or from $r = .13$ to $r = .75$, with median $r = .22$ (Lane et al., 2007). Only a handful of studies have shown strong correlations between the IAT and explicit measures (e.g., Asendorpf et al., 2002; Cunningham et al., 2001; Greenwald & Farnham, 2000; Wiers, Van Woerden, Smulders, & De Jong, 2002). See Section 3.3.2 in the Literature Review chapter for a summary about studies of the psychometric properties of the IAT. In the present study, the reliability and validity of the SE-IAT will be examined against a series of criteria as well as compared with previous findings.

5.2.5. Data preparation

In this section, the procedure of how to calculate the scores for each measure is explained in detail. The IAT scores were calculated using an improved algorithm to reduce the influence of construct-unrelated variance. Comparisons between the conventional algorithm and the improved algorithm for the IAT are presented. As for the self-report explicit measures, scores were calculated in accordance with the IAT results with positive scores representing stereotype-congruent attitudes (i.e., associate *science* with *rationality* and *liberal arts* with *empathy*) and negative scores representing counter-stereotype attitudes (i.e., associate *science* with *empathy* and *liberal arts* with *rationality*).

5.2.5.1. SE-IAT scoring

The so-called improved D scoring algorithm (Greenwald et al., 2003) has been selected to calculate the SE-IAT effect in the present study. Unlike the conventional algorithm, which only uses data from the correct trials in the critical testing blocks, the improved D scoring algorithm takes all trials in both the practice and testing blocks into account, therefore displaying better resistance to construct-unrelated variance such as order effect and cognitive fluencies (Greenwald et al., 2003). Table 5.8 compared the steps of the conventional algorithm and the improved D scoring algorithm in detail.

Table 5.8 Conventional and improved IAT scoring algorithms compared

Step	Conventional Algorithm	Improved D Algorithm
1	Use data from B4 & B7	Use data from B3, B4, B6, & B7
2	Nonsystematic elimination of subjects for excessively slow responding and/or high error rates	Eliminate trials with latencies > 10,000 ms; eliminate subjects for whom more than 10% of trials have latency less than 300 ms
3	Drop first two trials of each block	Use all trials
4	Recode latencies outside 300/3000 ms boundaries to the nearer boundary value	No extreme value treatment (beyond Step 2)
5		Compute mean of correct latencies for each block
6		Compute one pooled SD for all trials in B3 & B6; another for B4 & B7
7		Replace each error latency with block mean (computed in Step 5) + 600 ms
8	Log-transform the resulting values	No transformation
9	Average the resulting values for each of the two blocks	Average the resulting values for each of the four blocks

Step	Conventional Algorithm	Improved D Algorithm
10	Compute the difference B7-B4	Compute two differences: B6-B3 and B7-B4
11		Divide each difference by its associated pooled-trials SD from Step 6
12		Average the two quotients from Step 11

Note. Block numbers (e.g., B1) refer to the procedure shown in Table xxx. The conventional algorithm has no procedures corresponding to Steps 5 – 7 or Steps 11 – 12 of the improved D algorithm. SD = standard deviation. Adapted from Greenwald et al., (2003).

For both algorithms, the basic logic for calculating the IAT effect is to compare the reaction times between the congruent pairing blocks and the incongruent pairing blocks. The conventional scoring algorithm (Greenwald et al., 1998) log-transforms the response latencies as well as recodes latencies greater than 3000 ms or smaller than 300 ms to 3000ms and 300ms respectively. The conventional scoring algorithm has been criticized for (1) being contaminated by extraneous variables such as cognitive fluency, thus inflating the IAT scores for slow responders (Cai, Sriram, Greenwald, & McFarland, 2004; Hilgard, Bartholow, Dickter, & Blanton, 2014), (2) requiring a large sample size to yield power of .80 to reject the null hypothesis for studies looking at the correlations between IAT and other explicit measures (Greenwald et al., 2003), (3) the reduction in magnitudes of IAT scores with repeated administrations (Greenwald & Nosek, 2001), and (4) low internal consistency (Greenwald et al., 2003).

In response to such criticisms, Greenwald et al. (2003) developed a new set of scoring algorithm, the so-called D scoring algorithm that has three substantial changes from the conventional procedure: (1) use of practice-block data (Step 1 in Table 5.6), (2) use of error penalties (computed in Steps 5 and 7), and (3) use of individual-respondent standard deviations to provide the measure's scale unit (Computed in Step 6 and applied in Step 11). The D scoring algorithm therefore displays better internal consistency, and better resistance to a number of extraneous procedural influences such as order effects and cognitive fluencies than the conventional scoring method (Cai et al., 2004; Greenwald et al., 2003; Hilgard et al., 2014). The D scoring algorithm has become the most commonly used scoring method for the IAT and is used in the present study for the SE-IAT as well. The SPSS syntax for computing IAT effects using the improved D algorithm can be obtained from http://faculty.washington.edu/agg/iat_materials.htm.

For the SE-IAT, a positive D score represents a stereotype-congruent implicit belief that *science* is associated with *rationality* and *liberal arts* with *empathy*, but a negative D score represents a counter-stereotype implicit belief that *science* is associated with *empathy* and *liberal arts* with *rationality*. The value of the D score shows the strength of the associations. The bigger the value of the D score is, the stronger the associations between the concepts are. The D score ranges from -2 to +2. The strength of the IAT effect can be described as 'strong' (.80), 'medium' (.50), 'slight' (.20), or 'little or no' when interpreting the D scores. These cut-offs are in correspondence to results meeting conventional criteria for small, medium, and large effect size of Cohen's *d* (1988) measure (Nosek et al., 2005).

5.2.5.2. SE-explicit scoring

For the SE-explicit questionnaire, responses were collected on 7-point Likert scales ranging from "strongly disagree" (coded as -3) to "strongly agree" (coded as +3). The middle point "neither agree nor disagree" was coded zero. Consistent with the SE-IAT scores, stereotype-congruent responses associating science with rationality and liberal arts with empathy were coded with positive integers while counter-stereotype responses associating science with empathy and liberal arts with rationality were coded with negative integers. The sum of the scores of the three questions in the scale was regarded as the final score of the SE-explicit, ranging from -9 to +9.

5.2.5.3. ISSOS scoring

For the ISSOS scale, responses were also collected on 7-point Likert scales ranging from "strongly disagree" (coded as -3) to "strongly agree" (coded +3). The middle point "neither agree nor disagree" was coded zero. Negatively worded items were reverse scored. Positive ISSOS scores thus represented counter-stereotype responses (i.e. that scientists have good social skills) and the negative ISSOS scores represented stereotype-congruent responses (i.e. that scientists do not have good social skills). The sum of the scores of the nine items in the questionnaire was regarded as the final score of the ISSOS, ranging from -27 to +27. However, in order to make the valence of scores consistent with the other two measures, we then reversed the total scores such that counter-stereotype responses were reflected by negative ISSOS scores and the stereotype-congruent responses were reflected by positive ISSOS scores.

5.2.6. Data analysis

In order to preserve credibility of the results, rigorous procedures of data analysis were followed in the present study. For the statistical tests of the present study, sources of bias (e.g., missing data and outliers) were treated using techniques justified with reference to a variety of literatures. Assumptions of tests (e.g., normality and homogeneity of variance) were also checked before using any parametric test with the data (e.g., t-test and correlation; Field, 2013).

5.2.6.1. Missing data

Firstly, Missing Value Analysis (MVA) in SPSS was performed with the dataset. The results suggest that in the present dataset 7% of the cases have missing values in academic identification, meaning that 7% of participants did not report their identification with science or liberal arts. In this case, independent t-tests were conducted to see if there are significant differences between those who have reported their academic identification and those who hadn't in terms of their SE-IAT, SE-explicit and ISSOS performances. No significant difference was found on any of the three measures, indicating that the patterns of missing values do not depend on the data values. The missing values can therefore be deemed as missing at random (Little & Rubin, 2002). Moreover, as for other variables (e.g., age, gender, individual item in the three measures), the MVA results suggest less than 5% of the cases have missing values and are missing at random as well.

According to Field (2013), there are several methods to deal with missing data under this condition, such as case deletion, mean imputation, and regression imputation. Each method has its advantages and disadvantages. Case deletion, especially listwise deletion, is one of the most commonly used methods, which can avoid the drawback of introducing bias to the data. However, it inevitably reduces the sample size, which results in the decreased power of analysis (Field, 2013). In contrast, imputations retain the original sample size by replacing missing data with values generated in different ways. For example, mean imputation uses the variable mean of all other cases to replace the missing value. However, this poses a potential problem of attenuating correlations between variables. Given that the present study proposes to examine the implicit-explicit correlations, this method is deemed inappropriate. Regression imputation, on the other hand, makes up the pitfall of mean imputation by using a predicted regression model to generate possible values for the missing data. Yet, this may mask the uncertainty of imputed values and lead to over identified relationships between variables (Field, 2013). In this case, the present study adopted the simpler and more straightforward way of handling missing data by deleting the cases with missing values. Besides, as the sample size of the present study is relatively large (= 485) and the number of cases with missing values is very small (< 7%) the effect on sample size is acceptable. Therefore, listwise deletion was adopted in the present study as the technique to handle missing data.

5.2.6.2. Outliers

One thing that usually affects normality is outliers. An outlier is a score very different from the rest of the data (Field, 2013). To spot the outliers, z-scores were calculated for each measure and cases with z-scores greater than 3.29 in absolute value were regarded as potential outliers (Field, 2013). No outlier was spotted for the SE-IAT and SE-explicit results and only 3 outliers were found for the ISSOS results. In terms of how to deal with outliers, one can either delete the case or transform data values. Given that there were only a very small number of outliers identified in the present study, these cases were removed from the dataset on the grounds that they might have represented rather idiosyncratic situations.

5.2.6.3. Normality

Before using any parametric tests with the data (e.g., t-test and correlation), the assumption of normality is checked by examining the Skew and Kurtosis of the data. The Kolmogorov-Smirnov test or Shapiro-Wilk test is not used because in a large sample of the present study (n = 485) these tests can be significant even when the scores are only slightly different from a normal distribution (Field, 2013). Instead, Histograms, P-P plots and the values of skew and kurtosis are examined. Data for SE-IAT, SE-explicit and ISSOS are all normally distributed with skew and kurtosis values very close to 0.

5.2.6.4. Homogeneity of variance

Another important assumption for parametric tests is the homogeneity of variance because unequal variances could create bias and inconsistency in the estimate of the standard error associated with the parameter estimates in linear models (Hayes & Cai, 2007). Statisticians used to recommend testing for homogeneity of variance using Levene's test, if the assumption was violated, using an adjustment to correct it (Field, 2013). However, like the aforementioned problem with significant tests of normality, Levene's test can also be significant even for small and unimportant effects in large samples. As suggested by Field (2013), in a large sample like the present study, the assumption of homogeneity of variance is pretty much irrelevant, and can be ignored (Field, 2013).

5.2.6.5. Criteria for evaluating the SE-IAT

As discussed in the literature review, controversial evidence has been found for the psychometric properties of different IATs. Before applying the newly developed SE-IAT to investigate implicit stereotypes about empathy in scientists, it is important to assess whether the SE-IAT met relevant psychometric criteria. Selected criteria, hypotheses and applied statistical analyses to evaluate the reliability and validity of the SE-IAT are displayed in Table 5.9 below.

Table 5.9 Summary of the criteria, hypotheses and statistical analyses to evaluate the SE-IAT

Criteria	Hypotheses	Statistical analyses
Replicating the IAT effect	H1: Both the reaction time (H1a) and errors (H1b) should be increased for the incompatible task as compared to the compatible task due to the theory that the incompatible task requires more cognitive capacities.	Paired sample T-test
Internal consistency	H2: SE-IAT is expected to show similar internal consistency with other IATs ranging from .70 to .90.	Split-half reliability
Relationship with explicit measures	H3: SE-IAT is expected to show little or no correlation with SE-explicit (H3a) or ISSOS	Pearson's correlation

Criteria	Hypotheses	Statistical analyses
	(H3b). However, if correlated, the correlation is expected to be stronger for SE-explicit than ISSOS because the SE-explicit has better *structural fit* with the SE-IAT than ISSOS (H3c).	
Ability to capture individual differences	H4: Both gender and major are expected to have influence on the SE-IAT performance. Women are expected to have weaker SE-IAT effect than men due to their advantage in social sensitivity (H4a). Students identified with sciences are expected to have weaker SE-IAT effect than liberal arts due to the *ingroup favouritism* (H4b). If there is an interaction between gender and major, individuals with unconventional identities (women in science and men in liberal arts) are expected to have weaker SE-IAT effect than those with conventional identities (women in liberal arts and men in science) due to the *role incongruity theory* (H4c).	Two-way ANOVA
Resistance to order effect	H5: No difference is expected in SE-IAT results between participants who did the incompatible task first and the compatible task later and those who completed the test in a reversed order.	Independent t-test
Resistance to prior IAT experience effect	H6: No difference is expected in their SE-IAT results between participants who had done an IAT before and those who had no prior experience with IAT.	Independent t-test

5.3. Results

5.3.1. Descriptive statistics

First of all, descriptive data of the involved variables are presented so the readers can gain general knowledge of the status of the current sample on each variable. Table 5.10 illustrates the mean (M), standard deviation (SD) and the range of current participants' SE-IAT, SE-explicit and ISSOS scores.

Table 5.10 Descriptive data of the SE-IAT, SE-explicit and ISSOS results

				Range	
Variable	n	M	SD	Potential	Actual
SE-IAT	485	0.60	0.36	-2 − 2	-.79 − 1.51
SE-explicit	475	2.26	2.87	-9 − 9	-6.0 − 9.0
ISSOS	476	-10.34	6.49	-27 − 27	-26 − 14

Note. The variation in sample size is due to the variation in the number of participants who completed each measure. Negative scores reflect stereotype-congruent views and positive scores reflect counter-stereotype views.

The data showed that the average SE-IAT score was positive (M = .60, within the medium effect range of .50 to .80), meaning that current participants held a moderate stereotype-congruent implicit mental association of *science – rationality* and *liberal arts – empathy*. For the explicit stereotypes, the data showed that the average SE-explicit score was also positive (M = 2.26, within the medium effect range of 2.25 to 4.5), meaning that current participants explicitly endorsed moderate stereotype-congruent associations of *science – rationality* and *liberal arts – empathy*. On the contrary, the data showed that the average ISSOS score was negative (M = -10.34, within the slight effect range of 6.75 to 13.5), meaning that current participants explicitly held a weak counter-stereotype view about scientists' interpersonal skills. In other words, current participants reported slight positive attitudes about scientists' interpersonal skills, but they still believed scientists were less empathetic but more rational than humanists, both implicitly and explicitly.

5.3.2. Replicating the IAT effect

As explained in the literature review, the incompatible combined blocks of the IAT are supposed to draw more heavily on cognitive capacities than the compatible combined blocks and as such cause higher response latencies (i.e., reaction time and errors) than compatible combined blocks. Hypothesis 1a predicted that participants would take longer to complete the incompatible task (i.e., *science – empathy* and *liberal arts - rationality*) than compatible task (i.e., *science – rationality* and *liberal arts – empathy*) of the SE-IAT. The results of a two-tailed paired sample t-test supported this hypothesis, ($t(484) = 28.57$, ***$p < .001$, Cohen's $d = .46$, marking a small effect size.) Reaction time was significantly increased in the incompatible combined blocks (M = 1205.12 (ms)) as compared to the compatible combined blocks (M = 897.67(ms)). Figure 5.1 depicts the means and standard errors for reaction time of the compatible and incompatible combined blocks.

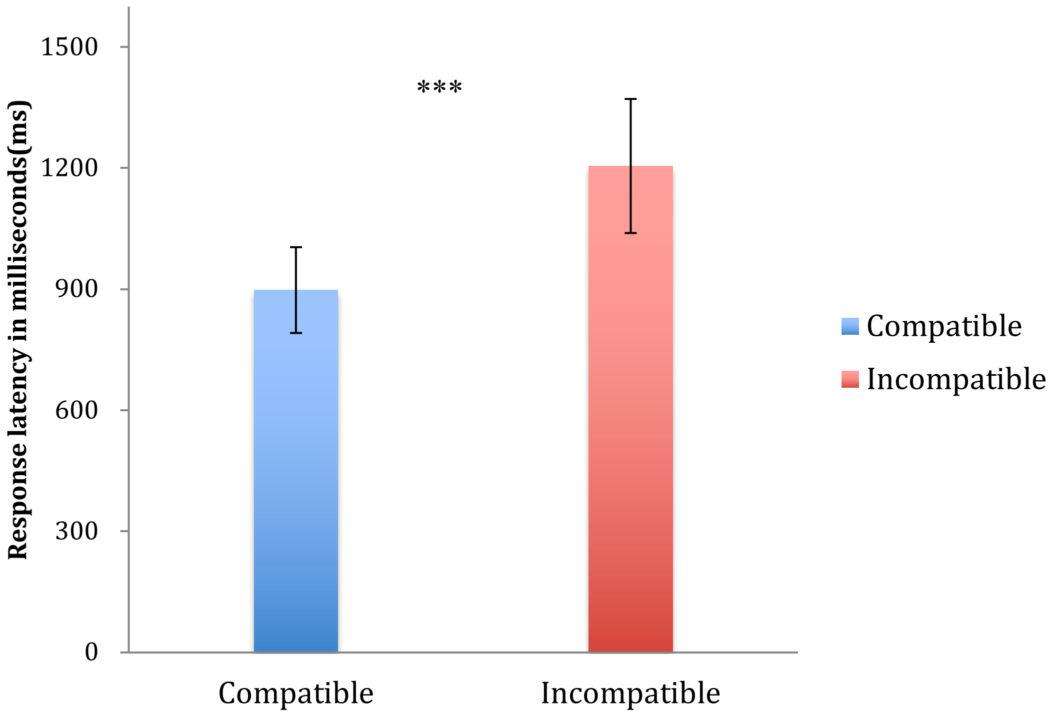


Figure 5.1 Mean reaction times of the compatible and incompatible SE-IAT blocks

Error bars are standard error of the mean. Compatible task is to associate science with rationality and liberal arts with empathy. Incompatible task is to associate science with empathy and liberal arts with rationality.

Hypothesis 1B predicted that the participants would make more mistakes (i.e., failed to sort the item into the affiliated category) in the incompatible combined blocks than the compatible combined blocks. Results of a two-tailed paired sample t-test also supported this hypothesis, ($t(484) = 9.85$, ***$p < .001$, Cohen's $d = .53$, marking a medium effect size.) Error rate was significantly increased in the incompatible combined blocks (M = 10.34 (%)) as compared to the compatible combined blocks (M = 7.33 (%)). Figure 5.1 depicts the means and standard errors for error rates of the compatible and incompatible combined blocks.

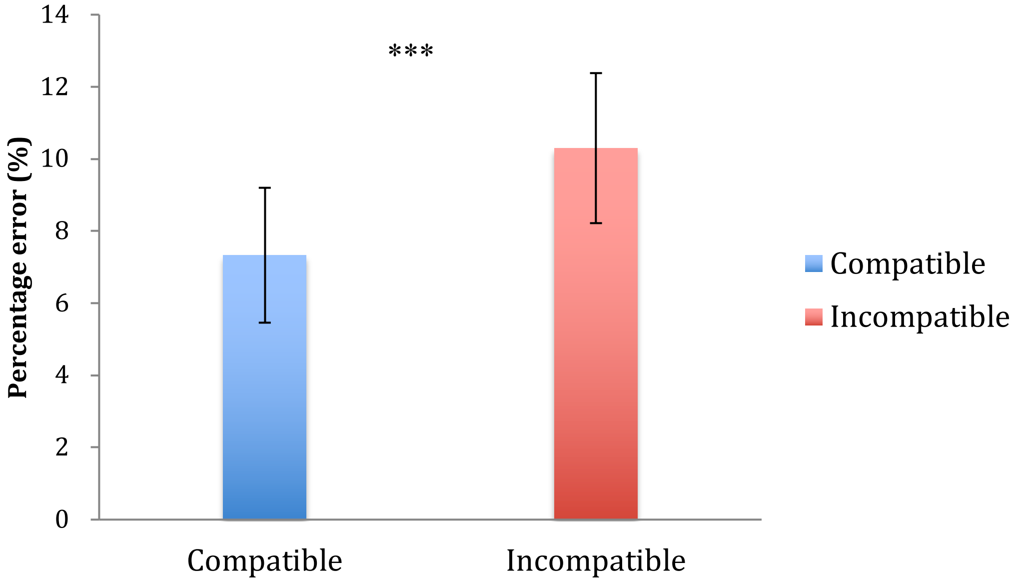


Figure 5.2 Mean error rates of the compatible and incompatible SE-IAT blocks

Error bars are standard error of the mean. Compatible task is to associate science with rationality and liberal arts with empathy. Incompatible task is to associate science with empathy and liberal arts with rationality.

5.3.3. Internal consistency

The internal consistency is examined using the split-half reliability method. To calculate the SE-IAT's split-half reliability, the D scoring algorithm was applied to two mutually exclusive subsets of the critical trials, using an odd-even divide. Hypothesis 2 predicted that the SE-IAT would display an internal consistency ranging from .70 to .90 according to a meta-analysis of existing IATs (Nosek, Greenwald, et al., 2007). However, the result of the Pearson's correlation between the two sub tests did not support this hypothesis ($r(485) = .44$, $*p < .05$). Although the two subtests were significantly correlated, the coefficient is smaller than the expected .70, meaning that the SE-IAT has weak internal consistency.

5.3.4. Relationship with explicit measures

Hypothesis 3 predicted that SE-IAT would show no or low correlations with SE-explicit (H3a) and ISSOS (H3b). The results of two Pearson's correlations supported these two predictions. There was neither a significant relationship between SE-IAT and SE-explicit ($r(475) = .087$, $p = .06$), nor a significant relationship between SE-IAT and ISSOS ($r(474) = .039$, $p = .395$).

Moreover, Hypothesis 3c also predicted that SE-IAT would show stronger correlation with SE-explicit than with the ISSOS (H3c). However, given that no significant correlation was detected, it was deemed pointless to compare the correlation coefficients.

5.3.5. Ability to capture individual differences

Hypothesis 4 predicted that both gender and major would have an influence on the SE-IAT performance. Women were expected to show a weaker SE-IAT effect than men due to their advantage in social sensitivity (H4a). Students identified with science were expected to have weaker SE-IAT effect than those identified with liberal arts due to the ingroup favouritism (H4b). However, the results of the 2 (female vs. male) * 2 (science vs. liberal arts) ANOVAs did not support these two predictions. There was neither a significant main effect of gender, $F(1, 450) = .086$, $p = .77$, $\omega^2 = .000$ nor a significant main effect of major, $F(1, 450) = 3.03$, $p = .082$, $\omega^2 = .000$. There was no significant interaction between gender and major either, $F(1, 450) = .054$, $p = .816$, $\omega^2 = .000$.

5.3.6. Resistance to undesired influence of order of combined tasks

Hypothesis 5 predicted no difference in their SE-IAT results between participants who did the incompatible task first and the compatible task later and those who completed the test in a reversed order. The results of an independent t-test support this prediction.

On average, participants who completed the incompatible task first and the compatible task later showed slightly greater SE-IAT scores (M = .64, SE = .023), than those who completed the compatible task first and incompatible task later (M = .56, SE = .023). This difference, .08, BCa 95% *CI* [0.0105, 0.1366], was not significant ($t(483) = .089$, $p = .766$), and it represented only a small-sized effect, Cohen's $d = 0.208$.

5.3.7. Resistance to undesired influence of prior IAT experience

Hypothesis 6 also predicted no difference in their SE-IAT results between participants who had done an IAT before and those who had had no prior experience with any IAT. The results of an independent t-test also support this prediction. On average, participants who had no prior IAT experience showed slightly greater SE-IAT scores (M = .60, SE = .018), than those who had (M = .57, SE = .040). This difference, .03, BCa 95% *CI* [-0.1131, 0.0502], was not significant ($t(474) = .489$, $p = .485$), and it represented a trivial effect, Cohen's $d = 0.09$.

5.4. Discussion

The aim of this chapter was to test the newly developed SE-IAT against certain psychometric criteria to see if it would be eligible for investigating individual differences in stereotypes of empathy in scientists in the next stage. Specifically speaking, SE-IAT was tested for its 1) ability to replicate an IAT effect, 2) internal consistency, 3) relationship with explicit measures, 4) ability to capture individual differences, 5) resistance to an undesirable order effect, and 6) resistance to prior experience of IAT.

Results showed that the SE-IAT successfully replicated an IAT effect (i.e., significantly greater reaction time and errors in incompatible blocks than in compatible blocks), captured implicit attitudes that were different from the explicit attitudes assessed by self-report measures, and resisted construct-unrelated variance such as order effect and prior IAT experience. However, results also showed that the SE-IAT displayed low internal consistency and had failed to detect any gender or major difference in implicit attitudes. Some improvements to the test's internal consistency, in particular, would be desirable in the final version.

In this section, the results of the current study will be synthesized and discussed with reference to the wider literature. Problems with the SE-IAT and limitations of the current study are identified to inform the future directions of research in this field. Finally, on the basis of the identified problems, corresponding modifications are made on the SE-IAT and lead to a single-category SE-IAT (SSE-IAT) before applying it on the next stage to investigate individual differences.

5.4.1. Validated aspects of the SE-IAT

The present study found significant differences in reaction time and error rates between compatible and incompatible blocks of the SE-IAT, with the incompatible blocks resulting in significantly greater reaction time and error rates than the compatible blocks. Hypothesis 1 has been successfully confirmed and in line with previous research which states that incompatible blocks draw more heavily on cognitive capacities and as such result in higher response latencies than the compatible blocks (e.g., Conrey, Sherman, Gawronski, Hugenberg & Groom, 2005; Klauer, Schmitz, Teige-Mocigemba, & Voss, 2010; McFarland & Crouch, 2002; Mierke & Klauer, 2003). As depicted in Table 5.9, the positive SE-IAT effect showed that there was a significant stereotype-congruent implicit preference for liberal arts over scientists in terms of their empathy among participants in the present study. As such, the SE-IAT has shown its ability to replicate an IAT effect and reflect implicit associations between concepts.

Furthermore, confirming Hypothesis 3, results showed non-significant low correlations between the SE-IAT and explicit measures (*r* = .087 with SE-explicit and *r* = .039 with ISSOS), indicating that the SE-IAT reflected implicit automatic mental representations that were different from attitudes reported in a rather controlled or reflective manner in explicit measures (Back, Jordan, & Thomas, 2009; Lane et al., 2007). Such findings are in accordance with a previous meta-analysis that revealed a somewhat low correlation between IATs and explicit measures with an average of *r* = .19 (range: *r* = -.25 to *r* = .60; Hofmann et al., 2005). However, it is also worth mentioning that although both the SE-explicit and ISSOS showed non-significant low correlations with SE-IAT, SE-explicit (M = 2.26, out of 9) yielded the same positive mean score with SE-IAT (M = 0.60, out of 2) while ISSOS yielded a negative mean score in the opposite direction (M = -10.34, out of 27). The directional convergence between SE-IAT and SE-explicit, and the directional divergence from ISSOS, supports the aforementioned *structural fit* account for the implicit-explicit correlations (Payne et al., 2008). As discussed before, SE-explicit was seen to have a better *structural fit* with SE-IAT than the ISSOS. This is because both SE-explicit and SE-IAT measure relative attitudes based on the comparisons between science and liberal arts in terms of their empathy and rationality. In contrast, ISSOS involves only absolute propositional judgments focusing on social skills in scientists. Therefore, cautions must be stressed when interpreting implicit-explicit correlations as evidence for underlying cognitive processes. The low correlation between SE-IAT and ISSOS might therefore be partly indicative of structural misfit of the measures themselves in addition to the underlying cognitive processes (Hofmann et al., 2005; Payne et al., 2008). Nevertheless, SE-IAT still showed low correlation with SE-explicit regardless of the similarity in their task demands.

As regard to construct-unrelated variance, SE-IAT showed good resistance to both order effect and the influence of prior experience with other IATs. One of the most commonly observed types of construct-unrelated variance in the IAT is that participants are often slower in the subsequent combined blocks than in the preceding combined blocks, regardless of whether it is compatible or incompatible (Lane et al., 2007; Nosek, Greenwald, et al., 2007). However, such order effect was not observed in the present study, thereby confirming Hypothesis 5, and showing alignment with previous research suggesting that D scoring algorithm could reduce the block order effect (Greenwald et al., 2003). This finding also adds to the evidence that adding extra practice trials to the later combined blocks could effectively reduce the order effect of combined blocks in IATs (Nosek et al., 2005). Several researchers have proposed that the order effect is due to the impact of cognitive inertia (i.e., the tendency for beliefs or set of beliefs to endure once formed; Klauer, 2005). Once the preceding task categorization rule has been activated in one's mind it maintains a heightened state of activation for substantial amounts of time, as such makes it difficult for participants to suppress it and switch to the opposite categorization in the second task (Baker, Friedman, & Leslie, 2010; Klauer, 2005; Klauer et al., 2010). In the present study, the number of practice trials in the precedent combined blocks doubled the amount of practice trials in the subsequent combined blocks, allowing more time for participants to switch the categorisation rules between tasks, thus successfully reduced the order effect.

Moreover, confirming Hypothesis 6, the influence of prior IAT experience was not significant in the present study either. The concern about prior IAT experience comes from the possibility that subjects could attempt to exert cognitive control in order to either suppress or overcome their automatic associations if they know how IAT works in the first place (e.g., Asendorpf et al., 2002; Banse, Seise, & Zerbes, 2001; Steffens, 2004). It was known from previous analyses that prior experience with IAT could lead to reduction in IAT scores using conventional scoring algorithms (Greenwald & Nosek, 2001). Consistent with previous findings, using the improved D-scoring algorithm, the present study also found that participants who had prior IAT experience showed slightly smaller SE-IAT scores (M = .57) than those who had no IAT experience before (M = .60). But the difference was not statistically significant and can be ignored. Thus, the D-scoring algorithm has been shown to be an effective method to reduce sensitivity to prior IAT experience for SE-IAT.

To conclude, findings from the present study have validated several aspects of the SE-

IAT as an effective measurement for assessing implicit stereotypes about empathy in scientists. Firstly, it has successfully replicated an IAT effect thus showing the ability to detect implicit attitudes by measuring their underlying automatically activated associations between concepts. Moreover, the low correlations between SE-IAT and explicit measures even after controlling for the structural fit effect suggests that the SE-IAT indeed captures unconscious attitudes that are very likely to be independent from the self-reported beliefs gauged by explicit measures. Last but not least, construct-unrelated variances such as order effect and prior IAT experience were found to exert no significant influence on SE-IAT effect, thus cast little harm to the validity of SE-IAT.

5.4.2. Limitations of the SE-IAT and modifications

5.4.2.1. Limitations of the SE-IAT

Unfortunately, findings from the present study also suggest several problems with the SE-IAT. First of all, unlike most existing IATs which displayed satisfactory internal consistency ranging from .70 to .90 (Hofmann et al., 2005), the newly developed SE-IAT presented here showed an unacceptable internal consistency of .44, thereby violating Hypothesis 2. The poor internal consistency indicates that the items in the SE-IAT do not contribute equally to what is being measured and items that are unrelated to the target construct should either be removed or modified (Robson, 2002).

Furthermore, SE-IAT did not produce any individual difference as most existing IATs did (e.g., Gender-Science IAT, Race IAT and Age IAT, see the review paper by Nosek, Smyth, et al., 2007), therefore not supporting Hypothesis 4. It is important for SE-IAT to be able to capture variations in implicit mental representations for two reasons. Firstly, IAT needs to serve as more than a mirror of the culture (Lane et al., 2007; Nosek, Greenwald, et al., 2007). That is, it cannot be simply a tally of "the associations a person has been exposed to in his or her environment" (Karpinski & Hilton, 2001, p. 774). The lack of variations in the SE-IAT scores makes it plausible that the test may not reflect a person's own attitude toward empathy in scientists, but merely the popular belief residing in the culture. Secondly, the known-group approach to construct validity also argues that SE-IAT should reliably distinguish between members of different groups, based on a priori predictions or knowledge about those groups (Lane et al., 2007). Hypothesis 4a, that predicted women to have weaker SE-IAT effect than men due to their advantages in social sensitivity, was not supported. Hypothesis 4b, that predicted students identified with sciences to have weaker SE-IAT effect than those with liberal arts due to the ingroup favouritism (i.e., the tendency to favouring one's ingroups over out-group members; Tajfel & Turner; 2001), was also not supported. Thus, known-group validity for SE-IAT was not established. In this case, SE-IAT should not be used in its present form in the next stage, but instead needs modifications to be useful in evaluating individual differences in stereotypes of empathy in scientists.

These two psychometric issues with the SE-IAT may arise from a fundamental problem with the features of the SE-IAT. Notably, the attribute category "rationality" is not a natural counterpart to "empathy" like "female" to "male" or "good" to "bad" in other psychometrically sound IATs. A widely recognized limitation to the traditional IAT is the assessment of contrasting categories (Karpinski & Steinman, 2006; Lane et al., 2007; Nosek, Greenwald, et al., 2007; Schnabel, Asendorpf, & Greenwald, 2008). In the SE-IAT, it is unclear if an individual's score is due to a strong preference toward empathy in one group, or an aversion toward rationality in the other group, or a combination of both preference and aversion. Unlike "male" and "female" which are the two dimensions of one concept "gender", "rationality" and "empathy" are two independent concepts, therefore leading the SE-IAT to measure more than one construct. Instead of focusing on associations between *empathy* and *science*, it lumps the evaluations of associations of *science-empathy*, *science-rationality, liberal arts-empathy* and *liberal arts-rationality* altogether, thus creating much "noise" in the test, and potentially reducing its reliability and validity. As such, the unsatisfactory internal consistency could very likely be a result of the extraneous *rationality* items in the SE-IAT.

Furthermore, some native participants also reported having doubts about the use of *liberal arts* as the companion category label for *science*. They pointed out that "*humanities"* is the more commonly used word in the UK education system whereas "*liberal arts"* is more familiar with people in the US. Also, Fleischhauer et al. (2013) have found evidence that using more than one word as labels or stimuli in IATs may increase the complexity of processing, leading to unwanted sources of variance. Therefore, the use of the two-word category label *liberal arts* may also play a role in violating psychometric soundness of the SE-IAT.

To conclude, findings from the present study revealed two main psychometric problems with the SE-IAT, which are 1) the unsatisfactory internal consistency and 2) the lack of ability to capture individual differences in implicit cognition. As such, SE-IAT should be modified before being used in the next stage of the work. In the light of existing literature, two features of the SE-IAT were considered inappropriate that might have led to construct-unrelated variance. One is the artificially selected contrasting category *rationality* leading to ambiguity about the SE-IAT construct, and the other is the two-word category label "liberal arts" adding extra processing burden to participants when completing the test.

5.4.2.2. Modified Single-category SE-IAT (SSE-IAT)

To address the above-mentioned problems with SE-IAT, modifications were made leading to a Single-category SE- IAT (SSE-IAT; Karpinski & Steinman, 2006). Rather than presenting two pairs of contrasting concepts in the traditional IAT, a single-category IAT consists of only one pair of bipolar concepts and one unipolar concept. In a *single-target IAT*, there is 1 target (e.g., flower), and 2 attributes (pleasant - unpleasant), whereas in a *single-attribute IAT*, there are 2 targets (e.g., flower and insect) and 1 attribute (e.g., pleasant). By adopting the single-category IAT paradigm, the attribute category *rationality* and the items belonging to it were all dropped. The label *liberal arts* was replaced by *humanities*. The new modified Single-category SE-IAT is abbreviated as SSE-IAT for the convenience of interpretation in this thesis. It is worth mentioning that there are also other implicit procedures that may be adopted to assess unipolar concept. The detailed rationale for using the SSE-IAT will be explained in section 6.2.4.3 in the next stage. Table 5.11 presents the stimulus items in the new SSE-IAT.

Table 5.11 Stimulus items in the SSE-IAT

Science	**Humanities**	**Empathy**
Chemistry	Fine arts	Affectionate
Physics	Linguistics	Considerate
Mathematics	Philosophy	Caring
Engineering	History	Emotional
Computer science	Literature	Empathetic
Astronomy	Sociology	Feeling
Biology	Politics	Thoughtful
Geology	Music	Sensitive

As a variant to the traditional IAT, the procedure design of the new SSE-IAT is very similar to SE-IAT except dropping two blocks, resulting in a 5-block design. First, the *rationality - empathy* discrimination block (Block 2, see Table 4.4) in SE-IAT was removed. Given that SSE-IAT uses only *empathy* as the unipolar attribute concept, the discrimination between attribute concepts will no longer be needed. Second, the reversed target discrimination (Block 5, see Table 4.4) in SE-IAT was also removed. In the SSE-IAT, positions of target labels (i.e., *science* and *humanities*) remain the same, it is only the attribute label *empathy* needs to change position. As such, the reversed discrimination of target concepts will not be required either, and therefore should be removed. Table 5.12 presents the 5-block design of the SSE-IAT.

Table 5.12 Sequence of blocks in the SSE-IAT

			Response Key Assignment	
Block	N trials	Task	Left key (E)	Right key (I)
1	20	Target discrimination	Humanities	Science
2	40	Initial combined block	Humanities, Empathy	Science
3	50	Initial combined block	Humanities, Empathy	Science
4	40	Reversed combined block	Humanities	Science, Empathy
5	50	Reversed combined block	Humanities	Science, Empathy

It is worth noting that the number of trials in the critical combined blocks in the SSE-IAT has almost doubled the amount of trials in the corresponding blocks in the original SE-IAT to keep the length of the SSE-IAT the same with the standard IAT procedure (200 trials in total). It is important to do so because the full version of Single-category IAT has previously shown higher reliability than the short version of only 120 trials (Penke, Eichstaedt, & Asendorpf, 2006).

5.4.2.3. Limitations of the present study and suggestions for future research

The present study has a number of limitations. Firstly, the dichotomization of participants' major subjects when examining the individual differences in SE-IAT by academic discipline may lead to potential loss of study power. As discussed in the literature review, there are potential variations among different subjects under the general science label. For example, recent evidence has shown that students in life science could have comparable empathy with (Thomson et al., 2015) or even superior empathy over (Khorashad et al., 2015) those in social science, therefore they are very likely to differ from physics or engineering students in terms of the attitudes toward empathy in scientists. Future research that is interested in individual difference in stereotypes of scientists should address this issue and reveal what the variations are like inside the STEM world.

Secondly, participants from different academic fields were not truly representative of the university population in the UK. In general, female participants (60%) were overrepresented in the current sample as compared with the female proportion (49%) of university students from the national data (UCAS, 2017). Only 29% of humanities participants were men in the present study, which is lower than the national male proportion (39%) in humanities (UCAS, 2017). The overrepresentation of women in the current sample is in line with previous findings that university student research pools were usually overrepresented by women, students early in their studies and psychology majors (Barlow & Cromer, 2006), as they were often given the opportunity to self-select into particular research studies (Miller, 1981). As suggested by Dickinson, Adelson, and Owen (2012), knowing that male university students are often less willing to volunteer in psychology studies than females, researchers should make extra efforts to boost response rates among male students (e.g., additional reminders, targeted incentives) or seek other sources to recruit participants (e.g., another university) to achieve a sample that is representative of the target population.

Thirdly, the present study examined only one SE-IAT version, but it would be more rigorous to compare multiple versions of preliminary IATs using different category labels and stimulus that purport to measure the same construct to choose the final version. For example, when developing an IAT to assess individual differences in implicit cognitive motivation, Fleischhauer et al., (2013) developed and tested four preliminary IAT versions to validate the final one. As emphasized by many IAT experts, characteristics of selected category labels and stimulus materials would exert great influence on the IAT effects (Lane et al., 2007; Nosek, Greenwald, et al., 2007). For example, in the present study, *rationality* was selected as the counterpart to *empathy*. It would be interesting to see how things would work if we use *not empathy* as the counterpart to *empathy* as a different version of the SE-IAT. Moreover, in the last IAT development stage, we had changed all stimulus items representing *science* and *liberal arts* from professions (e.g., physicist, writers, mathematicians) to subjects (e.g., physics, literature, mathematics) simply because some participants reported having difficulty sorting certain professions into one discipline for they thought such professions could be both (e.g., some participants regarded mathematicians also as philosophers). But it would actually produce more informative findings by keeping and testing both versions. As such, it is suggested that future studies with more resources should conduct a more thorough investigation among different versions of potential IATs to illuminate their distinctiveness in psychometric properties and to determine the most appropriate version to fulfill the research purpose.

Last but not least, only a limited set of psychometric criteria was tested, leaving several other psychometric properties of the SE-IAT unexamined. Regarding measurement consistency, the test-retest reliability was not examined so it remains unknown if the SE-IAT could reliably replicate the result more than once in the same situation with the same population. Moreover, due to the limitations of time and resources of a PhD project, the IAT paradigm is the only applied method in the present study to assess the implicit stereotypes of empathy in scientists, therefore we did not check the convergent validity among SE-IAT and other implicit measures that purport to measure the same construct. More importantly, we did not test the predictive validity of the SE-IAT to see if it can predict any meaningful behaviour such as the academic performance or aspiration in science. For example, a meta-analysis of 184 independent samples (Greenwald et al., 2009) revealed that in studies that assess discrimination toward certain social groups, IATs did a superior job than explicit measures of prediction, but in studies that assess life style preferences (e.g., smoking, eating, and political candidate choices), explicit measures often outperformed IATs in prediction. Future studies may establish the psychometric soundness of a newly developed IAT by testing a full set of criteria including: 1) the ability to replicate the IAT effect; 2) Internal consistency; 3) test-retest reliability; 4) relationship with explicit measures; 5) relationship with implicit measures; 6) ability to predict meaningful behaviours; and 7) resistance to construct-unrelated variance (e.g., order effect, prior IAT experience, etc.).

5.4.3. Conclusion

In this stage, the newly developed SE-IAT was tested against certain psychometric criteria. It showed the ability to replicate an IAT effect, to capture implicit attitudes that were different from the explicit attitudes, and to resist construct-unrelated variance such as order effect and prior IAT experience. However, SE-IAT also showed poor internal consistency and was unable to detect any gender or major difference in implicit attitudes. Problems yielded by the original SE-IAT were analysed, leading to a modified Single-category SE-IAT (SSE-IAT) by dropping the *rationality* dimension. This modified SSE-IAT would be used in the next stage to explore individual differences in empathy in scientists.

Chapter 6 Stage Three: Individual differences of the implicit and explicit stereotypes of empathy in scientists and their relationships with career aspirations in science

6.1. Introduction

As the IAT to measure implicit stereotypes of empathy in scientists was established, tested and modified in the previous Stage 1 and 2, this new IAT as well as self-reported explicit measures are finally applied to investigate individual differences in stereotypes of emapathy in scientists in the present Stage 3. The three main research questions were investigated in Stage 3: 1) what are the implicit and explicit stereotypes of empathy in scientsits among UK university students nowadays; 2) What are the indivdiual differences in the implicit and explicit stereotypes of empathy in scientists by gender and major subejct; and 3) How do the implicit and explicit stereotypes of empathy relate to students' career aspirations in science?

6.2. Method

6.2.1. Sampling and participants

All participants were recruited through their departments or colleges from eight Russell Group universities located in England. As explained in the previous stage, the Russell Group universities were selected because, unlike many other universities with lower rankings that usually do not provide math-intensive subjects, Russell Group universities provide subjects in both STEM and non-STEM fields. Under such circumstances, it is important to bear in mind that the generalizability of the present study is limited.

The opt-in sampling method was adopted to recruit participants. A participant recruitment email containing a brief introduction to the project and the link to the new study website was spread out to all the students in the target departments or colleges by their department administrators or college student representatives from 22nd May to 29th November, 2017. It is worth mentioning that the four target universities from the previous stage were still included in the present study. However, given that this round of data collection was conducted almost a year after, it was the fresh students from some of the same target departments or colleges that were recruited in the present study. Gatekeepers from the previous target universities were asked to avoid sending the email to the same students last year. A reminder asking participants who had done the IAT test before not to retake the new test was also included in the study advertisement (See Appendix III).

Moreover, learning from previous experience, extra effort was made to recruit male participants in the present study. By checking the female to male proportions of each academic discipline from the national data, the precise number of women and men needed in the current sample from each academic field to achieve a similar gender ratio to the national data was calculated. The researcher had access to the daily raw data collected. Therefore, it was possible for the researcher to monitor the progress of data collection and identify the number of participants with particular characteristics in need during the process.

After three months of data collection, it was found that the recruited participants were overrepresented by women, as they were often more willing than men to volunteer in psychology studies (Dickinson et al., 2012). The sample was especially short of men from Engineering & Technology as well as Arts & Humanities fields. Therefore, additional advertisements (See Appendix IX) addressing men in particular were sent to the engineering and humanities departments from target universities. The male participant recruitment process lasted around two months. Once the desired number of participants was achieved, the data collection was finished by the end of November, 2017.

Among 3243 clicks on the study link, 1701 students continued to participate in the study.

However, 238 participants who did not finish the IAT test were removed from the final sample. Moreover, according to guidelines created by Greenwald, Nosek, & Banaji (2003), 15 participants who made too many errors in the IAT test, meaning that they sorted the items into the wrong categories too many times (error rates > 20%), were also dropped from the final sample. No participant completed the SSE-IAT too fast (reaction time in each trial less than 300 milliseconds > 10%) nor was any participant who completed the SSE-IAT too slow (reaction time in certain trial(s) in the SSE-IAT test greater than 10,000 ms) in the present study.

Consequently, the final sample is comprised of 1448 participants. The female to male ratio is 52.9% to 45.5% (766 females to 659 males, 23 participants who did not report their gender were still included, occupying 1.6%). Unlike the previous stage in which participants were asked to identify with only Science or Humanities, the current stage adopted a more detailed classification of participants' major subjects. Participants were identified with Arts & Humanities (n = 326, occupying 22.5%), Social Science & Management (n = 241, occupying 16.6%), Engineering & Technology (n = 263, occupying 18.2%), Life Science & Medicine (n = 252, occupying 17.4%), and Physical Science & Maths (n = 301, occupying 20.8%). There are 40 participants identified with other disciplines (occupying 2.7%) and 25 missing data (occupying 1.7%). The respective female to male ratio of participants identified with each discipline in the current sample as well as from national data is presented in Table 6.1. As can be seen from the table, the gender ratio of each academic discipline is not balanced. For example, the majority of participants from Engineering & Technology were male (77.9%) but the majority from Life Science & Medicine were female (76.2%) in present study. However, such imbalanced gender ratio of current sample is consistent with that from the national data. Therefore, the current sample is considered representative of the wider student population in UK higher education so that ensures the generalizability of the present study at least on the basis of gender.

Table 6.1 Female to male ratio in different academic fields of Stage 3 final sample and national higher education data

Academic Fields	Current Number			Current Gender Ratio		National Gender ratio	
	Total	Female	Male	Female	Male	Female	Male
Arts & Humanities	326	216	110	66.3%	33.7%	63%	37%
Social Science & Management	241	148	93	61.4%	38.6%	61%	39%
Life Science & Medicine	252	192	60	76.2%	23.8%	70%	31%
Engineering & Technology	263	58	205	22.1%	77.9%	19%	81%
Physical Science & Maths	299	126	173	42.1%	57.9%	40%	60%
Total	1381	740	641	53.6%	46.4%	57.5%	42.5%

Note. 40 participants identified with other academic disciplines and 27 participants who did not report their academic discipline were excluded. The national data is generated from the "Higher Education Student Statistics: UK, 2016/17 - Student numbers and characteristics" report by Higher Education Statistics Agency (2018).

However, given that we were interested in differences of stereotypes between participants majoring in the general science and humanities academic fields, the dichotomous groupings of major subjects were still applied as well. Participants identified with Engineering & Technology, Life Science & Medicine, and Physical Science & Maths were grouped as the *science majors*, while participants majoring in Arts & Humanities and Social Science & Management were grouped as the *humanities majors*. The science to humanities ratio is 56.4% to 39.2% (816 identified with sciences to 567 identified with humanities, the remaining 40 participants identified with other academic fields as well as 27 who did not report their identified academic field). The respective female to male ratio of participants identified with science and humanities majors is presented in Table 6.2. As can be seen from this table, the female to male ratio for humanities majors in the sample is slightly imbalanced with only 35.8% participants being male. The gender ratio is balanced for participants identified with science majors.

Table 6.2 Female to male ratio in science and humanities majors of Stage 3 final sample

Gender	Science majors (n = 814)		Humanities majors (n = 567)	
	N	Percentage	N	Percentage
Female	376	53.8%	364	64.2%
Male	438	46.2%	203	35.8%

Note: 25 participants identified with other academic fields and 9 participants did not report their academic field were excluded from the table.

The majority of participants (95.7%) were in the age range of 18 to 35 years old (See Table 6.3 for details). Also, in terms of education level, almost half of the participants were pursuing bachelor's degrees whereas the other half were postgraduates in the final sample (See Table 6.4 for details).

Table 6.3 Age range of Stage 3 final sample

Age range	N	Percentage
< 18	3	0.2%
18 - 20	427	30.1%
21 - 25	608	42.8%
26 - 30	240	16.9%
31 - 35	84	5.9%
36 - 45	32	2.3%
> 45	26	1.8%

Note: There are 28 missing values.

Table 6.4 Education level of Stage 3 final sample

Education level	N	Percentage
Bachelor's degree	614	44.8%
Master's degree	356	26.0%
Doctoral degree	399	29.2%

Note. 56 participants who were not currently university students and 6 who did not report their education level were excluded from the table.

In terms of ethnicity, participants were asked to choose their ethnic group from four options based on the classification of ethnicity by the Higher Education Statistics Agency (HESA, 2018). The majority of participants (72.2%) were White (See Table 6.5 for details). The proportions of different ethnic groups are inconsistent with national data (HESA, 2018). Finally, regarding the language ability, the majority of participants (70.0%) were native English speakers as shown in Table 6.6.

Table 6.5 Ethnic group of Stage 3 final sample

Ethnic group	N	Current percentage	National percentage
White	1022	72.2%	76.0%
Asian/Asian British	263	18.6%	10.3%
Black/African/Caribbean/Black British	18	1.3%	6.9%
Mixed/Multiple/Other	113	7.9%	5.1%

Note. There are 32 missing values

Table 6.6 English level of Stage 3 final sample

English level	N	Percentage
Native English speaker	996	70.0%
Non-native speaker	427	30.0%

Note. There are 25 missing values

6.2.2. Procedure

Similar to the previous stage, all tests and questions were self-administered by participants online with their own computers. See previous Section 5.2.2 for detailed steps.

Below is the link to the Stage 3 study website (See Appendix X for screenshots of the SSE-IAT procedures):

https://app-prod-03.implicit.harvard.edu/implicit/Launch?study=/user/nlofaro/contract.qin.empathy2/manager.expt.xml&refresh=true

6.2.3. Ethical considerations

Ethical approval has been obtained from the Ethics Committee of the Faculty of Education, University of Cambridge. Participants were asked to give consent by reading the instructions and clicking the "continue" button if they agree to proceed to the test and questions. All collected data were kept strictly confidential. See previous Section 5.2.3 for detailed steps addressing potential ethical issues.

6.2.4. Measures

To have a thorough investigation about stereotypes of empathy in scientists, both implicit and explicit measures were applied to examine the unconscious and conscious biases toward scientists' empathy and social skills. As discussed in the previous chapter, the Science Empathy-Implicit Association Test (SE-IAT) had been modified to the Single-category Science-Empathy Implicit Association Test (SSE-IAT) by dropping the *rationality* category to focus on the evaluation of empathy. This updated SSE-IAT was applied in the present study to see if it was able to capture individual differences in the implicit stereotypes of empathy in scientists.

In terms of selecting and adapting the self-report measures for explicit stereotypes of empathy in scientists, the same two questionnaires used in Stage 2 were applied. The only exception was that questions relating to *rationality* were dropped so that the Science-Empathy Explicit (SE-Explicit) scale has been adapted to the Single-category Science-Empathy Explicit (SSE-Explicit) scale as well. The absolute explicit stereotype was still measured by the Interpersonal Subscale of the Stereotype of Scientists (ISSOS) questionnaire. It is important to include both self-report measures for they have different structures and assess different aspects of explicit stereotypes of empathy in scientists (See Section 5.2.4 for detailed rationale). The SSE-Explicit has similar structure to the SSE-IAT and measures relative attitudes by assessing potential preferences between science and humanities in terms of their relations with empathy.

The ISSOS measures absolute propositional opinions toward the interpersonal skills in scientists. It is deemed necessary to include both relative and absolute self-report measures in the present study for the purpose of getting a fuller picture of stereotypes of empathy in scientists.

Moreover, a brief questionnaire assessing participants' science aspirations was also included. This questionnaire was included for the purpose of assessing the interplay between implicit and explicit stereotypes as factors in career choices. One of the most important hypotheses of the present study is that the stereotypical attitudes against empathy in scientists may prevent certain individuals, especially women who usually considered themselves of good empathy, from pursuing scientific careers. As such, it is considered of great interest to examine the relations between gender, stereotypes, and science aspirations. Table 6.7 illustrates the measures applied in Stage 3. Detailed information for each measurement is presented in the following sections.

Table 6.7 Applied measures in Stage 3

What to measure?	Applied measures
Implicit stereotypes of empathy in scientists	• Single-category Science-Empathy Implicit Association Test (SSE-IAT)
Explicit stereotypes of empathy in scientists	• Relative self-report: Science - Empathy Explicit Scale (SE-Explicit) • Absolute self-report: Interpersonal Subscale of the Stereotype of Scientists (ISSOS) questionnaire
Aspirations in science careers	• Career Aspiration in Science (SCAS) scale

6.2.4.1. Absolute explicit measure: Interpersonal Subscale of the Stereotype of Scientists questionnaire (ISSOS)

The ISSOS was applied again in the current stage for similar reasons presented in Section 5.2.4.1. It has been selected because: 1) it is suitable to capture stereotypes among adults in the UK context; 2) it assesses up-to-date stereotypes and 3) it targets interpersonal relationships in scientists in particular. Moreover, in previous Stage 2, the ISSOS displayed good internal consistency (Cronbach's $\alpha = .72$) and successfully captured explicit attitudes that were different from the implicit attitudes. Therefore, it is considered of good psychometric soundness.

The ISSOS is comprised of 9 items tapping attitudes toward various aspects of scientists' social life. Participants are required to rate on a 7-point Likert scale from "strongly disagree" to "strongly agree" on statements describing either a positive or a negative view about scientists' interpersonal relationships, such as "Scientists have fun with colleagues at work" or "scientists are out of touch with what is happening in the world" (See Appendix XI for the full scale). The internal consistency of the ISSOS items in current Stage 3 is acceptable (Cronbach's $\alpha = .75$).

6.2.4.2. Relative explicit measure: Single-category Science-Empathy Explicit scale (SSE-Explicit)

The SSE-Explicit was applied to measure the explicit relative attitudes corresponding to the implicit attitudes captured by the SSE-IAT. As mentioned earlier, the relative explicit measurement is usually developed in a way to imitate the structure of the IAT (See Section 5.2.5.2). In the current study, the SSE-IAT was developed to compare the strengths of automatic mental associations between *science-empathy* and *humanities-empathy*. In order to capture the similar relative attitudes explicitly, the SSE-Explicit directly asks participants to rate how strong they associate empathy with science and humanities.

The SSE-Explicit is comprised of 3 items. For the first two items, participants are required to rate how strongly they associate *empathy* with *science* and *empathy* with *humanities* on a 4-point Likert scale from "do not associate" to "strongly associate". Then they are also required to rate the statement "scientists are less empathetic than humanities majors" on a 7-point Likert scale from "strongly disagree" to "strongly agree".

Similar to the original SE-Explicit, which resembles the relative structure of SE-IAT, the SSE-Explicit also resembles the structure of the modified SSE-IAT. It is worth mentioning that, unlike the propositional items in the ISSOS that claim to address the same construct (i.e., social skills in scientists), items in the SSE-Explicit measures attitudes toward associations between different concepts (i.e., *empathy-science* or *empathy-humanities*) and together they measure a relative attitude (i.e., scientists have better empathy than humanities majors or humanities majors have better empathy than scientists). In other words, to resemble the structure of SSE-IAT, the SSE-Explicit actually measures the comparison between strengths of attitudes toward two concepts. It is of great importance to include this relative explicit measure for the purpose of controlling the construct-unrelated variance of *structural fit* (i.e., the degree of methodological similarity between different tests, Payne, Burkley, & Stokes, 2008) when examining the relationship between implicit and explicit stereotypes (see previous Section 5.2.4 for detailed explanation).

6.2.4.3. Implicit measure: Single-category Science-Empathy IAT (SSE-IAT)

In the previous chapter, the newly developed SE-IAT has been tested and modified. The modified Single-category Science-Empathy IAT (SSE-IAT) was applied in the present study to capture variations in implicit stereotypes of empathy in scientists. The original SE-IAT was not applied for evidence from Stage 2 indicating two main psychometric issues with it, including 1) poor internal consistency and 2) the lack of ability to detect individual differences in implicit cognition. These issues were deemed largely owed to the artificially selected contrasting category *rationality* in the SE-IAT that could lead to ambiguity about the measured construct (See Section 5.4.2 for details). As such, the category *rationality* was dropped, so that the SSE-IAT is consisted of only one attribute (i.e., *empathy*) and two target categories (i.e., *Science* and *Humanities*). Figure 6.1 below is a screenshot of the incompatible task of the SSE-IAT.

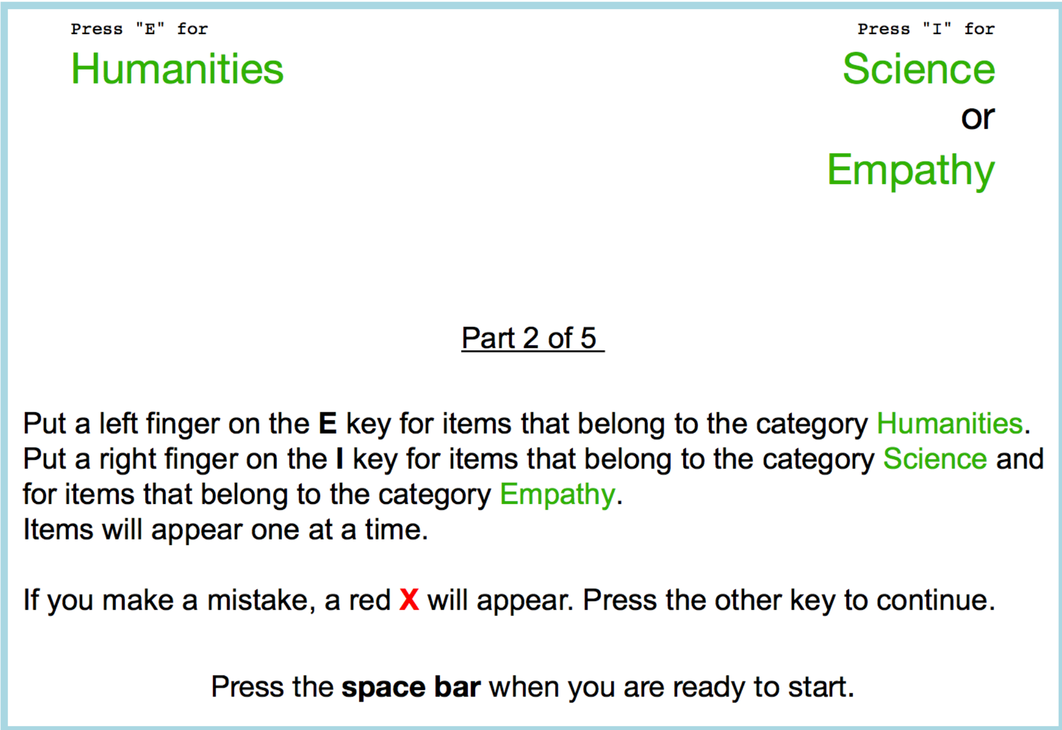


Figure 6.1 A screenshot of the incompatible task in the SSE-IAT

The Single-category IAT paradigm is utilised for three reasons. First, by dropping *rationality* as the attribute category, the SSE-IAT primes participants to focus on the evaluation of *empathy*. The category *rationality* may have introduced ambiguity about the meaning of the score of the test. It is unclear if a positive SE-IAT score is due to a stronger preference for scientists in terms of their rationality or a stereotypical view against scientists in terms of their empathy. SSE-IAT avoids such ambiguity and only assesses relations of empathy to science and humanities in particular.

Second, as a variant to the traditional IAT, it is very easy to modify the SE-IAT to the SSE-IAT by simply dropping one category. When there is no natural comparison category to the construct of interest, some researchers have suggested using an unrelated neutral category (e.g., furniture or animals). This type of alteration assumes that the neutral contrasting category contributes no meaningful variability to measurement and thus results in a score that can be interpreted as an uncontaminated assessment of the target concept (Jajodia & Earleywine, 2003; Sherman, Rose, Koch, Presson, & Chassin, 2003). Nevertheless, evidence has shown that the neutral category could induce systematic error variance and thus reduce validity to the IAT (Penke et al., 2006). Moreover, from a practical point of view, it also requires extra effort to select a neutral category as well as the items to represent the new category, not to mention the fact that there is no standard to decide if the category is truly neutral. On the contrary, SSE-IAT completely bypasses the requirement of selecting contrasting category and includes only one attribute category from the original SE-IAT.

Third, existing single-category IATs have shown satisfactory reliability and validity. In terms of reliability, single-category IATs have shown acceptable internal consistency ranging from r = .55 -.85 obtained by Karpinski and Steinman (2006) and r =.65 -.84 obtained by Bar-Anan and Nosek (2014). In a study comparing psychometric properties of seven implicit measures, single-category IAT has shown better discriminant and convergent validity than the Sorting Paired Features task, the Evaluative-Priming task, and often not far behind traditional IAT (Bar-Anan & Nosek, 2014). The existing evidence suggests that single-category IAT is a psychometrically sound tool, encouraging its further use, especially for its unique procedural features.

It is worth mentioning that the Go/No-go Association Task (GNAT; Nosek & Banaji, 2001) and the Extrinsic Affective Simon Task (EAST; De Houwer, 2003) may also serve as an alternative implicit measure to assess single attitudes in a non-relative manner. However, the GNAT task was not selected for it is more sensitive to extraneous influences, such as cognitive fluency, than IAT and single-category IAT (Bar-Anan &

Nosek, 2014). The EAST was not selected for it is less sensitive to individual differences in attitudes as compared to IAT (Houwer & Bruycker, 2007). Furthermore, both the GNAT and EAST measures tend to have low reliability (Bar-Anan & Nosek, 2014; De Houwer, 2003), thus were not regarded as appropriate for the present study. To sum up, the SSE-IAT was modified and selected because it 1) is able to focus on assessment of empathy; 2) is easy to implement by simply dropping one category from the original SE-IAT; and 3) has displayed better reliability and validity than other alternative implicit measures.

6.2.4.4. Career Aspirations in Science (CAS) scale

In order to investigate the interplay of implicit and explicit stereotype in career choices, a brief Career Aspirations in Science (CAS) scale was applied. The CAS was used to capture intentions of pursuing a career in science by asking participants to rate their likelihood of doing science-related jobs on a 7-point Likert scale from "very likely" to "very unlikely". The three items in the CAS are as below:

In your future career, how likely is it that you will:

- Get an advanced degree in science?
- Have a very successful career in science?
- Have a lifelong career in science?

The three items were adapted from the Intentions of Pursuing a Career in Science questionnaire (Schneider, 2010). These items were selected for two reasons. Firstly, the original questionnaire is comprised of 12 items and most of them are quite repetitive, so that it might invoke the fatigue effect if all items were included in the current study. It has been found that tired and bored participants may more often answer "don't know", engage in "straight-line" responding (i.e., choosing answers down the same column on a page), give more perfunctory answers, or give up answering the questionnaire altogether, and as such this could deteriorate the data quality (Bloom, 2008). Secondly, all items in the CIS were found to load on the same factor (Schneider, 2010). The top three items with the highest loadings on the one-factor model greater than .90 were selected (Schneider, 2010). As such, the selected items should be able to tap on the same construct that the original scale assesses.

6.2.5. Data preparation

In this section, the procedure of how to calculate the scores for each measure is explained in detail. The same D scoring algorithm was applied to calculate the SSE-IAT effect with data from blocks that are slightly different from the standard IAT. Scores for the self-report explicit measures were calculated in accordance with the SSE-IAT results with positive scores representing stereotype-congruent beliefs against empathy in scientists and the negative scores representing counter-stereotype beliefs about empathy in scientists. Positive intentions in pursuing science careers were scored with positive integers and vice versa.

6.2.5.1. SSE-IAT scoring

The same improved D scoring algorithm was applied to calculate the SSE-IAT effect in the present study. Given that the SSE-IAT has only 5 blocks, the data used for calculation is slightly different from the standard SE-IAT. In the SE-IAT, it was the data from Block 3,4,6,7 that were used for calculation (See Table 6.8); while in the SSE-IAT, it was the data from Block 2,3,4,5 that were used (See Appendix XII for an example of the recorded trial latencies in the SSE-IAT). Table 6.8 illustrates the steps to calculate the D score for SSE-IAT.

Table 6.8 D scoring algorithms for SSE-IAT

Step		Step	
1	Use data from B2, B3, B4, & B5	7	Replace each error latency with block mean (computed in Step 5) + 600 ms
2	Eliminate trials with latencies > 10,000 ms; eliminate subjects for whom more than 10% of trials have latency less than 300 ms	8	No transformation
3	Use all trials	9	Average the resulting values for each of the four blocks
4	No extreme value treatment (beyond Step 2)	10	Compute two differences: B4-B2 and B5-B3
5	Compute mean of correct latencies for each block	11	Divide each difference by its associated pooled-trials SD from Step 6
6	Compute one pooled SD for all trials in B2 & B4; another for B3 & B5	12	Average the two quotients from Step 11

Note. Block numbers (e.g., B1) refer to the procedures shown in Table 5.12. SD = standard deviation. Adapted from Greenwald et al. (2003).

Similar to the SE-IAT, a positive D score of SSE-IAT also represents a stereotype-congruent belief that *empathy* is more associated with *humanities* than *science*, but a negative D score represents a counter-stereotype implicit belief that *empathy* is more associated with *science* than *humanities*. The value of the D score shows the strength of the associations. The bigger the value of the D score is, the stronger the associations between the concepts are. The D score ranges from -2 to +2. The strength of the IAT effect can be described as 'strong' (.80), 'medium' (.50), 'slight' (.20), or 'little or no' when interpreting the D scores. These cut-offs are in correspondence to results meeting conventional criteria for small, medium, and large effect size of Cohen's *d* (1988) measure (Nosek et al., 2005).

6.2.5.2. SSE-Explicit scoring

Consistent with the SSE-IAT score, stereotype-congruent responses associating *empathy* more with *humanities* were coded with positive integers while counter-stereotype responses associating *empathy* more with *science* were coded with negative integers. For the first two items in the SSE-Explicit questionnaire, responses were collected on 4-point Likert scales ranging from "strongly associated" to "not associated". For the *science-empathy* item, responses were coded negatively with "strongly associated" coded as -3 to "not associated" coded as 0. In contrast, for the *humanities-empathy* item, responses were coded positively with "strongly associated" coded as +3 to "not associated" coded as 0.

For the third item "scientists are less empathetic than humanities majors", responses were collected on a 7-point Likert scale ranging from "strongly disagree" (coded as -3) to "strongly agree" (coded as +3). The middle point "neither agree nor disagree" was coded zero. The sum of the scores of the three questions in the scale was regarded as the final score of the SSE-Explicit, ranging from -6 to +6.

6.2.5.3. ISSOS scoring

The same scoring method was applied for ISSOS as in Stage 2 (See Section 5.2.5.3). The final score of the ISSOS ranges from -27 to +27. The counter-stereotype responses (i.e., that scientists have good social skills) were reflected by negative ISSOS scores and the stereotype-congruent responses (i.e., that scientists do not have good social skills) were reflected by positive ISSOS scores.

6.2.5.4. CAS scoring

For the CAS scale, responses were collected on a 7-point Likert scale ranging from "very unlikely" (coded as 1) to "very likely" (coded as 7). The sum of the scores of the three items in the scale was regarded as the final score of the CAS, ranging from 3 to 21. Final scores smaller than 12 indicate relatively low career aspirations in science (unlikely to pursue a career in science) while final scores bigger than 12 indicate relatively high career aspiration in science (likely to pursue a career in science). The bigger the final score is, the higher career aspirations in science the participant report.

6.2.6. Data analysis

In order to preserve credibility of the results, rigorous procedures of data analysis were followed in the present study. Sources of bias (e.g., missing data and outliers) as well as assumptions of tests (e.g., normality and homogeneity of variance) were checked and treated using similar techniques in Stage 2 (See Section 5.2.6 for detailed justifications) before applying any parametric test with the data (e.g., t-test and correlation).

The Missing Value Analysis results suggest less than 5% of the cases have missing values and are missing at random as well. In this case, missing data was deleted listwise (Field, 2013). No outlier was spotted for the SSE-Explicit and CAS scores, and only 2 outliers were spotted for SSE-IAT and 3 outliers for ISSOS scores. These cases with outliers were removed from the dataset on the grounds that they might have represented rather idiosyncratic situations. In terms of normality, histograms and P-P plots showed that data for all measures are normally distributed with skew and kurtosis values very close to 0. Given that the sample size of the present study is fairly large (n = 1448), the assumption of homogeneity of variance can be ignored (Field, 2013).

6.3. Results

The purpose of this section is to provide readers with the statistical facts and results of the quantitative analysis. The section starts with illustration of descriptive data of variables of the present study. Then, results of comparisons of stereotypes by gender and major subject are demonstrated with implicitly measured stereotypes presented first followed by explicitly measured stereotypes. Then correlations between implicit and explicit stereotypes are examined. Last but not least, the relationship between stereotypes and aspirations in science are also investigated. Later in the Discussion section, these results are integrated and discussed with reference to existing literatures in the field.

6.3.1. Descriptive data: the basic status quo of implicit and explicit stereotypes of empathy in scientists in science among UK university students

First of all, descriptive data of the involved variables are presented so that readers could gain a general knowledge of the status of the current sample for each variable. As mentioned in the previous methodology section, implicit stereotypes were measured by SSE-IAT, which intends to capture implicit bias about empathy in science majors as compared to empathy in humanities majors. Explicit stereotypes were measured by SSE-Explicit and ISSOS. SSE-Explicit intends to measure the corresponding explicit construct that SSE-IAT captures, that is the explicit bias toward empathy in scientists as compared to empathy in humanists. At the same time, ISSOS were included to measure the explicit attitude toward social skills in scientists. Table 6.9 illustrates the mean (M), standard deviation (SD) and the range of current participants' SSE-IAT, SSE-Explicit and ISSOS scores.

Table 6.9 Descriptive data of the SSE-IAT, SSE-Explicit, ISSOS results

				Range	
Variable	n	M	SD	Possible	Observed
SE-IAT	1448	0.28	0.25	-2 – 2	-.73 – 1.16
SE-explicit	1425	0.84	2.22	-9 – 9	-5.0 – 6.0
ISSOS	1434	-9.8	7.07	-27 – 27	-27 – 21

Note. The variation in sample size is due to the variation in the number of participants who completed each measure.

The data showed that the average SSE-IAT score was positive (M = .28, within the slight effect range of .20 to .50), meaning that participants held a weak stereotype-congruent implicit bias against empathy in scientists as compared with empathy in humanists. For the explicit stereotypes, the data showed that the average SSE-Explicit score was also positive (M = 0.83, within the little or no effect range of 0 to 2.25). However, the mean SSE-Explicit score was so small that could be ignored, indicating little or no explicit preference for scientists or humanists regarding their empathy in current sample. Furthermore, data showed that the average ISSOS score was negative (M = -9.8, within the slight effect range of 6.75 to 13.5), meaning that current participants explicitly held a weak counter-stereotype positive view about scientists' interpersonal skills.

In terms of participants’ career aspirations in science, those who have already chosen to study humanities subjects at the higher education level, their mean CAS score (M = 5.75) was smaller than 12, meaning that they reported no career aspiration in science Instead, for participants who already study sciences at the higher education level, the mean CAS score was bigger than 12 (M = 16.61), showing that they possess career aspirations in science[1]. Table 6.10 illustrates the mean (M), standard deviation (SD) of CAS scores for science majors and humanities majors, respectively.

Table 6.10 Mean and standard deviations of CAS scores for science majors and humanities majors

	Science		Humanities	
	M	SD	M	SD
CAS	16.61	3.54	5.75	4.33

To sum up, participants in general showed stereotype-congruent implicit bias against empathy in scientists for they automatically associated empathy more strongly with humanities than with sciences. However, participants did not explicitly report any preference for scientists or humanists regarding their empathy and they even reported counter-stereotype positive views about scientists' interpersonal skills. Moreover, participants majoring in humanities subjects showed no science career aspirations whereas those majoring in science subjects showed science career aspirations.

6.3.2. Comparisons of the *implicit* stereotypes of empathy in scientists

Secondly, variations in the implicit stereotypes of empathy in scientists were under examination. In this part, individual differences in the implicit stereotypes of empathy in scientists by gender and major subject were examined using the SSE-IAT scores. The SSE-IAT assesses the strengths of automatic associations of *empathy* with *science* and with *humanitie*s.

Comparisons between women and men’s implicit stereotypes were conducted first

[1] Note: The cut-point of 12 indicates "not sure about pursuing a career in science". Scores smaller than 12 indicate "unlikely to pursue a career in science" and scores bigger than 12 indicate "likely to pursue a career in science". The bigger the score is, the higher career aspiration in science the participant has.

followed by comparisons between science majors and humanities majors' implicit stereotypes. Then the interaction between gender and major subject effects on the implicit stereotypes of empathy in scientists were examined. Afterwards, the dichotomous groupings of science/humanities academic fields were broken down to five specific major subjects and variations in the implicit stereotypes by these specific majors were examined. Finally, gender differences of the implicit stereotypes of empathy in scientists within each specific major subject were investigated as well.

6.3.2.1. Women versus men's *implicit* stereotype of empathy in scientists (SSE-IAT)

In terms of the gender effect on the implicit stereotypes of empathy in scientists, averaged across the entire sample, women (M = .27, SD = .25) displayed smaller mean SSE-IAT effect than men (M = .29, SD = .26). However, this difference, -.02, BCa 95 *CI* [-.008, .044], was not significant $t(1421) = -1.34$, $p = .18$, two-tailed; and represented a very small effect, $d = 0.07$. Such results indicate that women and men in general had similar implicit bias against empathy in scientists. Figure 6.2 illustrates the means of SSE-IAT scores of women and men.

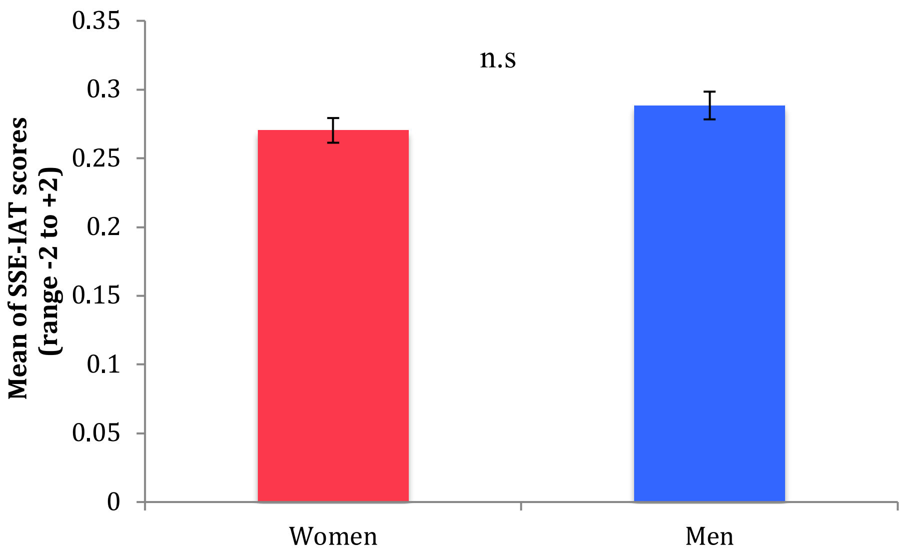


Figure 6.2 Comparisons of women and men's SSE-IAT scores.

Non-significant difference is marked by n.s.

6.3.2.2. Science versus humanities majors' *implicit* stereotype of empathy in scientists (SSE-IAT)

In the present study, we refer to the dichotomous groupings of Science and Humanities as participants' *academic fields*. Participants identified with Engineering & Technology, Life Science & Medicine, and Physical Science & Maths were grouped as the *science majors*, while participants majoring in Arts & Humanities and Social Science & Management were grouped as the *humanities majors*.

In terms of the academic field effect on the implicit stereotypes of empathy in scientists, the result of independent-samples t-test showed that science majors (M = .27, SD = .26) displayed slightly smaller SSE-IAT effect than humanities majors (M = .29, SD = .24). However, this difference, -.017, BCa 95 *CI* [-.044, .010], was not significant $t(1379) = -1.269$ (two-tailed), $p = .196$; and represented very little effect, $d = 0.07$. Such results suggest that participants majoring in STEM field and those majoring in Non-STEM field had similar levels of implicit prejudice against empathy in scientists. Figure 6.3 illustrates the means of SSE-IAT scores of participants majoring in STEM and Non-STEM fields.

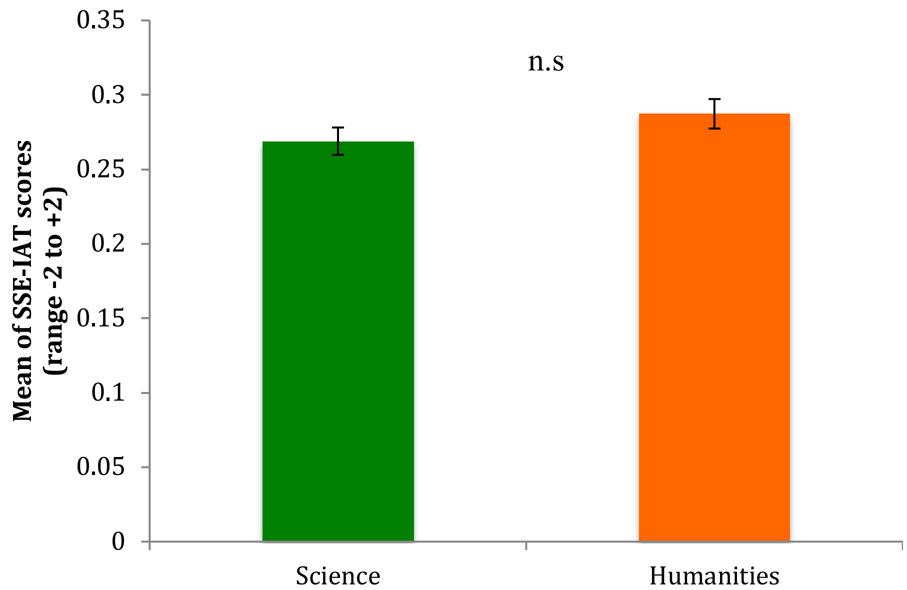


Figure 6.3 Comparisons of science and humanities majors' SSE-IAT scores

Non-significant difference is marked by n.s.

6.3.2.3. Interactions of gender and academic field effects on the *implicit* stereotype (SSE-IAT)

In terms of the interactions between gender and academic field effects on the implicit stereotypes of empathy in scientists, the results of 2 (women, men) * 2 (science, humanities) ANOVA indicate neither significant effect of gender, F (1, 1379) = .368, p = .544 nor significant effect of academic major field, F (1, 1379) = .690, p = .406 on SSE-IAT scores. However, there is a significant interaction between gender and academic field, F (1, 1379) = 25.574, p = .000, ω = 0.26. This effect indicates that the gendered patterns of implicit bias toward empathy in scientists differ by academic field. Figure 6.4 illustrates the comparisons between women and men's SSE-IAT scores in science and humanities academic fields.

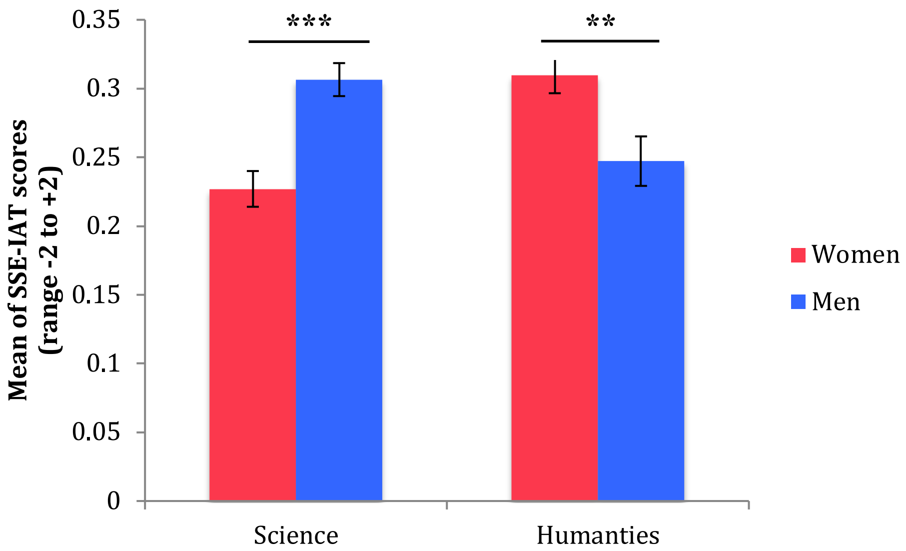


Figure 6.4 Comparisons between women and men's mean SSE-IAT scores by academic field

*Significant gender differences are marked by asterisks (***p < .001; **p < .01)*

Data indicates that for science majors, women had weaker implicit bias about empathy in scientists than men. However, such gendered pattern was reversed for humanities majors, with men showing weaker implicit bias against empathy in scientists than women.

Specifically, among science majors, women (n = 376, M = .23, SD = .26) showed smaller mean SSE-IAT score than men (n = 437, M = .31, SD = .26). This difference, -.08, BCa 95% *CI* [-.12, -.04], was significant $t(811) = -4.35$ $p = .000$, Cohen's $d = .31$, representing a small-sized effect. On the contrary, for humanities majors, women (n = 364, M = .31, SD = .22) showed bigger mean SSE-IAT score than men (n = 202, M = .25, SD = .25). This difference, .06, BCa 95% *CI* [.02, .10], was also significant $t(564) = 3.05$ $p = .002$, Cohen's $d = .25$, representing a small-sized effect.

6.3.2.4. Variations of the *implicit* stereotype (SSE-IAT) by specific major subject

As mentioned in the Sample section, participants in the present study were asked to identify their specific major subjects, namely, Arts & Humanities (N = 326, 23.6%), Social Science & Management (N = 241, 17.5%), Life Science & Medicine (N = 252, 18.2%), Engineering & Technology (N = 263, 19.0%) and Physical Science & Maths (N = 299, 21.7%). In the present study, we refer to this more specific groupings as participants' *major subject*. As discussed in the literature review, variations may also exist among participants majoring different subjects within the science and humanities fields. As such, comparisons among participants with different major subject backgrounds were also examined using one-way analyses of variance (ANOVA) of the SSE-IAT scores. Meanwhile, given that the sample size for each group is slightly different, Gabriel's procedure was selected as the post hoc test to do pairwise comparisons between those groups (Field, 2013).

In terms of the variations in implicit stereotypes of empathy in scientists, the results of

ANOVA showed no significant difference of SSE-IAT performance among participants from different major subjects, F (4, 1376) = 1.982, p = .095. Such results suggest that participants with varied academic subject backgrounds did not differ in their implicit bias about empathy in scientists, indicating they all associated empathy more with humanities than with sciences. Figure 6.5 shows the mean SSE-IAT scores of different major subjects.

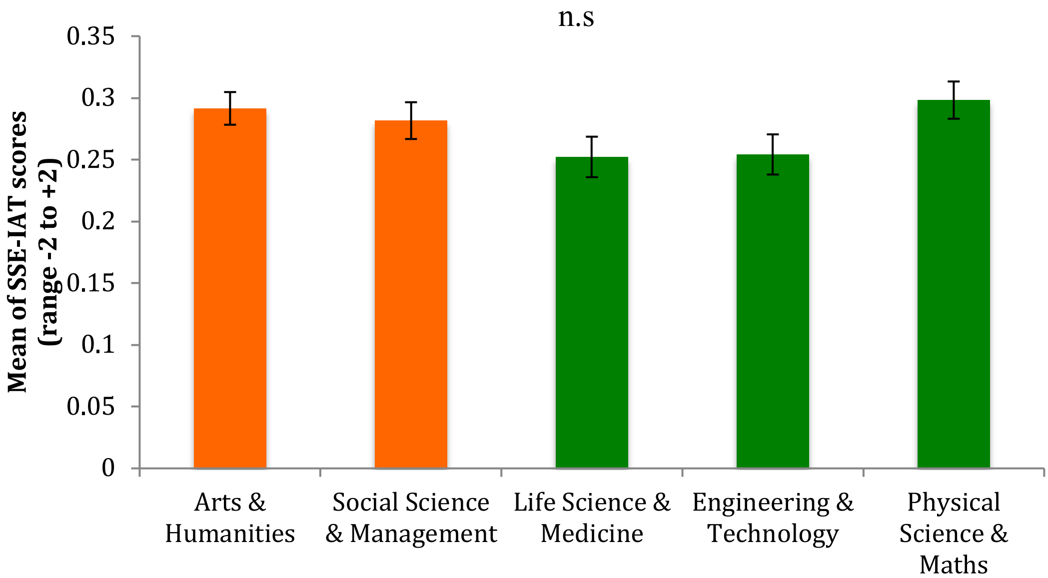


Figure 6.5 Means of the SSE-IAT scores of different major subjects

Non-Significant difference is marked by n.s

6.3.2.5. Gender difference of the *implicit* stereotype (SSE-IAT) within each major subject

To examine the gender difference within each specific subject major, independent t-tests were conducted with SSE-IAT scores between women and men majoring in each major subject separately. Significant gender difference of SSE-IAT was only found for participants majoring in Physical Science & Maths, but not for any other major subject. Specifically, for participants majoring in Physical Science & Maths, women (n = 126, M = .22, SD = .27) scored significantly lower than men in SSE-IAT (n = 173, M = .35, SD = .25). This difference, -.13, BCa 95% *CI* [-.18, -.07], was significant t(297) = -

4.20, p = .000, Cohen's d = .50, representing a medium-sized effect. Such results indicate that, for those majoring in Physical Science & Maths, women showed significantly weaker stereotype-congruent implicit bias against empathy in scientists. However, as for participants majoring in other subjects, women and men showed similar implicit bias against empathy in scientists. Figure 6.6 below illustrates the mean SSE-IAT scores of women and men by specific major subject.

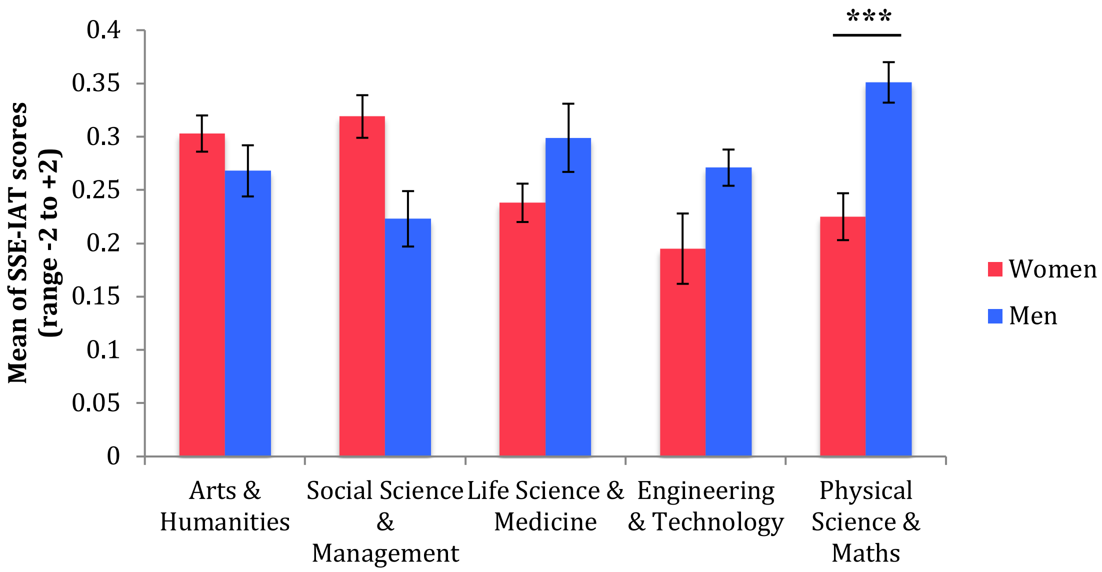


Figure 6.6 Means of the SSE-IAT scores of women and men by specific major subject

*Significant gender difference is marked by asterisks (***p < .000)*

6.3.3. Comparisons of the *explicit* stereotypes of empathy in scientists

Thirdly, variations in the *explicit* stereotypes of empathy in scientists were also under examination. In this part, individual differences in the explicit stereotypes of empathy in scientists by gender and major subject were examined using both the SSE-Explicit scores and the ISSOS scores, respectively. The SSE-Explicit scores can reflect participants' self-report preference for scientists or humanists in terms of their empathy. The ISSOS scores reflect participants' self-report attitudes about social skills in scientists.

Comparisons between women and men's explicit stereotypes were conducted first followed by comparisons between science majors and humanities majors' explicit stereotypes. Then the interaction between gender and major subject effects on the explicit stereotypes of empathy in scientists were examined. Afterwards, the dichotomous groupings of the science/humanities academic fields were broken down to five specific major subjects and variations in the explicit stereotypes by these specific major subjects were examined. Finally, gender difference of the explicit stereotypes within each specific major subject were investigated as well.

6.3.3.1. Women versus men's *explicit* stereotypes of empathy in scientists

6.3.3.1.1. Women versus men's SSE-Explicit scores

In terms of the gender difference in participants' explicit bias about empathy in scientists, women (M = .66, SD = 2.24) displayed smaller mean SSE-Explicit score than men (M = 1.04, SD = 2.18). This difference, -.38, BCa 95 *CI* [-.148, .612], was significant $t(1407) = -3.22$ (two-tailed), $p = .001$; however, it only represented a small-sized effect, $d = 0.17$. Such results indicate that women reported significantly weaker explicit bias against empathy in scientists than men. Figure 6.7 illustrates the means of SSE-Explicit scores of women and men.

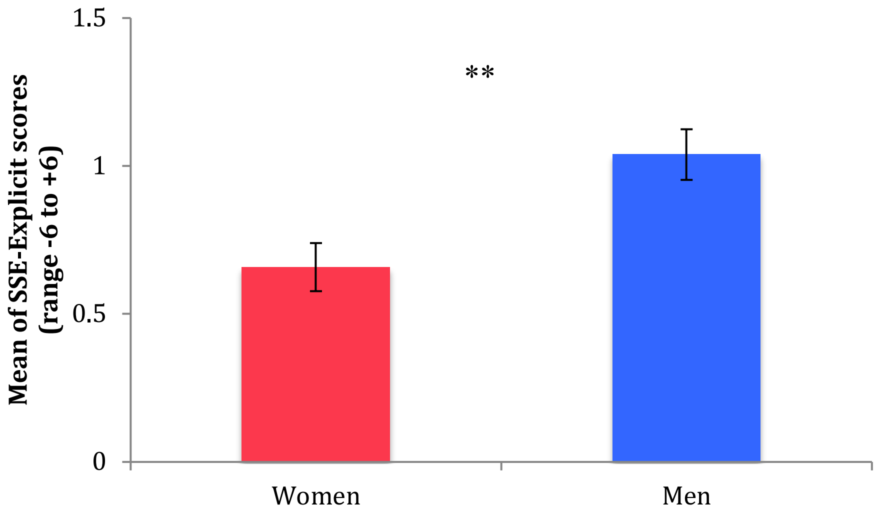


Figure 6.7 Comparisons of women and men's SSE-Explicit scores

*Significant difference is marked by asterisks (**p < .01)*

6.3.3.1.2. Women versus men's ISSOS scores

In terms of the gender difference in participants' explicit attitudes about social skills in scientists, women (M = -10.30, SD = 7.02) showed more negative ISSOS scores than men (M = -9.22, SD = 7.07). This difference, -1.08, BCa 95% CI [.345, 1.821], was significant $t(1411) = -.004$ (two-tailed); however, it only represented a small-sized effect, $d = 0.15$. Such results suggest that women reported significantly stronger counter-stereotype positive views about scientists' social skills than men. Figure 6.8 illustrates the means of ISSOS scores of women and men.

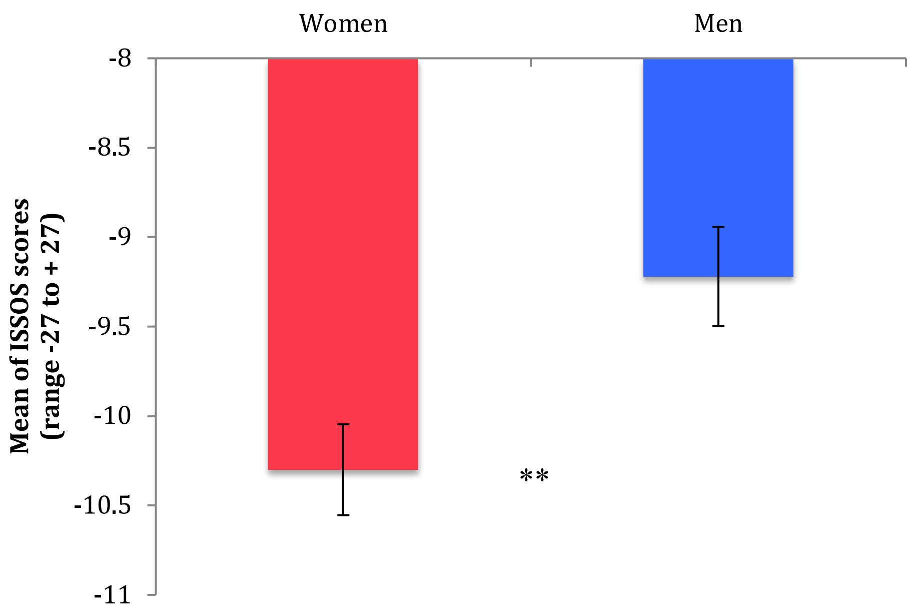


Figure 6.8 Comparisons of women and men's ISSOS scores

*Significant difference is marked by asterisks (**p < .01)*

6.3.3.2. Science versus humanities majors' *explicit* stereotype of empathy in scientists

6.3.3.2.1. Science versus humanities majors' SSE-Explicit scores

In terms of the academic field effect on the SSE-Explicit scores, science majors (M = .36, SD = 2.20) displayed smaller SSE-Explicit scores than humanities majors (M = 1.46, SD = 2.10). This difference, -1.10, BCa 95 *CI* [-1.337, -.873], was significant $t(1365) = -9.338$, $p = .000$ (two-tailed); and it did represent a medium-sized effect, $d = 0.52$. Such results indicate that science majors reported significantly weaker explicit bias against empathy in scientists than humanities majors. Figure 6.9 illustrates the means of SSE-Explicit scores of science and humanities majors.

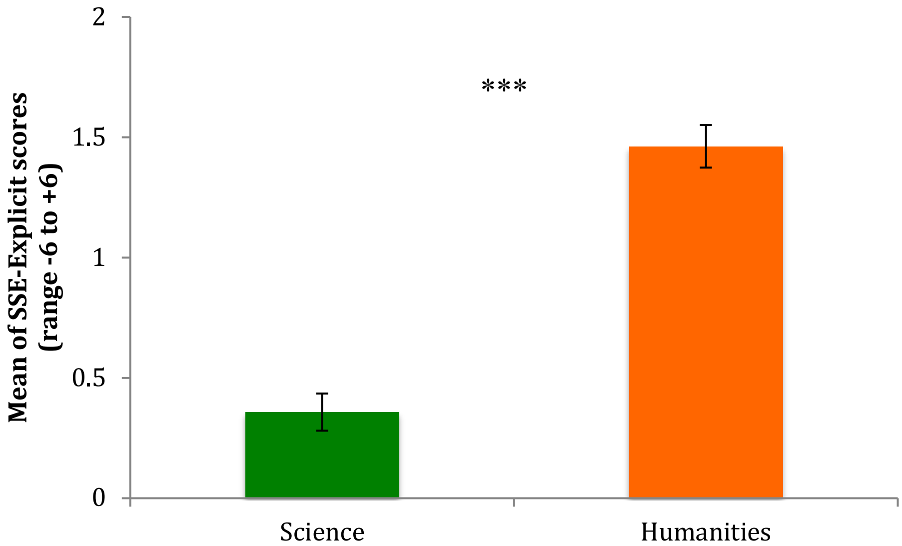


Figure 6.9 Comparisons of science and humanities majors' SSE-Explicit scores

*Significant difference is marked by asterisks (***p < .001)*

6.3.3.2.2. Science versus humanities majors' ISSOS scores

In terms of the academic field effect on the ISSOS scores, science majors (M = -11.84, SD = 6.62) showed more negative ISSOS scores than humanities majors (M = -7.18, SD = 6.76). This difference, -4.658, BCa 95% *CI* [.345, 1.821], was significant $t(1369) = -12.699$, $p = .000$; and it did represent a medium-sized effect, $d = 0.70$. Such results indicate that science majors reported significantly stronger counter-stereotype positive views about social skills in scientists than humanities majors. Figure 6.10 illustrates the means of ISSOS scores of science and humanities majors.

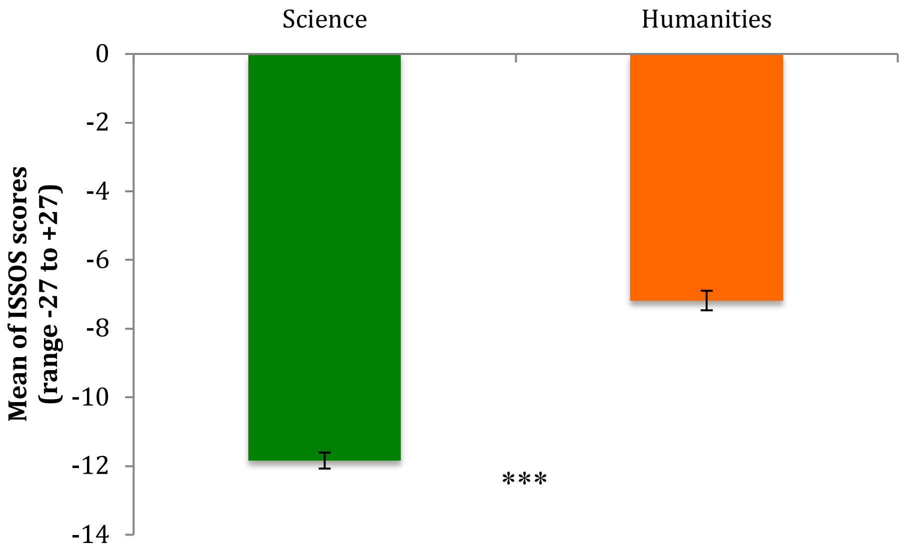


Figure 6.10 Comparisons of science and humanities majors' ISSOS scores

*Significant difference is marked by asterisks (***p < .001)*

6.3.3.3. Interactions of gender and academic field effects on the *explicit* stereotypes

6.3.3.3.1. Interactions of the gender and academic field effects on the SSE-Explicit scores

In terms of the interaction between gender and academic field effects on the SSE-Explicit scores, the results of 2 (women, men)* 2 (science, humanities) ANOVA indicate both significant effects of gender, F (1, 1365) = 19.51, $p = .000$, $\omega = 0.11$ and academic field, F (1, 1379) = .94.16, $p = .000$, $\omega = 0.25$, on SSE-Explicit scores, but no significant interaction effect between them, F (1, 1365) = 5.54, p = .019.

For science majors, women reported significantly weaker explicit bias against empathy in scientists than their men majoring in science. However, such gender difference was not significant for humanities majors. Specifically, for science majors, women (n = 372, M = -.08, SD = 2.13) showed smaller mean SSE-IAT score than men (n = 431, M = 0.74, SD = 2.17). This difference, -.82, BCa 95% CI [-1.12, -.52], was significant t(801) = -5.35 p = .000, Cohen's d = .38, representing a small-sized effect. Nevertheless, for humanities majors, though women (n = 363, M = 1.37, SD = 2.11) also showed smaller mean SSE-Explicit score than men (n = 199, M = 1.62, SD = 2.08), this difference, -.25, BCa 95% CI [-.61, .11], was not significant t(560) = -1.34 p = .18. It also worth noting that only women majoring in science scored negatively in SSE-Explicit, showing a weak counter-stereotype explicit preference for scientists in terms of empathy. All the other groups scored positively in SSE-Explicit, showing stereotype-congruent explicit bias against empathy in scientists. Figure 6.11 illustrates the comparisons between women and men's SSE-Explicit scores in the science and humanities academic fields.

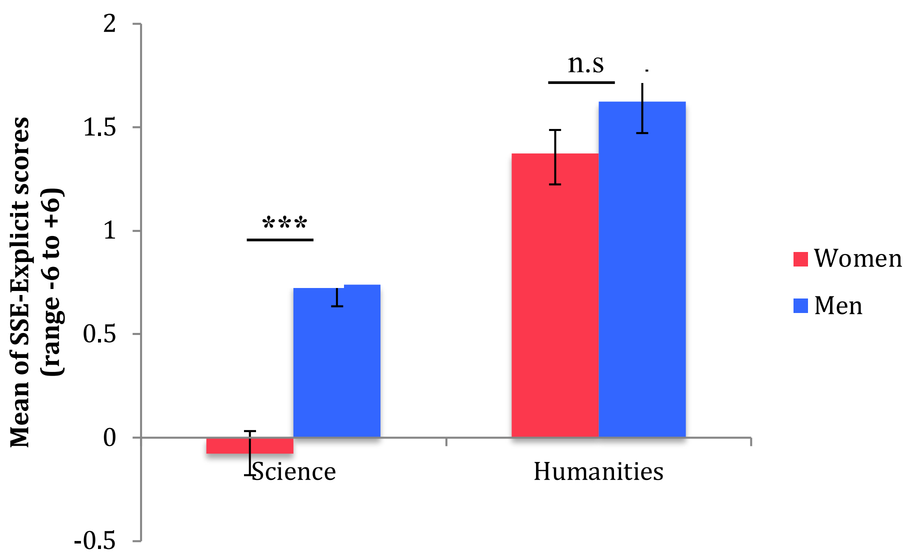


Figure 6.11 Gender differences in the SSE-Explicit scores by academic field

*Significant gender difference is marked by asterisks (***p < .001)*

6.3.3.3.2. <u>Interactions of the gender and academic field effects on the ISSOS scores</u>

Likewise, in terms of the interactions between gender and academic field effects on the ISSOS scores, the results of 2 (women, men)* 2 (science, humanities) ANOVA also indicate significant effect of gender, F (1,1365) = 28.61, p = .000, ω = 0.13, and academic major field, F (1, 1365) = 179.39, p = .000, ω = 0.34, but no significant interaction between them, F (1,1365) = .105, p = .746.

For science majors, women (n = 374, M = -12.97, SD = 6.37) scored more negatively in ISSOS than men (n = 434, M = -10.85, SD = 6.70). This difference, -2.12, BCa 95% *CI* [-3.03, -1.21], was significant t(806) = -4.59 p = .000, *Cohen's d* = .32, representing a small effect size. Similarly, for humanities majors, women (n = 363, M = -7.84, SD = 6.76) also scored more negatively in ISSOS than men (n = 198, M = -5.96, SD = 6.59). This difference, -1.88, BCa 95% *CI* [-3.04, -.72], was also significant t(559) = -3.17 p = .002, *Cohen's d* = .28, representing a small effect size. Figure 6.12 illustrates the comparisons between women and men's ISSOS scores in Science and Humanities academic fields.

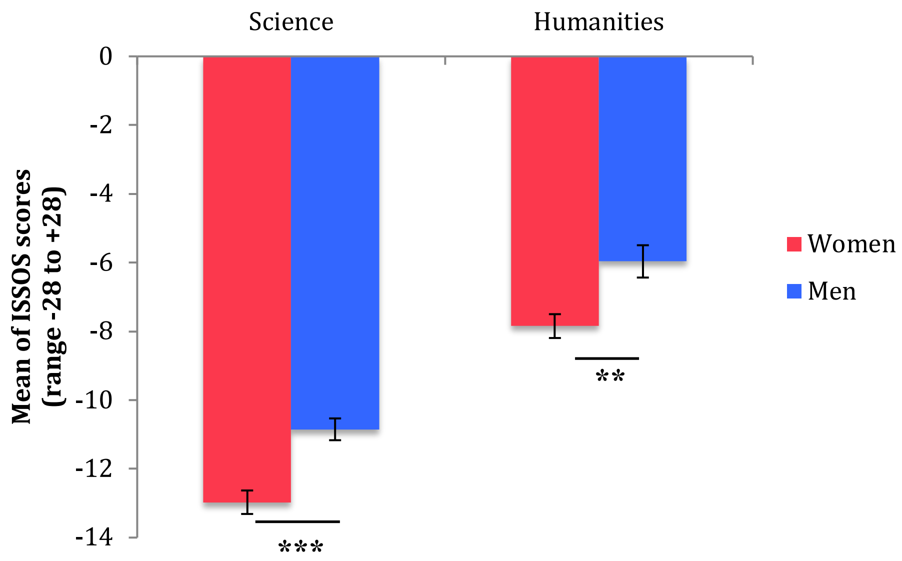


Figure 6.12 Gender differences in the ISSOS scores by academic field

*Significant differences are marked by asterisks (**p < .01, ***p < .001)*

6.3.3.4. Variations of the *explicit* stereotypes by specific major subject

6.3.3.4.1. Variations of the SSE-Explicit scores by specific major subject

In terms of the variations in the explicit bias about empathy in scientists, significant differences among participants with different major subjects were found for the SSE-Explicit scores, F (4, 1362) = 33.717, p = .000, ω = .29. As can be seen from Figure 6.13 below, only participants majoring in Life Science & Medicine (M = -.33, SE = .13) showed a negative mean score in SSE-Explicit, meaning that they reported attitudes that were not congruent with the stereotype, and were deemed to hold little explicit bias against empathy in scientists. However, participants majoring in all the other subjects scored positively in the SSE-Explicit (ranging from .48 to 1.56), meaning that they reported stereotype-congruent explicit bias against empathy in scientists.

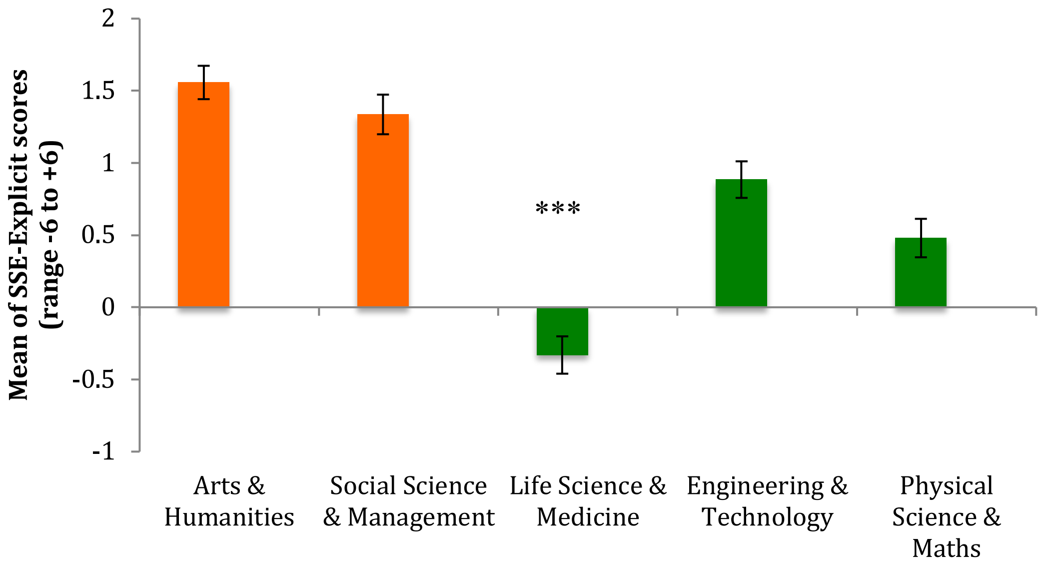


Figure 6.13 Means of the SSE-Explicit scores of different major subjects

*Significant difference is marked by asterisks (***p < .001)*

Specifically, the results of Gabriel's post hoc test of the SSE-Explicit scores indicate that participants majoring in Life Science & Medicine (M = -.33, SE = .13) reported significantly different explicit bias about empathy in scientists from their counterparts majoring in Arts & Humanities (M = 1.56, SE = .12), $p = .000$, *Cohen's d* = .92; those in Social Science & Management (M = 1.33, SE = .14), $p = .000$, *Cohen's d* = .80; those in Physical Science & Maths (M = .48, SE = .13), $p = .000$, *Cohen's d* = .37; as well as those in Engineering & Technology (M = .88, SE = .13), $p = .000$, *Cohen's d* = .59.

Moreover, there is neither significant difference of explicit bias about empathy in scientists between participants majoring in Arts & Humanities (M = 1.56, SE = .12) and those majoring in Social Science & Management (M = 1.33, SE = .14), $p = .915$ nor significant difference between participants majoring in Physical Science & Maths (M = .48, SE = .13) and Engineering & Technology (M = .88, SE = .13), $p = .223$.

6.3.3.4.2. Variations of the ISSOS scores by specific major subject

Likewise, regarding variations in the explicit attitudes about social skills in scientists, significant differences among participants with different major subjects were also found for the ISSOS scores, F (4, 1362) = 43.871, $p = .000$, $\omega = .33$. As can be seen from Figure 6.14 below, participants majoring in Life Science & Medicine had smallest negative ISSOS mean score among all the major subjects, meaning that they reported the strongest counter-stereotype positive explicit attitude about social skills in scientists.

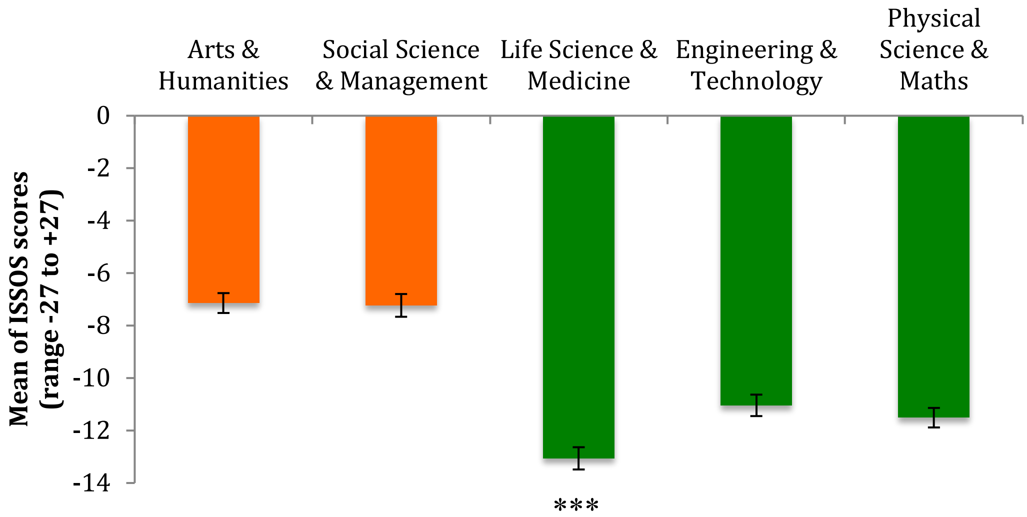


Figure 6.14 Means of the ISSOS scores of different major subjects

*Significant difference is marked by asterisks (***p < .001)*

Specifically, the results of Gabriel's post hoc tests show that participants majoring in Life Science & Medicine (M = -13.06, SE = .43) scored significantly lower in ISSOS than their counterparts majoring in Arts & Humanities (M = -7.14, SE = .38), *p* = .000, Cohen's *d* = .87; those in Social Science & Management (M = -7.23, SE = .43), *p* = .000, Cohen's *d* = .87 ; as well as those in Engineering & Technology (M = -11.04, SE = .41), *p* = .000, Cohen's *d* = .30. Nevertheless, participants majoring in Life Science & Medicine (M = -13.06, SE = .42) did not differ significantly from those in Physical Science & Maths (M = -11.51, SE = .37) in ISSOS, *p* = .06.

Similarly, there is again neither significant difference of explicit attitude toward social skills in scientists between participants majoring in Arts & Humanities (M = -7.14, SE = .38) and those majoring in Social Science & Management (M = -7.23, SE = .43), *p* = 1.0 nor significant difference between participants majoring in Physical Science & Maths (M = -11.51, SE = .37) and Engineering & Technology (M = -11.04, SE = .41), *p* = .99.

6.3.3.5. Gender difference of the *explicit* stereotypes within each major subject

6.3.3.5.1. Gender difference of the SSE-Explicit scores within each major subject

To examine the gender difference in the explicit bias about empathy in scientists within each specific subject major, independent t-tests were conducted with SSE-IAT scores between women and men majoring in each major subject, separately. Significant gender difference in the SSE-Explicit scores was only found for participants majoring in Physical Science & Maths, but not for any other major subject. Specifically, for participants majoring in Physical Science & Maths, women (n = 123, M = .08, SD = 2.20) scored lower than men in SSE-Explicit (n = 169, M = .78, SD = 2.31). This difference, -.70, BCa 95% *CI* [-1.23, -.17], was significant $t(290) = -2.60$, $p = .01$, *Cohen's d* = .31, representing a small effect size.

Such results indicate that women majoring in Physical Science & Maths reported significantly weaker explicit bias about empathy in scientists than their male counterparts. However, as for participants majoring in other subjects, women and men reported similar explicit bias against empathy in scientists. It is worth mentioning that for those majoring in Life Science & Medicine, women and men reported similar counter-stereotype explicit bias about empathy in scientists. Figure 6.15 illustrates the comparisons between women and men's SSE-Explicit scores by specific major subject.

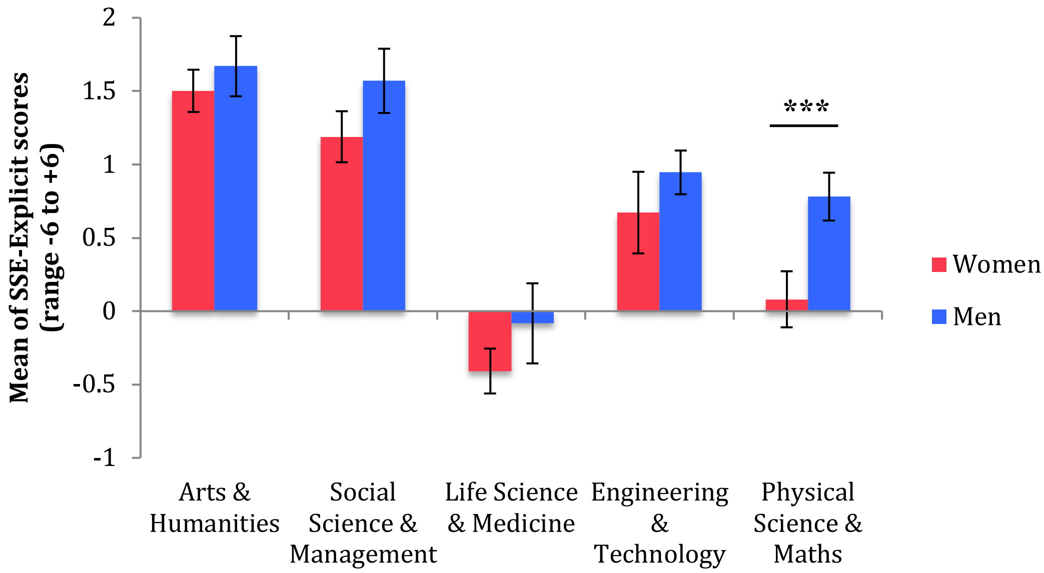


Figure 6.15 Means of the SSE-Explicit scores of women and men by major subject

*Significant difference is marked by asterisks (***p < .001)*

6.3.3.5.2. Gender difference of the ISSOS scores within each major subject

As for the gender difference in the explicit attitude about social skills in scientists within each specific subject major, significant gender difference of the ISSOS scores were found for participants majoring in Physical Science & Maths again and also for those majoring in Arts & Humanities. Specifically, for participants majoring in Physical Science & Maths, women (n = 125, M = -12.62, SD = 6.21) scored lower than men (n = 170, M = -10.65, SD = 6.48) in ISSOS. This difference, -1.98, BCa 95% *CI* [-3.45, -.50], was significant $t(293) = -2.60$, $p = .009$, Cohen's $d = .31$, representing a small-sized effect. Similarly, for participants majoring in Arts & Humanities, women also (n = 216, M = -7.99, SD = 6.82) scored lower than men (n = 108, M = -5.43, SD = 6.52) in ISSOS. This difference, -2.53, BCa 95% *CI* [-3.45, -.50], was significant as well $t(322) = -3.20$, $p = .002$, Cohen's $d = .38$, representing a small-sized effect. Such results indicate that women reported stronger counter-stereotype explicit positive views about social skills in scientists than their male counterparts for those majoring in Physical Science & Maths as well as those in Arts & Humanities. As for participants majoring in other subjects, women and men reported similar levels of positive attitudes about social skills in scientists. Figure 6.16 illustrates the comparisons between women and men's ISSOS scores by specific major subject.

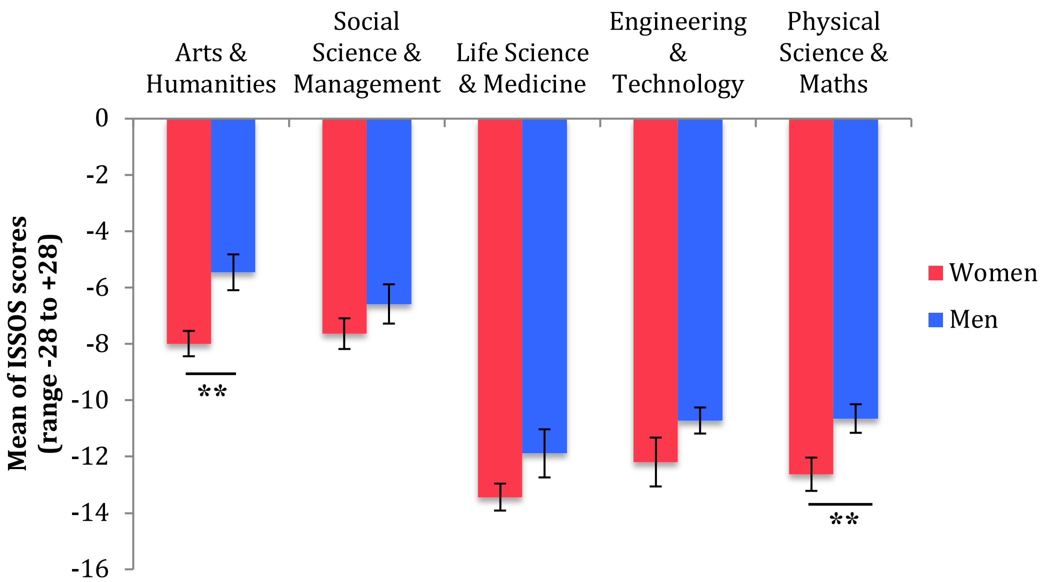


Figure 6.16 Means of ISSOS scores of women and men by specific major subject

*Significant differences are marked by asterisks (**p < .01)*

6.3.4. Comparisons of the career aspirations in science by gender and major subject

To examine the variations in career aspirations in science by gender and major subject, a 2 (women, men) * 5 (Arts & Humanities, Social Science & Management, Life science & Medicine, Engineering & Technology, Physical Science & Maths) ANOVA was conducted. The ANOVA results indicate no significant difference between women and men in their science career aspirations, F (1, 1375) = .379, p = .54. But significant differences were found by major subject, F (1, 1375) = 589.79, p = .000. There is no interaction effect of gender and major subject on the CAS scores, F (1, 1375) = .573, p = .68.

As explained in the Descriptive data, humanities majors in general reported no intention in pursuing science careers but science majors instead showed science career aspirations. Confirming this finding, the results of the Gabriel's post hoc test of the CAS scores indicate that participants majoring in Arts & Humanities (M = 4.84, SD = 3.47) and those in Social Science & Management (M = 6.89, SD = 5.03) reported significantly lower science career aspirations than participants majoring in Life Science & Medicine (M = 16.70, SD = 3.63), $p = .000$, those in Engineering & Technology (M = 16.47, SD = 3.35), $p = .000$, as well as those in Physical Science & Maths (M = 16.66, SD = 6.61), $p = .000$. Also, it is worth noting that there is no significant difference in science career aspirations between the Life Science & Medicine, Engineering & Technology and Physical Science & Maths majors. Figure 6.17 illustrates the means of the CAS scores by gender and specific major subject.

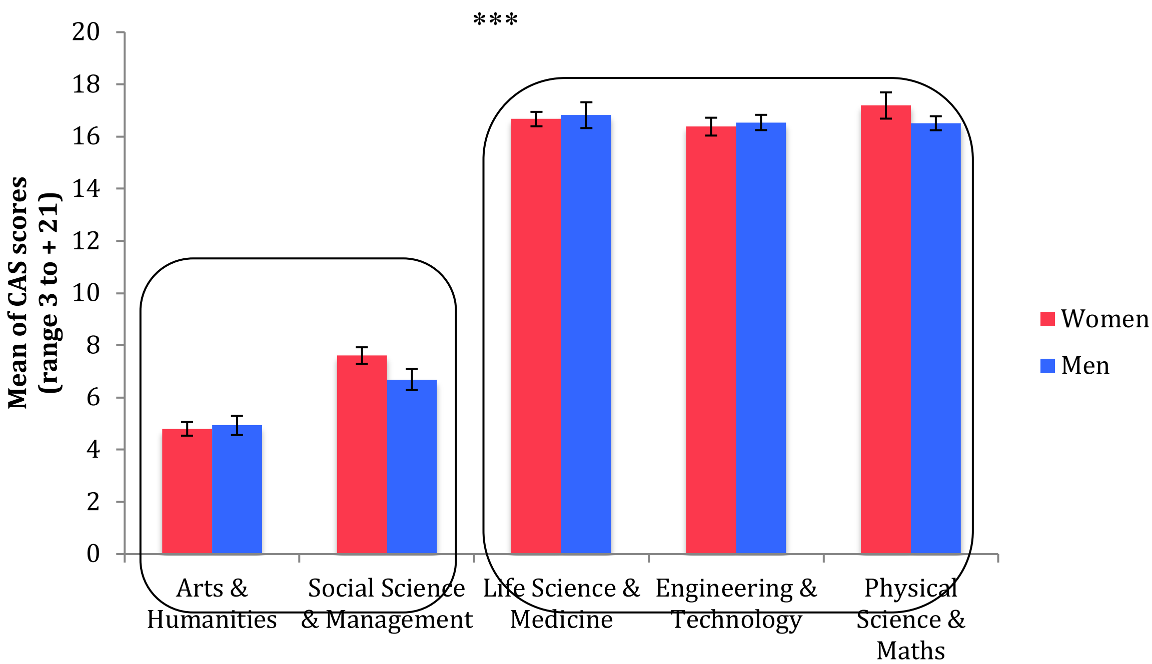


Figure 6.17 Means of the CAS scores by gender and specific major subject

*Significant differences are marked by asterisks (***p < .001)*

6.3.5. Correlations among variables

In order to examine the relationships among implicit and explicit stereotypes of empathy in scientists as well as the relationship between stereotypes and career aspirations, correlations among focal variables involved in current studies were examined. The results are illustrated in Table 6.11. The following correlations are worth paying attention to.

Table 6.11 Correlations between variables of the study

	1	2	3	4
1. SSE-IAT	--	.137**	.039	-.057*
2. SSE-Explicit	[.084, .192]	--	.409**	-.259**
3. ISSOS	[.027, -.015]	[.359, .457]	--	-.349**
4. CAS	[-.105, -.003]	[-.310, -.208]	[-.391, -.306]	--

*Note: Pearson's correlations are presented above the diagonal, and the corresponding bootstrap 95% CIs are presented below the diagonal. ** p< .001*

Firstly, the correlations among the three different kinds of stereotypes were under scrutiny. Results suggest that the implicit bias about empathy in scientists (SSE-IAT) has a significant and positive correlation with the explicit bias toward empathy in scientists (SSE-Explicit), $r(1421) = .37$. Statistically correlated and accelerated bootstrap 95% CIs are reported as follows: 95% BCa CI [.083, .195], p = .000, $r^2 = .019$. However, the implicit bias toward empathy in scientists (SSE-IAT) has *no* significant correlation with the explicit attitude about social skills in scientists (ISSOS), $r(1430) = .039$, $p = .117$. Meanwhile, there is a significant and positive correlation between explicit bias toward empathy in scientists and explicit attitude about social skills in scientists, $r(1411) = .409$, 95% BCa *CI* [.359, .457], $p = .000$, $r^2 = .167$.

Secondly, results show that career aspirations in science (CAS) has significant but negative correlations with all the three kinds of stereotypes: including implicit bias about empathy in scientists (SSE-IAT), $r(1431) = -.057$, 95% BCa *CI* [-.105, -.003], $p = .033$, $r^2 = .003$; explicit bias about empathy in scientists, $r(1417) = -.259$, 95% BCa *CI* [-.310, -.208], $p = .000$, $r^2 = .067$ and explicit attitude about social skills in scientists, $r(1421) = -.349$, 95% BCa *CI* [-.391, -.306], $p = .000$, $r^2 = .122$. It is worth noting that both explicit bias about empathy in scientists (SSE-Explicit) and explicit attitudes about social skills in scientists (ISSOS) displayed significant correlations with science career aspirations, but implicit bias about empathy in scientists (SSE-IAT) did not correlate with science career aspirations, therefore we suspect that explicit stereotypes may be more important than implicit stereotype for predicting career aspirations.

Thirdly, in order to identify the covariates for predicting career aspirations in science, it is also of necessity to examine whether demographic variables (i.e., gender, academic field, major subject, age, and degree level) are significantly correlated with the focal independent and dependent variables. Table 6.12 illustrates the correlations between demographic variables and the focal independent and dependent variables.

Table 6.12 Correlations between demographic variables and the focal variables

	Gender	Academic field	Major subject	Age	Degree level
SSE-IAT	-.027	.028	.024	-.047	.005
SSE-Explicit	-.085**	.240*	.144**	-.067*	-.015
ISSOS	-.073**	.332**	.256**	.064*	-.034
CAS	-.125**	-.760*	-.68**	.121**	.123*

*Note: For Degree level, Spearman's correlations are presented. For all the other demographic variables, Pearson's correlations are presented. * $p< .05$, ** $p< .001$*

All demographics were found significantly correlated with career aspirations in science. Nevertheless, major subject was dropped from the covariate selection for two reasons.

Firstly, given that academic field and major subject were closely correlated with each other, $r(1379) = .870$, $p = .000$, it would violate the multicollinearity assumption if both of them were included as covariates in the regression model (Field, 2013). Secondly, as presented before, the main difference in science career aspirations relied on whether one majored in humanities field or science field. There is no significant difference in career aspirations within subjects of each field. As such, only academic field was included as one of the covariates in the regression model. As a result, demographic variables that were significantly associated with the focal variables (namely, gender, age, academic field and degree level) were selected as the covariates in the regression model predicting career aspirations in science.

6.3.6. Multiple regression analyses

To answer the research question "if variations in the stereotypes of empathy in scientists can predict intentions in pursuing science careers", multiple regression analyses predicting career aspirations in science from implicit and explicit stereotypes were carried out.

In the present study, *hierarchical regression* method is selected because it allows the researcher to decide in which order to enter the predictors into the model based on existing work or theory (Field, 2013). Given that most of our predictors are correlated, the order of independent variables entry is crucial. As a general rule, covariates should be entered into the model first in order to control for their effects. After covariates have been entered, focal predictors can be entered hierarchically (such that the focal predictor suspected to be the most important is entered first).

In this regression model, predictors that explain career aspirations in science were entered in three steps. First of all, selected demographics, a control variable that consisted of age, gender, major subject, and degree level were entered in Step 1. After that, our focal predictors were entered hierarchically. It is important to note that given that both gender and academic field are dichotomous nominal variables, they had been dummy coded before entering the model. Women was coded as 0 and men as 1. Humanities majors were coded as 0 and science majors were coded as 1.

In the present study, explicit stereotypes were entered before the implicit stereotype for two reasons. Firstly, as explained in the literature review, recent *double dissociation* predictive models of implicit and explicit attitudes (e.g., Wilson, Lindsey, & Schooler, 2000) suggest that explicit attitudes are particularly valuable for the prediction of deliberate, controlled behaviours, while implicit attitudes are more important for the prediction of less controlled, more impulsive behaviours. As a vital decision that may affect all aspects of one's life, career aspirations in science were regarded as controlled processes that participants are sufficiently motivated to engage in deliberate reasoning. In this case, career aspirations were assumed more likely to be influenced by explicit than implicit stereotypes. Secondly, the previous correlations also showed that explicit stereotype measures (SSE-Explicit and ISSOS) had stronger correlations with our outcome variable (CAS) than the implicit stereotype measure (SSE-IAT). Under such circumstances, explicit stereotype measures (SSE-Explicit and ISSOS) therefore were entered into Step 2 followed by implicit stereotype measure (SSE-IAT) into Step 3.

Before the hierarchical multiple regression analysis was performed, the independent variables were examined for collinearity. Results of the variance inflation factor (all less than 2.0), and collinearity tolerance (all greater than .76) suggest that the estimated β (i.e., the standardised coefficients representing the relationship between predictors and outcome; Field, 2013) are well established in the following regression model. Table 6.13 summarized the full model by reporting partial correlations coefficients (*Partial r*), the unstandardized regression coefficients (*B*) and its standard errors (SE B), the standardised regression coefficients (β), and the R square change (*Block* ΔR^2) for each predictor.

Table 6.13 Hierarchical regression predicting career aspirations in science (CAS) from selected demographics, SSE-IAT, SSE-Explicit and ISSOS

Predictor	*Partial r*	*B*	*SE B*	*β*	*Block ΔR^2*
Step 1 Demographics					*.661*
Age	.007	.012	.011	.018	
Gender	.014	.112	.221	.008	
Academic field	.806***	10.93	.225	.814	
Degree level	.008	.038	.129	.005	
Step 2 Explicit stereotypes					*.008*
SSE-Explicit	-.052	-.099	.053	-.033	
ISSOS	-.116***	-.071	.017	-.076***	
Step 3 Implicit stereotype					*.000*
SSE-IAT	-.030	-.450	.427	-.017	

Note. N = 1297. Both gender and academic field are dummy coded. Total R2 = .036. F (9, 781) = 3.33, p =.001. Partial r shows the correlations between each predictor and the outcome variable, controlling for all other predictors in the model. Block ΔR^2 shows the change in R^2 *resulting from the inclusion of the new block of predictors. ***p < .001.*

As shown in Table 6.13, results of Step 1 indicate that the variance accounted for (R^2) with the first four control independent variables (gender, academic field, age, degree level) equaled .661 (adjusted R^2 = .660), which was significantly different from zero ($F_{(4, 1270)}$ = 618.81, *p* = .000). Among all the demographics, only major subject was a statistically significant predictor for science career aspirations.

In Step 2, explicit measures of stereotypes of empathy in scientists including SSE-IAT and ISSOS were entered into the regression equation. The change in variance accounted for (ΔR^2) was equal to .008, which was also significantly different from zero ($F_{(2, 1268)}$ = 14.60, *p* = .000) However, only ISSOS (*β* = -.125*, p* = .000) was a statistically significant predictor for science career aspirations.

In Step 3, the implicitly measured stereotype of empathy in scientists (SSE-IAT) was entered into the regression equation. The change in variance accounted for (ΔR^2) was equal to .000, which was not significantly different from zero ($F_{(1, 1267)} = 1.11, p = .292$). Therefore, SSE-IAT is not a statistically significant predictor for science career aspirations.

To conclude, the full model was responsible for 66.9% variance in career aspirations in science. Among all the demographic variables, only major subject was a significant predictor for science career aspirations. Explicit stereotypes about social skills in scientists (ISSOS) also significantly contributed to the prediction of career aspirations in science. However, neither implicit (SSE-IAT) nor explicit (SSE-Explicit) measure of stereotypes about empathy in scientists could significantly predict science career aspirations.

Inspection of the squared partial correlations in the prediction of career aspirations in science revealed that major subject predicts the most significant unique variance ($r_{sp}^2 = .65$, $p = .000$), indicating that studying a science major at the higher education level explains most of the variations in science career aspirations. ISSOS also predicts a small but significant unique variance in science career aspirations ($r_{sp}^2 = .01$, $p = .000$). The standardised β value for the ISSOS score is -.116, indicating that as the ISSOS score *decreased* in one standard deviation, the CAS score *increased* by .116 standard deviations. In other words, the less stereotypical/more positive attitude one reported about social skills in scientists, the higher career aspirations in science one would have had.

6.4. Discussion

In this section, the results of the quantitative analyses are synthesized to answer the research questions of the present study. Discussion of the findings is carried out with reference to existing theories and empirical studies in the field. The limitations of the current study are presented with suggestions for directions of future research.

6.4.1. The discrepancy between implicit and explicit stereotypes of empathy in scientists

6.4.1.1. What are the implicit and explicit stereotypes of empathy in scientists among UK university students?

The first aim of the present study is to investigate how people view empathy in scientists in contemporary UK society. In the present study, stereotypes of empathy in scientists were measured both implicitly and explicitly among UK university students. By combining relevant evidence produced by different measurements, a more comprehensive picture of people's perceptions about empathy in scientists as well as their attitudes toward social skills in scientists in UK higher education was generated. Evidence from the present study revealed discrepancies in implicit and explicit attitudes toward empathy in scientists among UK university students. Current findings were summarized in one sentence as below:

Finding 1: Participants showed implicit bias against empathy in scientists, but explicitly reported positive attitudes about scientists' empathy and social skills.

Specifically, the negative averaged IAT score suggests a weak stereotype-congruent implicit bias against empathy in scientists among participants. In other words, participants implicitly associated empathy more strongly with humanities than with sciences. This implicit bias was consistent with previous findings from the Public Attitudes to Science 2014 survey in the UK, showing that almost half of the surveyed

UK citizens perceived scientists as poor at communication (Ipsos MORI, 2014, p. 47). This implicit bias about empathy in scientists is also in accordance with the robust stereotypical depiction of scientists doing experiment alone in the lab found in the draw-a-scientist studies (Finson, 2002).

The average SSE-Explicit score was also positive but close to zero, representing a very weak explicit bias against empathy in scientists. However, such explicit bias was so trivial that can be ignored. Also, the negative averaged ISSOS score suggests counter-stereotype positive views about scientists' social skills among UK university students. These findings suggest that, in general, participants did not endorse the implicit bias about empathy in scientists and explicitly reported positive explicit views about empathy and social skills in scientists. Such findings are consistent with Archer and colleagues' (2013) study with UK adolescents as well as Schneider's (2010) study with US undergraduates. In their ASPIRES report, Archer and colleagues (2013) claimed that only 21% of Year 9 students in the UK agreed that scientists are “geeks” and even fewer (14%) believed that scientists are ‘odd’. Similarly, Schneider (2010) found that undergraduates in the US also explicitly perceived scientists as being good at interpersonal relations, especially with their colleagues in the workplace. But US college students remained sceptical about scientists' ability to handle relationships in personal life (Schneider, 2010). Evidence from the present study did not differ between personal and workplace relationships, thus could not detect if there is any variation in participants' attitudes about scientists' social skills in different situations. The discrepancy between negative implicit attitude and positive explicit attitude about empathy in scientists among UK university students suggests that the stereotypical image of scientists may still be deeply embedded in people’s unconscious mind, but young adults are unaware of such stereotype or they deliberately reject this stereotype when asked to report their perceptions.

6.4.2. Individual differences of stereotypes of empathy in scientists vary according to whether they are measured implicitly or explicitly

The second aim of the present study is to investigate individual differences in stereotypes of empathy in scientists by gender and major subject. In the present study, both implicit and explicit measures were used to assess stereotypes and we found that variations of stereotypes of empathy in scientists vary according to whether they were measured implicitly or explicitly. By comparing relevant evidence produced by implicit and explicit measurements, four main findings were generated. In this section, each finding is summarized and discussed with reference to existing social psychological theories as well as previous empirical findings.

6.4.2.1. Are women less stereotypical about empathy in scientists than men?

When comparing stereotypes between women and men, differences were only observed on explicit but not implicit stereotypes. Specifically, both women and men showed similar implicit bias against empathy in scientists. According to the *t*-test results, the average SSE-IAT score of women was not significantly different from those of men, indicating no gender difference in implicit stereotypes of empathy in scientists.

> *Finding 2a: Women and men showed similar levels of implicit bias against empathy in scientists, but women explicitly reported weaker stereotypes of empathy in scientists and more positive views about social skills in scientists than men.*

However, women scored significantly lower in SSE-Explicit and more negative in ISSOS than men, indicating that women reported weaker explicit bias against empathy in scientists and more positive self-reported views about social skills in scientists than men. The size of the gender effect, however, was small. These gender differences were summarized as below:

Such findings in the present study are in accordance with previous studies revealing that gender differences of attitudes vary according to the measures applied. The present study has successfully replicated the previous findings that men systematically reported higher explicit prejudice than women (Brewer, 2007; Dozo, 2015; Ekehammar et al., 2003; Hughes & Tuch, 2003). Nevertheless, the present study has added more mixed findings regarding gender differences in implicit attitudes to previous research. Ekehammar et al. (2003) found that women displayed systematically higher implicit racial prejudice than men. On the contrary, Dozo's (2015) meta-analysis of studies on prejudice against homosexuals and people with disabilities revealed that men consistently demonstrated a greater level of implicit prejudice as compared to women. Nevertheless, the current results suggest no gender difference in implicit stereotypes of empathy in scientists. It seems that gender differences in implicit attitudes may vary due to the content of the measured attitudes. Unlike the explicitly measured attitudes, there is no consistent female advantage in producing socially favourable implicit attitudes.

The ubiquitous and robust gender difference found in explicit attitudes with men reporting more prejudice than women may be explained by the theory that social gender roles have an influence on the channels that are available for women and men's expressions of attitudes (Ekehammar et al., 2003). According to gender stereotypes, men are expected to be more agentic – assertive, competitive, dominant and deference to facts over feelings. In contrast, women are expected to be more communal – selfless, caring, egalitarian, and emotionally expressive (e.g., Rudman & Glick, 2001; Rudman & Phelan, 2010; Witt & Wood, 2010). These traditional roles we ascribe to each gender reflect the patterns of behaviours we expect them to engage in. Some researchers claim that it is less acceptable for women to explicitly express prejudice and negative attitudes (e.g., Eagly, 2013; Wood & Eagly, 2012), yet men face greater social pressure to engage in prejudice and stereotyping (Dozo, 2015). Due to the different expectations about what is appropriate for each gender, women demonstrated more motivations to withhold prejudice and negative attitudes (Ratcliff et al., 2006). As such, it is understandable that in the present study women reported more positive attitudes toward empathy in scientists even though they harbour similar levels of implicit bias as their male counterparts.

6.4.2.2. Are science majors less stereotypical about empathy in scientists than humanities majors?

When comparing stereotypes between science majors and humanities majors, significant differences were only observed on explicit but not implicit stereotypes. Specifically, the averaged SSE-IAT scores of science majors was not significantly different from humanities majors', indicating similar levels of implicit bias against empathy in scientists between science majors and humanities majors. However, science majors scored significantly lower in SSE-Explicit and more negative in ISSOS than humanities majors, indicating weaker explicit bias against empathy and more positive views about social skills in scientists among science majors as compared with humanities majors. These differences by academic field were summarized as below:

> *Finding 2b: Science and humanities majors showed similar levels of implicit bias against empathy in scientists, but science majors explicitly reported weaker stereotypes of empathy in scientists and more positive views about social skills in scientists than humanities majors.*

Similar to the gender effect, the academic field effect has also been found to vary due to the measures of attitudes. As explained in the literature review, according to most theories on intergroup relations in social psychology (e.g., social identity theory; Tajfel & Turner, 2001; self-categorisation theory; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987), we human beings, as social animals, have a strong tendency to respond more positively to people from our ingroups than we do to people from outgroups, which is known as ingroup favouritism. On the basis of the ingroup favouritism theory, science majors were hypothesized to have weaker stereotypes of empathy in scientists than those studying humanities.

In the present study, the phenomenon of ingroup favouritism was only found for explicitly expressed attitudes but not implicit stereotypes. That is to say, although science majors reported positive attitudes about social skills in scientists, they still demonstrated an ingroup implicit bias against empathy in scientists. Such findings were consistent with the evidence from Nosek, Banaji and Greenwald's (2002) study of a large sample of White and Black participants' implicit and explicit racial attitudes via the Internet (Ns > 17,000). They found that African Americans exhibited no implicit in-group favouritism but White Americans exhibited strong implicit in-group favouritism. Nevertheless, African Americans explicitly expressed much stronger ingroup favouritism in self-reports. Furthermore, when playing a video game simulating a police chase, African American and White American participants were found equally likely to harbour implicit stereotypes associating Black with criminality and responded faster to shoot at the Black armed fictitious characters compared to the White (Correll et al., 2007). In some other studies, African Americans were even found to favour White Americans over their ingroup members implicitly but not explicitly (Ashburn-Nardo et al., 2003; Dasgupta, 2004). Similar findings were also obtained with regard to implicit gender stereotypes. For example, women were found as likely as men to implicitly associate women with communal traits and men with agentic traits (Banaji, Hardin, & Rothman, 1993; Blair, Ma, & Lenton, 2001) and women also implicitly favour male leaders over female leaders (Rudman & Kilianski, 2000). Together with all the existing evidence, the current findings of the equally negative implicit bias toward empathy in scientists between science and humanities majors add more support to the well-documented phenomenon of implicit outgroup favouritism (or sometimes, less ingroup favouritism) in the case of stigmatized or disadvantaged groups.

As introduced in the literature review, one potential explanation for the implicit outgroup favouritism is the system justification theory, which argues that people's intergroup attitudes and behaviours may sometimes reflect the tendency to legitimize existing social norms and hierarchies even at the expense of personal and group interest (Jost & Banaji, 1994; Jost et al., 2004). In other words, for the advantaged and dominant groups, the desire to preserve current social status (system justifying motive) and the desire to protect their self-esteem (ego-justifying motive) can work harmoniously together to produce ingroup favouritism implicitly and explicitly. However, for the disadvantaged and stigmatized group, the two motivations work in opposition – the desire to protect self-esteem should lead to ingroup favouritism, but the desire to maintain current social arrangements leads to outgroup favouritism. In this case, for members of the stigmatized groups, that are the participants studying science in the present study, there are two independent sources of implicit attitudes with opposite directions. The first source, consistent with social identity theory, relies on people's group membership as a meaningful source for self-esteem and should promote more positive attitudes toward ingroups relative to outgroups. The second source, consistent with the system justification theory, is the mainstream culture's imposition of common beliefs on particular groups (Dasgupta, 2004). Thus, for science majors, the supposed implicit likings for the ingroups (e.g., positive implicit attitudes toward empathy in scientists) could be attenuated or even transformed into negative views of their own empathetic ability by the cultural construal of their group (e.g., the image of socially deviant scientists), whereas for the humanities majors, implicit bias toward the outgroups (e.g., negative implicit attitudes toward empathy in scientists) might be exacerbated by the cultural construal of their group (e.g., people studying social sciences and humanities are better at interpersonal relations than scientists).

Such findings suggest that people don't always show more positive views about ingroup members as the social identity theory predicts. When implicit measures are applied, mixed attitudes toward ingroups can be found, suggesting that people, even those in the stigmatized groups, also possess a tendency to justify the existing status of social arrangements.

6.4.2.3. Are people with dissonant identities (i.e., women in science and men in humanities) less stereotypical about empathy in scientists than those with consistent identities (i.e., women in humanities and men in science)?

Moreover, when examining the interaction effect of gender and academic field on stereotypes of empathy in scientists, different gendered patterns were only observed in implicit, but not explicit, stereotypes between science and humanities majors.

Specifically, for science majors, women showed weaker implicit bias against empathy in scientists than their male counterparts; but for those majoring in humanities, it was men who instead showed weaker implicit bias against empathy in scientists than women.

Nevertheless, no interaction between gender and academic field was found on either SSE-Explicit or ISSOS scores. For both science and humanities majors, women scored lower in SSE-Explicit and more negative in ISSOS than men, meaning that women had weaker explicit biases as well as more positive views about empathy in scientists than men regardless of the academic field. These differences were summarized as below:

Finding 2c: Female science majors and male humanities majors showed weaker implicit bias against empathy in scientists than male science majors and female humanities majors, respectively. But for explicit attitudes, women consistently showed more positive views than men regardless of academic field.

As introduced in the literature review, according to the Social Identity Theory (SIT) of explicit cognitions (Tajfel & Turner, 2001), men majoring in science who have relatively lower self-evaluated empathy than their female counterparts should have been more motivated to show more positive views about empathy in scientists in order to promote their personal self-esteem in empathy. However, such prediction was not supported in the present study and women consistently displayed more positive views about empathy in scientists than men regardless of their identified academic field.

As discussed in previous Section 6.4.2.1, the consistent female advantage in producing favourable explicit attitudes may be attributed to the greater social pressure that women face to withhold their negative views and behave as more egalitarian and caring than men (Dozo, 2015; Eagly, 2013; Ratcliff et al., 2006; Wood & Eagly, 2012). Moreover, such female advantage in producing favourable explicit attitudes is also consistent with the robust female advantage in self-report empathy (Eisenberg & Lennon, 1983; Michalska et al., 2013; See Section 3.1.2.1 in the literature review for more details). If women, in general, perceive themselves as empathetic, it is not surprising that they would be reluctant to explicitly display negative attitudes towards others.

As for the implicit stereotypes, the current evidence revealed an interaction effect between gender and academic field on stereotypes, which adds more support to the Balanced Identity Theory (BIT) of implicit social cognition (Greenwald et al., 2002). As explained in the literature review, The BIT is grounded in the fundamental principle of cognitive consistency that individual's implicit attitudes, stereotypes, self-esteem, and self-concept need to be congruent (Cvencek et al., 2012). BIT anticipates that change in any one of the three sets of associations – self-attribute (e.g., *self* (*women) – empathetic*), self-identified group (e.g., *self (women)* - *science*), or stereotype of the identified group (e.g., *science – not empathetic*) – will induce balancing change in at least one of the others. BIT takes *associations* as its conceptual building blocks and treats self-evaluation as an associative connection of self to positive attribute, and the balance-congruity principle calls for the link between the self-associated group and positive attribute to be strengthened by the link of self to positive attribute (Cvencek et al., 2012; Greenwald et al., 2002). In other words, BIT emphasizes the transmit of the positive attribution of self-evaluation to self-identified groups. In contrast with the SIT's expectation that group members with low self-esteem should display stronger positive views about ingroups in order to promote their own social identity, BIT predicts the reverse – that those who have high self-esteem in a specific trait should develop stronger positive views about that trait for their ingroups.

Thus, for women in science majors, their self-attribute would be *self (female) – empathetic* and academic group identity would be *self (female) – science*. To keep their implicit perceptions about their academic group (i.e., *science*) consistent with their attribute gender identity (i.e., *empathetic*), BIT predicts the weakening of their implicit *science – not empathetic* stereotype. In other words, female science majors may view themselves as caring and empathetic (in order to confirm their social gender roles) thus automatically decline the negative evaluations of empathy in scientists at the unconscious level. Unlike their female colleagues, male science majors do not need to experience the clash of gender role and science identity, therefore it is easier for them to accept the traditional image of unsocial scientists. Evidence from the present study supports the BIT prediction and showed that female science majors had weaker implicit biases against empathy in scientists than their male counterparts. Such findings are also in accordance with Smyth and Nosek's (2015) study on gender-science implicit stereotypes in which they found that women majoring in more scientific disciplines had substantially weaker "scientist-is-male" implicit stereotypes.

As for humanities majors, women again transmit their positive self-evaluation of empathy to their identified humanities majors, thus when completing the SSE-IAT, they would associate *empathy* more strongly with *humanities* than with *science*, suggesting a stronger implicit bias against empathy in scientists than their male counterparts. As such, different gendered patterns were observed for science and humanities majors regarding their implicit stereotypes of empathy in scientists in the present study.

To sum up, the present study adds more mixed evidence for the Social Identity The0ry's predictions of intergroup explicit attitudes. The SIT hypothesis that men in science with low self-evaluated empathy would be more motivated to show positive attitudes toward empathy in scientists in order to promote their own self-esteem was not supported. Women consistently displayed more positive views about empathy in scientists in the present study. Nevertheless, evidence from the present study has lent more support for the Balanced Identity Theory of implicit cognition. Evidence about the variations in the implicit stereotypes by gender and major subject support the BIT theory that when individuals' self-evaluations conflict with the stereotypes of their identified academic fields, one may change their perceptions of the identified domain to keep a unified story of their implicit attitudes, stereotypes, and self-concept (Greenwald et al., 2002). That is to say, to balance out the clash between their gender roles (*self (women) – empathetic*) and the stereotypical image of their academic identities (*self (science) – not empathetic*), women with stronger science identity would decline the stereotypes and show weaker implicit bias against empathy in scientists than their male counterparts who do not need to experience the dissonance between their gender role and science identity. In addition, it is important to emphasize that evidence from the present study cannot imply any directional relations between science identity and perceptions of scientists. Though female scientists showed significantly weaker stereotypes of empathy in scientists both implicitly and explicitly, we still have no idea whether it is those who already possess positive views about scientists choose to enter science or it is the experience in studying science promotes their positive views about scientists.

6.4.2.4. Are there variations by specific major subject?

In the present study, participants were also asked to identify their specific major subject so that we could investigate the variations beyond the dichotomous groupings of science and humanities. As for the implicit stereotypes, no significant difference by major subject was found. However, regarding explicit attitudes, participants majoring in Life Science and Medicine showed significantly more positive explicit views about empathy and social skills in scientists than any other majors. Only Life Science & Medicine majors believed that scientists have comparable empathy with those who study humanities and displayed the strongest positive views about scientists' social skills. These differences by specific major subject were summarised as below:

Finding 2d: Unlike other majors who displayed stereotypical explicit bias against empathy in scientists, only participants majoring in Life Science & Medicine expressed a counter-stereotypical belief that scientists have comparable empathy with those who study humanities. They also reported strongest positive views about social skills in scientists. There is no significant difference by specific major subject in implicit bias.

Such findings may be interpreted as evidence for the assumption that the science-empathy stereotypes may change as the gender ratios in local environment change (Smyth & Nosek, 2015). As presented in the context of study chapter, national data has shown that in the UK higher education, women are well represented in Life Science & Medicine (61%) but remain underrepresented in Physical Science & Maths (40%) and Engineering & Technology (18%). Under such circumstances, when answering the question about empathy in scientists, given that students majoring in Life Science & Medicine have more opportunities to see female fellow scientists, who are usually perceived as more warm and friendly than male scientists in their environment (Steinke et al., 2007), it is not surprising that they would report more positive attitudes regarding empathy in scientists than participants majoring in other subjects.

Another possibility is that Life Science & Medicine majors indeed have better empathy than those majoring in other scientific subjects, as such they would easily spurn the stereotypical negative views about empathy in scientists. As a discipline focusing on different types of living creatures including human beings, trainings on empathy as well as ethics are an essential aspect of life science and medical education (Hirsch, 2007). As presented in the literature review, several empirical findings have documented an advantage in empathy for Life Science and Medicine students. For example, Thomson's et al. (2015) survey using self-report empathy questionnaire revealed equivalent levels of empathy among students in life science and social science, but lower levels of empathy in physics majors. Moreover, Khorashad et al. (2015) even found that medical students exceeded all the other majors in an emotion recognition task. These findings lend support to the claim that the positive explicit views of empathy in scientists among Life Science & Medicine majors may due to their own superiority in empathy.

However, though both explanations may have certain elements of truth, evidence from the present study cannot indicate any directional causality of these factors. It remains unknown whether it is the students who have an advantage in empathy choose to study Life Science & Medicine or it is their academic training in Life Science & Medicine improves their empathetic skills thus lead to reduced bias about empathy in scientists.

6.4.2.5. Does the gender gap in stereotypes of empathy in scientists vary by specific major subject?

Last but not least, gender differences in implicit and explicit stereotypes were also examined separately within each major subject. It was found that the gender gap in stereotypes of empathy in scientists was most pronounced for those majoring in Physical Science & Maths with women showing significantly weaker stereotype against empathy in scientists than men both implicitly and explicitly.

Specifically, although no gender difference in implicit stereotypes was observed in general, a significant gender gap was found for participants majoring in Physical Science in particular. Women majoring in Physical Science & Maths scored significantly lower in SSE-IAT than their male counterparts, meaning that they were less likely to automatically link empathy with humanities but sciences.

As for the explicit stereotypes, the only significant gender gap was also found for those majoring in Physical Science. Unlike their male counterparts who scored positively in SSE-Explicit, showing an explicit prejudice against empathy in scientists, women majoring in Physical Science scored close to 0 in SSE-Explicit, indicating that they believe scientists have comparable empathy with those who study humanities. As for the ISSOS results, women in Physical Science & Maths also showed significantly stronger positive views about social skills in scientists than their male counterparts. No such significant gender gap in explicit stereotypes of empathy was observed for other majors. The gender differences within major subjects were summarised as below.

Finding 2e: Gender gap in stereotypes of empathy in scientists was mostly pronounced for Physical Science & Maths majors. Women majoring in Physical Science & Maths showed significantly more positive views about empathy in scientists both implicitly and explicitly than their male counterparts. No such significant gender difference was observed for other majors.

Such findings may be explained by the evidence that, among women and men of comparably outstanding mathematical aptitude, women are more likely than men to have outstanding empathetic skills (Ceci & Williams, 2010). Valla et al., (2010) found that being in a scientific field of study was associated with poor empathising skills in men, but not in women. Based on this finding, Valla et al., (2010) proposed an interesting idea that men's choices of major subject may be determined more by their weakness than their strengths: men's choice of a science subject could be associated with weakness in empathising skills but women's choice of science subject are more associated with their strength in mathematical abilities. If this were the case, when women enter science field, they seem to do so because they are good at a particular scientific subject, whereas men may enter a scientific field more specifically when they have empathising difficulties regardless of their subject-related abilities (Valla et al., 2010).

Following this assumption, it is possible that men in Physical Science & Maths in fact have poor empathy while their female counterparts don't have such difficulty, therefore a gender disparity could emerge in their perceptions about empathy in scientists. Moreover, as presented before, Physical Science and Maths have remained male-dominant disciplines with almost 60% of the students are male. As such, men in these fields are more likely to identify with the stereotypical image of scientists, while their female counterparts have overcome the barriers of stereotypes of scientists thus are more likely to challenge the prejudice against empathy in scientists. Again, it is important to bear in mind that all these explanations need further empirical investigations and we cannot assume any directional causality between the environmental factors and personality traits.

To sum up, the present study revealed that individual differences in stereotypes of empathy in scientists varied by the measures of stereotypes used. Regarding implicit attitudes (SSE-IAT), no main effect of gender or major subject was detected, but an interaction effect was found with female science majors showed weaker implicit stereotypes of empathy in scientists than male science majors, whereas such gendered pattern was reversed for humanities majors. These findings have supported our hypothesis that people in untraditional roles may be less prone to stereotypes than those in traditional roles due to the fundamental principle of cognitive consistency among individuals' identities, beliefs and behaviours. Nevertheless, participants' implicit attitudes about empathy in scientists remained stereotypical regardless of their gender and major subject, showing that at the unconscious level, the tendency to legitimize existing social norms could be quickly internalized by people, even for those in the stigmatised groups at the expense of their personal and group interest (Ashburn-Nardo et al., 2003; Correll et al., 2007; Jost et al., 2004).

As for explicit attitudes, both main effects of gender and major subject were found. Women (and science majors) consistently expressed less stereotypical views about empathy in scientists (SSE-Explicit) and more positive views about social skills in scientists (ISSOS) than men (and humanities majors). The female advantage in producing socially favourable attitudes may be attributed to the social expectations for women to be egalitarian and caring. The positive attitudes toward empathy in scientists among science majors can be explained by ingroup favouritism.

In terms of variations among specific major subjects, we found that Life Science & Medicine majors expressed the most positive explicit attitudes about empathy in scientists among all the majors. Only the Life Science & Medicine majors reported that they believe science majors have comparable empathy with humanities majors, while all the other majors reported that scientists were less empathetic than those studying humanities. Such findings indicate that an environment with more female scientists as well as providing empathy training may efficiently reduce one's prejudice about empathy in scientists.

Last but not least, the most pronounced gender gap in attitudes toward empathy in scientists was found for Physical Science majors. Female physical science majors showed significantly more positive attitudes about empathy in scientists both implicitly and explicitly than their male counterparts, whereas no significant gender difference was discovered for other majors. Such findings may suggest that in a male-dominant scientific domain, men are more likely to identify with the stereotypical image of unsocial male scientists, while women in such area, in turn, may have overcome the barriers of stereotypes of scientists to study the traditionally "male" subject thus are more willing to challenge the prejudice against empathy in scientists.

6.4.3. Relationships between implicit and explicit stereotypes of empathy in scientists and career aspirations in science

6.4.3.1. Implicit-Explicit correlations

As presented in the literature review, there are three potential models regarding the relationship between implicit and explicit attitudes. At one extreme, some researchers argue that the implicit and explicit measures are completely independent of each other (i.e., double-dissociation model; Wilson et al., 2000). For example, Rudman & Phelan (2010) proposed that people can have dual attitudes toward the same object, one implicit and one explicit. From this perspective, IAT and self-reports tap exclusive underlying constructs, therefore, the correlations between them should be low to nonexistent. At the opposite extreme, implicit and explicit attitudes are viewed as measuring a single construct, only with different procedures (i.e., additive model; Fazio & Olson, 2003). From this point of view, attitudes can be compared to icebergs, with explicit attitudes residing above the surface of conscious control and implicit attitudes residing below it (Karpinski & Hilton, 2001). One implication of this argument is that, given the right conditions, implicit attitudes, explicit attitudes, and attitude-related behaviours should all correlate and divergence between them would be attributable to extraneous method-specific influences that are irrelevant to measured constructs (Nosek, 2007). The third intermediate possibility is that implicit and explicit measures assess constructs that are distinct, but related (i.e., interactive model; Conrey et al., 2005; Gawronski & Bodenhausen, 2014; Payne et al., 2008).

Evidence from the present study supports the third model: that the SSE-IAT and the self-reports assess distinct but related constructs. According to the statistical results, implicit bias toward empathy in scientists (SSE-IAT) showed a significant, weak, positive correlation (r = .14) with the explicit bias toward empathy in scientists (SSE-Explicit). Nevertheless, the contents of the implicit and explicit attitudes are actually different with implicitly measured attitudes reflecting negative views while the explicitly measured attitudes reflect positive views of empathy in scientists. Moreover, no significant correlation was found between implicit bias (SSE-IAT) and explicit attitudes about social skills in scientists (ISSOS). The lack of correlation between SSE-IAT and ISSOS may be attributed to the extraneous method-specific variance. As explained before, SSE-IAT and SSE-Explicit have similar structures that both of them measure a relative attitude toward empathy in scientists by comparing the strengths of associations between *science-empathy* and *humanities-empathy*. However, ISSOS assesses propositional opinions on statements such as "scientists are cooperative". Moreover, ISSOS focuses on the evaluations of *social skills* in scientists, which is a related but distinct concept, from *empathy* in SSE-IAT. Given that SSE-IAT and ISSOS differ in both methodological structures and conceptual substance, it is understandable that they were not correlated in the present study.

The varied correlations of SSE-IAT with different self-reports in the present study have added more evidence to the mixed findings of implicit-explicit correlations. Previous research has shown that, in some cases, the IAT and explicit measures can be strongly correlated (e.g., Asendorpf et al., 2002; Cunningham et al., 2001; Jellison et al., 2004; McConnell & Leibold, 2001; Wiers et al., 2002). But there are also a number of studies revealing only slight or moderate (but generally positive) correlations between the IAT and explicit measures (e.g., Egloff & Schmukle, 2002; Jong, Hout, Rietbroek, & Huijding, 2003; Karpinski & Hilton, 2001; Ottaway, Hayden, & Oakes, 2001). A recent meta-analysis revealed that across 126 independent correlations, implicit-explicit correspondence ranged from r = -.25 to r = .60, with an average implicit-explicit correlation of .19 (Hofmann et al., 2005). Given that the extent to which implicit and explicit attitudes are correlated varies widely across studies, issues of conditions under which implicit and explicit attitudes will correlate may be a more fruitful direction for future research.

In addition, it is worth noting that the current analysis also revealed a significant and positive correlation between explicit bias toward empathy in scientists (SSE-Explicit) and explicit attitude about social skills in scientists (ISSOS). Given that both of them are self-report measures and claim to assess similar constructs, it is not surprising that these two explicit measures correlate. This correlation can be interpreted as evidence confirming the convergent validity of the two questionnaires.

To sum up, SSE-IAT was found to significantly correlate with SSE-Explicit, but not ISSOS. Despite the distinctiveness in the content of implicit and explicit attitudes, the positive correlation between SSE-IAT and SSE-Explicit can be interpreted as evidence for the assumption that implicit and explicit attitudes are independent, but related, constructs. Moreover, the lack of correlation between SSE-IAT and ISSOS may be a result of the method-specific variance as well as the divergence in measured concepts. The varied implicit-explicit correlations lead to a new question about the conditions under which implicit and explicit measures will correlate for future research. Last but not least, a significant correlation between SSE-Explicit and ISSOS was found and can be seen as evidence supporting the convergent validity of the two questionnaires.

6.4.3.2. How do the implicit and explicit stereotypes of empathy in scientists predict science career aspirations?

In general, all the three measures of stereotypes of empathy were negatively correlated with science career aspirations, meaning that the stronger stereotypical views one possesses about empathy in scientists, the less likely one would pursue a scientific career in the future. To further understand which kind of stereotypes can serve as the better predictor for science career aspirations between the implicit and explicit ones, regression analyses were conducted.

Given that gender, age, academic field and degree level were also significantly correlated with science career aspirations in the present study, these variables were included as covariates. Among all the predictors including the covariates and the focal variables of implicit and explicit stereotypes of empathy in scientists, academic field was found the most important predictor for science career aspirations. That is to say, if one has already chosen to major in science in university, then it is highly possible for the person to pursue a scientific career in the future. In the present study, science majors showed significantly higher science career aspirations than humanities majors.

After controlling for the effect of academic field and other covariates (i.e., gender, age, and degree level), we found that it is the explicit stereotype, but not the implicit stereotype, of empathy in scientists, that serves as a significant predictor for science career aspirations. Among all the three measures of stereotypes, ISSOS was the only measure that could predict a significant unique variance in their career aspirations in science. This finding is consistent with previous evidence showing that ISSOS could positively predict US college students' science career intentions (Schneider, 2010). Neither SSE-IAT nor SSE-Explicit acted as a significant predictor for career aspirations.

These findings suggest that it is the endorsed explicit attitudes, but not the implicit bias at the unconscious level, that have an influence on one's career intentions. Specifically, for those who have already chosen to study science majors in university, regardless of the existence of the implicit prejudice against empathy in scientists, the more positive attitudes they express about social skills in scientists, the more likely they would want to continue a career path in the scientific field.

Such findings may also be interpreted as evidence supporting the *double dissociation* predictive model of implicit and explicit attitudes. As explained in the literature review, based on the three theoretical models of implicit and explicit cognitions, Perugini, (2005) generated three types of predictive models of implicit and explicit attitudes including: 1) *additive*, that the implicit and explicit measures reflect a single construct with different procedures, therefore attitudes inferred by the two types of measures explain different portions of variance in the behaviour (e.g., Fazio & Olson, 2003); 2) *double dissociation*, that the implicit and explicit attitudes are completely independent of each other, thus implicit attitudes predict spontaneous behaviour whereas explicit attitudes predict deliberative behaviour (e.g., Wilson et al., 2000); and 3) *multiplicative*, that the implicit and explicit attitudes are distinct but related, as such they interact in influencing behaviour (Strack & Deutsch, 2004). As making decisions about career choices is a controlled behaviour that people are sufficiently motivated to engage in deliberate reasoning, only the explicit attitude was found significantly predicting career aspirations in sciences in the present study, indicating that implicit attitudes were not interacting with explicit attitudes in predicting behaviours, therefore lending more support to the *double dissociation* predictive model.

However, it is important to note that although the SSE-IAT failed to predict science career aspirations in the present study, it may well predict some spontaneous behaviours in other situations. After all, the present study only assessed a single situation with only the controlled decision making as the outcome. It is certainly possible that SSE-IAT could predict some behaviours that tend to be out of one's conscious awareness such as people's tendency to interact with scientists as compared with humanities majors in a social situation when the information about major subject is primed.

Moreover, the current evidence should not be taken directly as evidence of the superiority of the *double dissociation* model. It is possible that there will be conditions where the *additive* or *multiplicative* models might provide a better explanation of the results. For example, research about people's defensive and secure self-esteem has revealed evidence to support the *multiplicative* model that implicit and explicit self-esteem work together to predict behaviours. Specifically, people with secure self-esteem (i.e., the congruence between high implicit self-esteem and high explicit self-esteem) have been found to be less narcissistic, to show less in-group bias, and to engage less in dissonance reduction compared to participants with defensive self-esteem (e.g., low implicit self-esteem but high explicit self-esteem; Jordan et al., 2003). Therefore, research needs to be extended to more behaviours and situations to test for alternative predictive models in the future.

6.4.4. Limitations of the present study and suggestions for future research

The present study has a number of limitations. Firstly, although SSE-IAT performed well in documenting individual differences in stereotypes in the present study, future research to particularly examine its psychometric properties would be desirable. Due to the time constraint of my PhD study, after we modified the SE-IAT to the SSE-IAT in Stage 2, the SSE-IAT was directly applied in Stage 3 without conducting a pilot study with an independent group of participants to establish its reliability and validity. Although the implicit-explicit correlations between SSE-IAT and SSE-Explicit found in the present study may be interpreted as evidence for the convergent validity of the SSE-IAT, there are many other important psychometric indicators that remain untested. For example, given that SSE-IAT failed to predict science career aspirations in the present study, future studies are needed to establish the predictive validity of the SSE-

IAT by examining its relationship with more spontaneous behaviours. Additional evidence supporting test-retest reliability, as well as its correlations with other implicit measures for SSE-IAT, is also desirable. It is also worth noting that "science and humanities" were used as the category labels throughout the present study. However, researchers should consider using the "sciences" as a more precise label for SSE-IAT in their future studies.

Moreover, it is worth noting that SSE-IAT still only revealed a relative implicit attitude that people automatically associate *empathy* more likely with *humanities* rather than *science*. Participants' absolute evaluation of the implicit association between *empathy* and *science* remains unknown. As demonstrated in Stage 2, we found a few psychometric problems with the standard SE-IAT for there is no clear opposite category to the focal attribute concept of *empathy*. Given that was the case, the controversial attribute *rationality* was dropped and the modified SSE-IAT with only the single attribute of *empathy* was applied in the last phase of the project (Stage 3). Nevertheless, the SSE-IAT still contains two target concepts (i.e., *science* and *humanities*), therefore it also only assesses a relative attitude by comparing the strengths of associations between *science-empathy* and *humanities-empathy*. In order to assess the absolute implicit attitude toward empathy in scientists, it would be necessary to assess the implicit *science-empathy* association independently from the implicit *humanities-empathy* association. This would require a reliable and valid single-attribute, single-target association test – that is, implicit procedures that purely assess associations between two concepts. However, reported reliabilities and validities of the existing available candidate methods such as the Go/No-Go Task (Bar-Anan & Nosek, 2014) and the Extrinsic Affective Simon Task (Houwer & Bruycker, 2007) are discouraging. To close this research gap, the development of new implicit measurements for the reliable evaluation of the strength of simple associations between one target concept and one attribute concept will be an important task for the years to come.

In addition, there are also limitations with the applied explicit measures. As for the

SSE-Explicit, though it was designed to resemble the structure of the SSE-IAT, it is comprised of only three questions and may not be a psychometrically sound tool to assess explicit stereotypes about empathy in scientists. An improved version of the SSE-Explicit may use all the stimulus items in the SSE-IAT to form a longer questionnaire. That is to say, instead of asking participants to rate the simple statement "scientists are less empathetic than those who study humanities" using only the category label of the attribute (i.e., *empathy*), a series of questions can be developed using the items representing the category that ask participants to specifically rate statements such as "scientists are less cooperative/considerate/affectionate/emotional than those who study humanities". Future studies using the improved version of the SSE-Explicit may reveal a stronger correlation with the SSE-IAT due to the closer structural correspondence between the two measures.

Moreover, as mentioned before, the present study only assessed a single situation with only one measurement for the implicit stereotypes. Though current evidence supports the theory that implicit and explicit attitudes are distinct but related constructs, we cannot exclude the possibility that studies using other implicit measures assessing different attitudes may tell a completely different story about the relations between implicit and explicit cognition. Also, when examining the predictive ability of the implicit and explicit stereotypes, we only included controlled decision making as the outcome behaviour. It would be desirable for future studies to apply different types of implicit and explicit measures to predict both related spontaneous and deliberate behaviours when examining the predictive models of the implicit and explicit attitudes.

Limitations also exist with regard to the study design. Given that scientists are often depicted as men (Cai et al., 2016; Miller et al., 2018) and men are often seen as socially insensitive (Hall & Mast, 2008), it remains unclear if the implicit bias against empathy in scientists found in the present study is actually a byproduct of the prejudice against empathy in men. To control for the potential confounding gender effect, future studies may use pictures of both female and male scientists to replace words as stimuli in the

IAT. Moreover, there may also exist an interactive gender effect between agent and target. That is to say, female and male participants' attitudes toward empathy in scientists may differ by the gender of the target scientists. It would be ideal to conduct a study that systematically assesses both female and male participants' implicit and explicit attitudes toward empathy in both female and male scientists.

In addition, it is necessary to point out that the present study, as a cross-sectional study, lacks the statistical power to substantiate the direction of the relationships under investigation. The regression analyses in the present study could only indicate how changes in the independent variables (i.e., stereotypes of empathy in scientists) are associated with changes in the dependent variables (i.e., science career aspirations), and did not imply any causal relations. It is very likely that stereotyping of scientists and science career aspirations are bi-directional in influence. In order to account for the reciprocal influences entailed in the stereotyping of scientists – aspirations in science interaction, future studies could adopt a longitudinal study design with more advanced statistical methods, such as growth curve modelling (Curran, Obeidat, & Losardo, 2010).

Chapter 7 General Discussion

In this final chapter, a brief summary of the main findings of the current study will be presented. Major contributions and implications will also be discussed. At last, limitations and directions for future research are noted.

7.1. Summary of findings

This thesis is based upon three stages of research that embraced a systematic development of a new Implicit Association Test (IAT), a test of the psychometric properties of the IAT in order to address its psychometric issues. The test was modified before being applied in the main research to ensure its soundness for gauging participants' unconscious bias towards empathy in scientists. In the third research stage, both the modified newly developed IAT and questionnaires were applied to assess individual differences in both implicit and explicit stereotypes of empathy in scientists as well as their potential relationship with aspirations for science careers among university students in the UK. Specifically, the present study sought to answer the following research questions:

1. What are the implicit and explicit stereotypes of empathy in scientists among UK university students nowadays?
2. What are the individual differences in the implicit and explicit stereotypes of empathy in scientists by gender and major subject?
3. How do the implicit and explicit stereotypes of empathy relate to students' career aspirations in science?

Regarding the first research question, evidence from Chapter 6 indicates that participants still held an implicit bias against empathy in scientists but explicitly reported positive views regarding empathy and social skills in scientists. The discrepancy found in the context of implicit and explicit perceptions of empathy in scientists suggests that university students possess strong default assumption of "socially clumsy" scientists but are aware that these are stereotypes, thus deliberately reject them in self-reports. This nuanced effect of students consciously being aware of stereotyping scientists' interpersonal skills while at the same time reproducing them in a spontaneous manner shows how difficult it is to change old, sturdy stereotypical image of scientists. As Tintori & Palomba (2017) has put forth in *Turn on the light on science*: "In spite of the reality…For centuries, the Western model of science has been simple: we relied on geniuses; our most revolutionary breakthroughs have typically emerged from individuals, working by themselves. This is not true anymore, but the false picture has remained in people's minds, becoming a deep-rooted stereotype" (Tintori & Palomba, 2017, p. 58).

In terms of the second research question, we found that variations in stereotypes of empathy in scientists also varied according to whether they were measured implicitly or explicitly. As for participants' unconscious bias towards empathy in scientists, findings from the present study indicated no significant difference by gender or major subject. That is to say, regardless of their gender or major subject, all participants exhibited similar levels of implicit bias against empathy in scientists. Nevertheless, when it comes to self-reporting explicit attitudes, women as well as science majors reported significantly more positive views of empathy in scientists than men and humanities majors, respectively. Such findings are in line with previous research showing that the female advantage in producing socially favourable attitudes may be partially owing to social expectations for them to be more egalitarian and caring than men (Doze, 2015). Although women explicitly reported more positive views about scientists than men, they harboured similar implicit biases against empathy in scientists to men.

Moreover, the present study also revealed that science majors showed similar levels of implicit bias against empathy in scientists as their counterparts majoring in humanities.

The well-documented in-group favouritism was found for explicit attitudes but not implicit attitudes. Although science majors demonstrated more positive views of empathy in scientists than humanities majors, both possessed similar levels of negative attitudes toward scientists' empathy in an uncontrolled, automatic and spontaneous way. Such findings provide evidence that, even being in the stigmatised group, science majors may have incorporated and internalised the prevailing cultural stereotypes of scientists at the unconscious level. As described by Jost and Banaji (1994), such in-group bias of implicit cognition may come from the system-justifying effect, which suggests a tendency to maintain and legitimise existing social systems and cultural views. Our findings on the implicit bias towards own groups infers that unconscious stereotypes may be particularly insidious.

Furthermore, when looking at the interactive effect of gender and major subject on stereotypes, we found that women majoring in science, especially those majoring in "hard science" (e.g., physics, mathematics), showed relatively stronger counter-stereotypical views about empathy in scientists than general participants, both implicitly and explicitly. As seen in much previous research, women usually rate themselves higher in self-reported empathy (e.g., Baron-Cohen & Wheelwright, 2004; Davis, 1983) and female scientists are also often viewed as more warm and caring than male scientists (e.g., Thomson et al., 2015; Willer, Wimer, & Owens, 2015), so it is not surprising that female science majors would be more willing to challenge the stereotypical idea of scientists being socially clumsy in order to reconcile their gender role and science identity (Rudman & Phelan, 2010). In addition, as described by Mujtaba and Reiss (2014), girls aspiring to take mathematics in school are more likely to have competitive personality traits than boys choosing mathematics. No wonder those competitive girls who continue to pursue science and mathematics at a higher education level would challenge stereotypes of scientists and demonstrate more positive views surrounding empathy in scientists in the present study.

In addition, we also observed that participants majoring in Life Science & Medicine, regardless of their gender, showed significantly weaker explicit stereotypes about empathy in scientists than all other majors. We speculated the positive attitudes among Life Science & Medicine majors may be attributed to two factors. Firstly, women are overrepresented in the Life Science & Medicine fields given that students from these fields have more opportunity to observe the counter-stereotypical female scientists, so they would more likely explicitly reject the traditional image of scientists. Secondly, considering training on empathy and ethics are an essential part of their education, students majoring in Life Science & Medicine may indeed be more empathetic than other majors (Billington et al., 2007; Khorashad et al., 2015), therefore would have more positive views on empathy in scientists. However, it is worth mentioning that evidence from the present study only demonstrated that certain individuals had more positive views about empathy in scientists than others, and hence why they held these counter-stereotypical views merits further empirical investigation - all the potential explanations herein are just hypotheses.

Finally, regarding the third research question, we also examined the relationship between implicit and explicit stereotypes and career aspirations in science. Confirming the double-dissociation theoretical model of implicit and explicit cognition, only the explicit stereotype of social skills in scientists was found significantly correlated with participants' career aspirations in science. According to the double-dissociation theory, implicit attitudes predict spontaneous behaviour, whereas explicit attitudes predict deliberative behaviour (Asendorpf et al., 2002). Taking into account that selecting a career path is no doubt an important decision such that every individual in such a scenario would engage in considerable amount of introspection, it is understandable that only the explicit stereotype is related. However, it is important to bear in mind that it was the participants' major subject choice at the higher education level that accounted for the most variance in their career choice in the present study (66%), while explicit stereotypes of empathy in scientists only accounted for a very small proportion (1%). That is to say, if participants have already chosen to study science in universities, they are very likely to continue their career path in science. Therefore, a more promising research direction may be to investigate the correlation between stereotypes of empathy in scientists and major choice for A-level students.

7.2. Contributions and implications

The present study has made contributions to the research fields of the IAT, social psychology of implicit and explicit stereotypes as well as the gender gap in math-intensive STEM fields.

Regarding contributions to the research of the IAT, the current work developed and modified a new IAT to test implicit stereotypes of empathy in scientists. This thesis summarised a systematic procedure of developing an IAT. As shown in the present study, the IAT effect can be contaminated by many construct-unrelated variables, and many problems may arise during the process of developing an IAT, obliging researchers to show scientific responsibility in using the IAT paradigm to customize their own implicit tests. Therefore, it is deemed important for researchers to provide a holistic approach for designing an IAT and highlights several important remedies for the respective contaminants. A three-step approach to develop and computerise an IAT has been presented in this thesis as a robust example for future researchers to follow the relevant steps in order to create their own IAT.

In addition, when examining the psychometric properties of the preliminary SE-IAT in Stage 2, a major limitation of the traditional IAT in the assessment of contrasting categories has been highlighted (Karpinski & Steinman, 2006; Lane et al., 2007; Nosek, Greenwald, et al., 2007). When using the traditional IAT, it is unclear if an individual's score is based on a strong preference toward the attribute of interest (e.g., empathy) in one group, or an aversion toward the contrast attribute (e.g., rationality) in the other group, or a combination of both preference and aversion. Unlike "male" and "female", which are the two dimensions of one concept "gender", "rationality" and "empathy" are two independent concepts, therefore leading the SE-IAT to measure more than one construct. Instead of focusing on associations between *empathy* and *science*, it lumps the evaluations of associations of *science-empathy*, *science-rationality, humanities-empathy* and *humanities-rationality* altogether, thereby creating much "noise" in the test and potentially reducing its reliability and validity. To address this limitation, we employed a variant of the IAT – the Single-category IAT – to modify the SE-IAT to the SSE-IAT by dropping the category of *rationality*. As a relatively novel implicit measurement, there are only a limited number of studies using the Single-category IAT method to date (Bluemke & Friese, 2008; Karpinski & Steinman, 2006; Penke et al., 2006).

The present Stage 3 study has indicated that the Single-category IAT can serve as a reliable measurement to investigate individual differences in attitudes. The newly-developed SSE-IAT adds to the instruments afforded to other researchers looking at unconscious attitudes that may affect subject and thence career choice at university. In addition, SSE-IAT can be translated into different languages and used to examine cross-cultural differences in attitudes toward empathy in scientists. For example, students from more industrialised countries (e.g., UK, USA, Japan) often report weaker interest in learning science than those from developing country (e.g., Nigeria, India, Ghana; Sjøberg & Schreiner, 2012). It is very likely that the varied scientific aspirations are related to cross-cultural differences in stereotypes of scientists. Future studies utilising SSE-IAT and questionnaires to investigate cross-cultural differences in both implicit and explicit attitudes may reveal many interesting findings.

Regarding theoretical contributions, the present study has examined the correlations between implicit and explicit stereotypes and revealed evidence supporting the *interactive* theoretical model that claims implicit and explicit cognitions are distinct but related constructs. Despite the divergence in the directions of implicit and explicit stereotypes (negative implicit views vs. positive explicit views), we found a weak but significant correlation between the SSE-IAT and SSE-Explicit scores. Such a correlation may be interpreted as evidence that implicit and explicit attitudes are distinct but related constructs.

Moreover, we found that individual differences in stereotypes of empathy in scientists varied according to whether it is measured implicitly or explicitly. Predictions based on some classic social psychological theory, such as the seemingly ubiquitous in-group favouritism according to the social identity theory (Tajfel & Turner, 2001), was only supported for explicit but not implicit stereotypes in the present work. Although researchers have proposed different models for the implicit and explicit cognitions, there is consensus that implicit manifestations of attitudes and stereotypes do exist and are different from explicit attitudes and stereotypes (e.g., Dovidio et al., 2010; Fazio & Olson, 2003; Gawronski et al., 2007). It is of great necessity for future psychologists who are interested in studying stereotypes and prejudices to differentiate between implicit and explicit biases. Furthermore, classic social psychological theories about stereotyping proposed before the emergence of implicit measures also need to be revisited using both implicit and explicit measurements of cognitions.

Last but not least, regarding contributions to the current knowledge, the present study uncovered that although students reported positive views about social skills in scientists, there remained a pervasive and stable implicit bias against empathy in scientists among university students in contemporary UK society. Such findings indicate that efforts over the past decades to break down the stereotypes of scientists have raised awareness of the existence of stereotypes of scientists. However, it is uneasy to change the old, stable, implicit bias against empathy in scientists, deeply rooted in long-established socialization experiences.

As noted by Leon (2014), even though there are more and more women playing STEM literate characters in movies or television series, they are often portrayed as merely a feminized version of their male counterparts, with all the stereotypical traits often associated with scientists: wearing glasses with thick lenses, obsessed with difficult scientific questions, and isolated. These negative traits may even be amplified when depicted as incompatible with "feminine" traits, such as beauty, fashion, social skills, and sexual desirability in mass media (Tintori & Palomba, 2017). With this, there is often a marked difference between the "clever" and the "good-looking" female protagonists. For example, the popular TV sitcom, *The Big Bang Theory* (Cendrowski, 2007), offers a contrasted image between the well-dressed, funny, and intuitive waitress/actress, Penny, and the neuroscientist, Amy, who is depicted as smart, quirky, out-of-fashion, spectacled, and working alone with monkeys in her lab. These portrayals of women in science may also reinforce negative stereotypes regarding scientists' empathy.

Existing initiatives and interventions (e.g., the Women in Science and Engineering (WISE) campaign), which is based on the message, "STEMs are for women too", may imply that women who study STEM subjects are unusual and, in turn, bolsters the existing stereotypes of scientists being "special" (smart, isolated, focused on their work, crazy, and dangerous). A more accurate message that should be conveyed to the public is that "STEMs are for everyone". A diverse set of images of scientists, not just "quirky geniuses", depicted in all forms of media (e.g., textbooks, children's literature, TV shows, movies) may help expand people's idea about who can be scientists as well as realise that scientists are normal people too. Another more direct way to dismantle the stereotypes of scientists is to create more opportunities for scientists themselves to get in touch with the public. As stated in the Public Attitudes to Science report, considerable numbers of the UK citizenry think scientists are poor at communication and that they believe they see and hear too little about science, preferring scientists to talk more about the social implications of their work (Ipsos, 2014). If people are able to have chances to meet scientists in person, such interactions are very likely to change the narrative with respect to scientists and provide strong role models for students aspiring to a scientific career.

7.3. Limitations and directions for future research

Considering the time and resource constraints of the PhD project, it is inevitable that a number of limitations of the present study would need to be addressed in the future. It is of great importance to take these limitations into account when interpreting the findings of the current work. The following sections explain the limitations of this investigation and also offers suggestions regarding directions for future research.

First of all, the present study adopted the stratified random sampling method and participants were selected based on their major subject from the Russell Group universities in the UK. Consequently, it is sensible to consider the demographic characteristics of the current sample when interpreting the findings of the present study, instead of over-generalising the results without mentioning the formation of the sample. Given that all participants were recruited from elite universities, it is possible that very different findings would be yielded if the research were to be carried out with the general public or even with other student populations. According to the *Public Attitudes to Science* 2014 survey report, the UK public felt "uninformed with science" and reported negative attitudes surrounding the communication skills of scientists and engineers (Ipsos, 2014). As such, it is reasonable to hypothesize that those who are not affiliated with higher education institutions might have fewer opportunities to interact with scientists and possess stronger stereotype-congruent views about empathy in scientists than the current sample. Future studies may extend the present research to participants with varied education levels and professional backgrounds.

Moreover, further research may also be beneficial if we were to target younger students. The present study has found that stereotypes of empathy in scientists varied by major subject in university. Given that the university subject options are dependent on school students' selected A-level subjects in the UK education system, we can guess that the link between subject choices and stereotypes of empathy in scientists may stem from earlier phases in school. It is not unrealistic to suggest that perceptions regarding empathy in scientists may have an impact on A-level students' school subject choices as well as their future major choices in university. Thus, further empirical evidence is required to examine the hypothesized relationships between stereotypes of empathy in scientists, A-level subject choices and university course intentions among secondary school students.

Secondly, throughout the thesis, we talked about gender as binary concepts of women and men. However, this traditional dichotomous model of gender has been criticised for failing to represent the experience of individuals who claim neither an unambiguously female nor male identity (Diamond & Butterworth, 2008). Gender binarism classifies gender into two distinct entities, assuming men adhere to traditional masculine traits and women to those that are feminine (Kay & Stewart, 2001). Nevertheless, not all women are feminine and not all men are masculine. It is possible that certain people identify with traits opposite to their biological sex (e.g., tomboys, which refer to girls who exhibit characteristics or behaviours considered typical of boys; Paechter, 2010) and others identify with both masculine and feminine traits such that they regard themselves as the "third gender" (Diamond & Butterworth, 2008). Future studies may differentiate between participants' biological sex and their feminine/masculine characteristics to gain more detailed profiles about participants' gender identities. As a symbolically masculine field, it is possible that STEM may not simply deter all women, but actually deter women who are hyper feminine (DeWitt et al., 2013). Similarly, there is also the potential that it may not be women who possess stronger counter-stereotypical explicit views than men, but feminine individuals (including men who identify themselves as more feminine) who are less stereotypical than masculine individuals (including women who identify themselves as more masculine). Future investigations adopting more specific conceptualisations of gender may reveal a more nuanced picture about individual differences in stereotyping and scientific aspirations.

Moreover, regarding limitations with respect to the applied measurements, it is important to note that the implicit measures are not a process-pure reflection of activated associations and only provide a proxy for participants' implicit cognition (Gawronski et al., 2007). In their quad model, Conrey et al. (2005) proposed a total of four different mental processes that all contribute to performance regarding implicit measures: 1) the activation of an association; 2) the ability to distinguish between correct and incorrect responses; 3) the ability to successfully overcome the automatically activated bias in favour of the correct response; and 4) the influence of a general response bias that might lead responses in the absence of other available response guides. Several studies using the quad model have found that overcoming bias has a significant impact on participants' performance on implicit measures and the use of a simple scoring algorithm may conceal various underlying processes (e.g., Gonsalkorale, Sherman, & Klauer, 2014; Klauer, 2005). For example, although younger and older participants demonstrated equivalent levels of anti-old bias on an IAT, application of the quad model revealed that the anti-old associations were actually less activated among older participants versus those that were younger, - the older participants were less able to overcome these activations when performing the task (Gonsalkorale et al., 2014). Thus, when interpreting the implicitly measured attitudes, the lack of process purity of the measurements should be taken into consideration. Future studies using the quad model to analyse IAT effects may be desirable in order to discuss the alternative interpretations of the outcomes.

Finally, the current research of stereotypes of empathy in scientists can be extended in multiple directions. Based on the current findings of the persistence of implicit prejudice against empathy in scientists, further questions would be why and how such implicit stereotypes remain unchanged. Qualitative research would be of value to seek in-depth understanding about how people form their stereotypes of empathy in scientists. Moreover, several explanations have been proposed for the individual differences in stereotypes found in the present study, but empirical studies are required to examine these hypotheses. For example, it was proposed that the significant gender gap in stereotypes of empathy in scientists among Physical Science & Math majors might be attributable to the gender gap in their actual empathetic ability. As such, a study assessing relationships between physical scientists' empathetic capacity and their attitudes about scientists' social skills might uncover some intriguing findings. However, such study can still not imply any directional causality between empathetic ability and stereotyping. Experiments involving empathy training may uncover whether improvement in empathy can result in a weakening in stereotyping. Finally, a recent meta-analysis of Draw-a-Scientist studies found that children's depictions of scientists have become more gender diverse over time, but children still associate scientists with men as they grow older (Miller et al., 2018), indicating that the change in the sociocultural context within which participants are reared can impact their stereotypes about scientists. Future research could investigate whether different generations have varying stereotypes about empathy in scientists. Will stereotypes of empathy in scientists increase with age? This is an important avenue of research to pursue to understand factors related to the formation as well as development of stereotypes.

References

Ajzen, I., & Fishbein, M. (1977). Attitude-Behavior Relations: A Theoretical Analysis and Review of Empirical Research. *Review of Empirical Research, Psychological Bulletin*, *84*(5), 888–918.

Allemand, M., Steiger, A. E., & Fend, H. A. (2015). Empathy development in adolescence predicts social competencies in adulthood. *Journal of Personality*, *83*(2), 229–241.

Allport, G. W. (1954). The nature of prejudice. Oxford, England: Addison-Wesley.
Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: judging by features and categories. Journal of Personality and Social Psychology, 83(1), 5–25.

Alper, J. (1993). The Pipeline Is Leaking Women All the Way Along. *Science*, *260*(5106), 409–411.

Archer, J., & McDonald, M. (1991). Gender roles and school subjects in adolescent girls. *Educational Research*, *33*(1), 55–64.

Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2012). Science Aspirations, Capital, and Family Habitus: How Families Shape Children's Engagement and Identification With Science. *American Educational Research Journal*, *49*(5), 881–908.

Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2013). 'Not girly, not sexy, not glamorous': primary school girls' and parents' constructions of science aspirations 1. *Pedagogy, Culture & Society*, *21*(1), 171–194.

Archer, L., Osborne, J., DeWitt, J., Wong, B., & Willis, B. (2013). *ASPIRES Young people's science and career aspirations, age 10 - 14* (ASPIRES Project Report). Department of Education & Professional Studies, King's College of London.

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or "would Jesse Jackson 'fail' the Implicit Association Test?". *Psychological Inquiry*, *15*(4), 257–278.

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: the case of shy behavior. *Journal of Personality and Social Psychology*, *83(*2), 380–393.

Ashburn-Nardo, L., Voils, C. I., & Monteith, M. J. (2001). Implicit associations as the seeds of intergroup bias: how easily do they take root? *Journal of Personality and Social Psychology*, *81(*5), 789–799.

Ashburn-Nardo, Leslie, Knowles, M. L., & Monteith, M. J. (2003). Black Americans' Implicit Racial Associations And Their Implications for Intergroup Judgment. *Social Cognition*, *21*(1), 61–87.

Astin, H. S. (1975). Sex differences in mathematical and scientific precocity. *The Journal of Special Education*, *9*(1), 80–91.

Auyeung, B., Allison, C., Wheelwright, S., & Baron-Cohen, S. (2012). Brief Report: Development of the Adolescent Empathy and Systemizing Quotients. *Journal of Autism and Developmental Disorders*, *42*(10), 2225–2235.

Back, E., Jordan, T. R., & Thomas, S. M. (2009). The recognition of mental states from dynamic and static facial expressions. *Visual Cognition*, *17(*8), 1271–1286.

Back, M. D., Schmukle, S. C., & Egloff, B. (2005). Measuring Task-Switching Ability in the Implicit Association Test. *Experimental Psychology*, *52*(3), 167–179.

Baenninger, M., & Newcombe, N. (1995). Environmental input to the development of sex-related differences in spatial and mathematical ability. *Learning and Individual Differences*, *7*(4), 363–379.

Baker, S. T., Friedman, O., & Leslie, A. M. (2010). The Opposites Task: Using General Rules to Test Cognitive Flexibility in Preschoolers. *Journal of Cognition and Development*, *11*(2), 240–254.

Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger, I. Nairne J. S., I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 117–149). Washington, DC: American Psychological Association.

Banaji, M. R., Hardin, C., & Rothman, A. J. (1993). Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology*, *65*(2), 272–281.

Banerjee, M. (1997). Hidden Emotions: Preschoolers' Knowledge of Appearance-Reality and Emotion Display Rules. *Social Cognition*, *15*(2), 107–132.

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes towards homosexuality: reliability, validity, and controllability of the IAT. *Zeitschrift Für Experimentelle Psychologie: Organ Der Deutschen Gesellschaft Für Psychologie*, *48*(2), 145–160.

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods*, *46*(3), 668–688.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*(2), 230–244.

Barlow, M. R., & Cromer, L. D. (2006). Trauma-relevant characteristics in a university human subjects pool population: gender, major, betrayal, and latency of participation. *Journal of Trauma & Dissociation: The Official Journal of the International Society for the Study of Dissociation*, *7*(2), 59–75.

Barman, C. R. (1997). Students' Views of Scientists and Science: Results from a National Study. *Science and Children*, *35*(1), 18–23.

Barnhoorn, J. S., Haasnoot, E., Bocanegra, B. R., & van Steenbergen, H. (2015). QRTEngine: An easy solution for running online reaction time experiments using Qualtrics. *Behavior Research Methods*, *47*(4), 918–929.

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes. Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science*, *17*(1), 53–58.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences*, *6*(6), 248–254.

Baron-Cohen, S., Knickmeyer, R. C., & Belmonte, M. K. (2005). *Sex Differences in the Brain: Implications for Explaining Autism. Science*, *310*(5749), 819–823.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46.

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., & Wheelwright, S. (2003). The systemizing quotient: an investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1430), 361–374.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. *Journal of Child Psychology and Psychiatry*, *42*(2), 241–251.

Barrett, G. V., Phillips, J. S., & Alexander, R. A. (1981). Concurrent and predictive validity designs: A critical reanalysis. *Journal of Applied Psychology*, *66*(1), 1–6.

Beardslee, D. C., & O'Dowd, D. D. (1961). The College-Student Image of the Scientist Scientists are seen as intelligent and hard-working but also as uncultured and not interested in people. *Science*, *133*(3457), 997–1001.

Benbow, C. P., Lubinski, D., Shea, D. L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning ability at age 13: their status 20 years later, *Psychological Science*, *11*(6), 474–480.

Berg, K., Blatt, B., Lopreiato, J., Jung, J., Schaeffer, A., Heil, D., … Hojat, M. (2015). Standardized Patient Assessment of Medical Student Empathy: Ethnicity and Gender Effects in a Multi-Institutional Study. *Academic Medicine*, *90*(1), 105–111.

Bethlehem, J., & Biffignandi, S. (2011). *Handbook of Web Surveys*. John Wiley & Sons.

Beyer, S. (2015). Women and Science, Technology, Engineering, and Math (STEM). In D. S. Dunn (Ed.), *Oxford Bibiographies in Psychology*. New York: Oxford University Press.

Billington, J., Baron-Cohen, S., & Wheelwright, S. (2007). Cognitive style predicts entry into physical sciences and humanities: Questionnaire and performance tests of empathy and systemizing. *Learning and Individual Differences*, *17*(3), 260–268.

Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: the moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, *81*(5), 828–841.

Blair, Irene V. (2002). The Malleability of Automatic Stereotypes and Prejudice. *Personality and Social Psychology Review*, *6*(3), 242–261.

Blair, Irene V., Judd, C. M., Havranek, E. P., & Steiner, J. F. (2010). Using Community Data to Test the Discriminate Validity of Ethnic/Racial Group IATs. *Zeitschrift Fur Psychologie, 218*(1), 36–43.

Bloom, P. B.-N. (2008). Respondent Fatigue. In P. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods*. California: Sage Publications, Inc.

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, *77*(1), 25–31.

Bluemke, M., & Friese, M. (2008). Reliability and validity of the Single-Target IAT (ST-IAT): assessing automatic affect towards multiple attitude objects. *European*

Journal of Social Psychology, *38*(6), 977–997.

Bosson, J. K., Swann Jr., W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology*, *79*(4), 631–643.

Brewer, M. B. (2007). The social psychology of intergroup relations: Social categorization, ingroup bias, and outgroup prejudice. In *Social psychology: Handbook of basic principles*, 2nd ed (pp. 695–715). New York: Guilford Press.

British Educational Research Association. (2018). *Ethical Guidelines for Educational Research* (4th Edition). London.Retrieved from https://www.bera.ac.uk/researchers-resources/publications/ethical- guidelines-for-educational-research-2018

Brown, J. R., Donelan-McCall, N., & Dunn, J. (1996). Why Talk about Mental States? The Significance of Children's Conversations with Friends, Siblings, and Mothers. *Child Developm*ent, *67*(3), 836–849.

Brunel, F. F., Tietje, B. C., & Greenwald, A. G. (2004). Is the Implicit Association Test a Valid and Valuable Measure of Implicit Consumer Social Cognition? *Journal of Consumer Psychology*, *14*(4), 385–404.

Buchner, A., & Wippich, W. (2000). On the reliability of implicit and explicit memory measures. Cognitive Psychology, 40(3), 227–259.

Cai, H., Luo, Y. L. L., Shi, Y., Liu, Y., & Yang, Z. (2016). Male = Science, Female = Humanities Both Implicit and Explicit Gender-Science Stereotypes Are Heritable. *Social Psychological and Personality Science*, *7*(5), 412–419.

Cai, H., Sriram, N., Greenwald, A. G., & McFarland, S. G. (2004). The Implicit Association Test's D measure can minimize a cognitive skill confound: Comment on McFarland and Crouch (2002). *Social Cognition*, *22*(6), 673–684.

Calero, C. I., Salles, A., Semelman, M., & Sigman, M. (2013). Age and gender dependent development of Theory of Mind in 6- to 8-years old children. *Frontiers in Human Neuroscience*, 7, 281.

Callegaro, M. (2008). Social Desirability. In P. Lavrakas, *Encyclopedia of Survey Research Methods*. 2455 Teller Road, Thousand Oaks California 91320 United States of America: Sage Publications, Inc.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

Capage, L., & Watson, A. C. (2001). Individual Differences in Theory of Mind, Aggressive Behavior, and Social Skills in Young Children. *Early Education and Development*, *12*(4), 613–628.

Carli, L. L., Alawa, L., Lee, Y., Zhao, B., & Kim, E. (2016). Stereotypes About Gender and Science: Women ≠ Scientists. *Psychology of Women Quarterly*, *40*(2), 244–260.

Carroll, J. M., & Chiew, K. Y. (2006). Sex and Discipline Differences in Empathising, Systemising and Autistic Symptomatology: Evidence from a Student Population. *Journal of Autism and Developmental Disorders*, *36*(7), 949–957.

Ceci, S. J., & Williams, W. M. (2010). Sex Differences in Math-Intensive Fields. *Current Directions in Psychological Science*, *19*(5), 275–279.

Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences*, *108*(8), 3157–3162.

Ceci, Stephen J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in Academic Science: A Changing Landscape. *Psychological Science in the Public Interest*, *15*(3), 75–141.

Cek, I. (2015). *Modernizing the Implicit Assocaition Test* (Unpublished doctoral dissertation). University of Cambridge, Cambridge, United Kingdom.

Chaiken, S., & Trope, Y. (1999). *Dual-process Theories in Social Psychology*. Guilford Press.

Chakrabarti, B., Dudbridge, F., Kent, L., Wheelwright, S., Hill-Cawthorne, G., Allison, C., … Baron-Cohen, S. (2009). Genes related to sex steroids, neural growth, and social-emotional behavior are associated with autistic traits, empathy, and Asperger syndrome. *Autism Research*, *2*(3), 157–177.

Chambers, D. W. (1983). Stereotypic images of the scientist: The draw-a-scientist test. *Science Education*, *67*(2), 255–265.

Chapman, E., Baron-Cohen, S., Auyeung, B., Knickmeyer, R., Taylor, K., & Hackett, G. (2006). Fetal testosterone and empathy: evidence from the empathy quotient (EQ) and the 'reading the mind in the eyes' test. *Social Neuroscience*, *1*(2), 135–148.

Charman, T., Ruffman, T., & Clements, W. (2002). Is there a Gender Difference in False Belief Development? *Social Development*, *11*(1), 1–10.

Cheryan, S., Siy, J. O., Vichayapai, M., Drury, B. J., & Kim, S. (2011). Do Female and Male Role Models Who Embody STEM Stereotypes Hinder Women's Anticipated Success in STEM? *Social Psychological and Personality Science*, *2*(6), 656–664.

Cheryan, Sapna. (2012). Understanding the Paradox in Math-Related Fields: Why Do Some Gender Gaps Remain While Others Do Not? *Sex Roles*, *66*(3–4), 184–190.

Cheryan, Sapna, Master, A., & Meltzoff, A. N. (2015). Cultural stereotypes as gatekeepers: increasing girls' interest in computer science and engineering by diversifying stereotypes. *Developmental Psychology*, *6*, 49.

Cheryan, Sapna, Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: how stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology*, *97*(6), 1045–1060.

Coaley, K. (2010). *An Introduction to Psychological Assessment and Psychometrics* (Second Edition).

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2 edition). Hillsdale, N.J: Routledge.

Cohen, L., Manion, L., & Morrison, K. (2003). *Research Methods in Education*. Taylor & Francis.

Connellan, J., Baron-Cohen, S., Wheelwright, S., Batki, A., & Ahluwalia, J. (2000). Sex differences in human neonatal social perception. *Infant Behavior and Development*, *23*(1), 113–118.

Conner, T., & Barrett, L. F. (2005). Implicit Self-Attitudes Predict Spontaneous Affect in Daily Life. *Emotion*, *5*(4), 476–488.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating Multiple Processes in Implicit Social Cognition: The Quad Model of Implicit Task Performance. *Journal of Personality and Social Psychology*, *89*(4), 469–487.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2007). The influence of stereotypes on decisions to shoot. *European Journal of Social Psychology*, *37*(6), 1102–1117.

Couper, M. P., Traugott, M. W., & Lamias, M. J. (2001). Web survey design and administration. *Public Opinion Quarterly*, *65*(2), 230–253.

Coutts, E., & Jann, B. (2011). Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT). *Sociological Methods & Research*, *40*(1), 169–193.

Crocker, J., Thompson, L. L., McGraw, K. M., & Ingerman, C. (1987). Downward comparison, prejudice, and evaluations of others: effects of self-esteem and threat. *Journal of Personality and Social Psychology*, *52*(5), 907–916.

Cundiff, J. L., Vescio, T. K., Loken, E., & Lo, L. (2013). Do gender–science stereotypes predict science identification and science career aspirations among undergraduate science majors? *Social Psychology of Education*, *16*(4), 541–554.

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

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve Frequently Asked Questions About Growth Curve Modeling. *Journal of Cognition and Development : Official Journal of the Cognitive Development Society*, *11*(2), 121–136.

Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2012). Balanced Identity Thoery: Review of Evidence for Implicit Consistency in Social Cognition. In B. Gawronski & F. Strack (Eds.), *Cognitive Consistency: A Fundamental Principle in Social Cognition* (pp. 157–177). New York, NY 10012: The Guilford Press.

Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2016). Implicit measures for preschool children confirm self-esteem's role in maintaining a balanced identity. *Journal of Experimental Social Psychology*, *62*, 50–57.

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–Gender Stereotypes in Elementary School Children. *Child Development*, *82*(3), 766–779.

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(5), 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*(5), 800–814.

Dasgupta, Nilanjana. (2004). Implicit Ingroup Favoritism, Outgroup Favoritism, and Their Behavioral Manifestations. *Social Justice Research*, *17*(2), 143–169.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126.

Davis, M. H., Mitchell, K. V., Hall, J. A., Lothert, J., Snapp, T., & Meyer, M. (2001). Empathy, Expectations, and Situational Preferences: Personality Influences on the

Decision to Participate in Volunteer Helping Behaviors. *Journal of Personality*, *67*(3), 469–503.

De Houwer, J. (2001). A Structural and Process Analysis of the Implicit Association Test. *Journal of Experimental Social Psychology*, *37*(6), 443–451.

De Houwer, J. (2003). The extrinsic affective Simon task. *Experimental Psychology*, *50*(2), 77.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, *135*(3), 347–368.

Department for Digital, Culture, Media & Sport. (2018). Data Protection Act 2018. Retrieved from http://www.legislation.gov.uk/ukpga/2018/12/contents/enacted

Devine, R. T., & Hughes, C. (2013). Silent Films and Strange Stories: Theory of Mind, Gender, and Social Experiences in Middle Childhood. *Child Development*, *84*(3), 989–1003.

Diamond, L. M., & Butterworth, M. (2008). Questioning gender and sexual identity: Dynamic links over time. *Sex Roles*, *59*(5–6), 365–376.

Dickinson, E. R., Adelson, J. L., & Owen, J. (2012). Gender Balance, Representativeness, and Statistical Power in Sexuality Research Using Undergraduate Student Samples. *Archives of Sexual Behavior*, *41*(2), 325–327.

Diekman, A. B., Brown, E. R., Johnston, A. M., & Clark, E. K. (2010). Seeking congruity between goals and roles: a new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychological Science*, *21*(8), 1051–1057.

Diekman, A. B., Clark, E. K., Johnston, A. M., Brown, E. R., & Steinberg, M. (2011). Malleability in communal goals and beliefs influences attraction to stem careers: evidence for a goal congruity perspective. *Journal of Personality and Social Psychology*, *101*(5), 902–918.

Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (2010). Prejudice, Stereotyping and Discrimination: Theoretical and Empirical Overview. In *The SAGE Handbook of Prejudice, Stereotyping and Discrimination* (pp. 3–28). London: SAGE Publications Ltd.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*(1), 62–

68.

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the Nature of Prejudice: Automatic and Controlled Processes. *Journal of Experimental Social Psychology*, *33*(5), 510–540.

Dozo, N. (2015). *Gender differences in prejudice: a biological and social psychological analysis.* PhD Thesis, School of Psychology, The University of Queensland, Australia.

Dunham, Y., Baron, A. S., & Banaji, M. R. (2006). From American City to Japanese Village: A Cross-Cultural Investigation of Implicit Race Attitudes. *Child Development*, *77*(5), 1268–1281.

Eagly, A. H. (2013). *Sex Differences in Social Behavior: A Social-role interpretation*. Psychology Press.

Eagly, A. H., & Diekman, A. B. (2005). What is the problem? Prejudice as an attitude-in-context. *On the Nature of Prejudice: Fifty Years after Allport*, 19–35.

Egloff, B., & Schmukle, S. C. (2002). Predictive validity of an implicit association test for assessing anxiety. *Journal of Personality and Social Psychology*, *83*(6), 1441–1455.

Einolf, C. J. (2012). Is Cognitive Empathy More Important than Affective Empathy? A Response to “Who Helps Natural-Disaster Victims?”: Cognitive Empathy and Overcoming Injustice. *Analyses of Social Issues and Public Policy*, *12*(1), 268–271.

Eisenberg, N. (2002). Empathy-related emotional responses, altruism, and their socialization. In R. J. Davidson & A. Harrington (Eds.), *Visions of compassion: Western scientists and Tibetan Buddhists examine human nature* (pp. 131–164). New York, US: Oxford University Press.

Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and other capacities. *Psychological Bulletin*, 100–131.

Eisenberg, N., Morris, A. S., McDaniel, B., & Spinrad, T. L. (2009). Moral Cognitions and Prosocial Responding in Adolescence. *Handbook of Adolescent Psychology*.

Ekehammar, B., Akrami, N., & Araya, T. (2003). Gender differences in implicit prejudice. *Personality and Individual Differences*, *34*(8), 1509–1523.

Erb, T. O., & Smith, W. S. (1984). Validation of the attitude toward women in science scale for early adolescents. *Journal of Research in Science Teaching*, *21*(4), 391–397.

Etzioni, A., & Nunn, C. (1974). The Public Appreciation of Science in Contemporary America. *Daedalus*, *103*(3), 191–205.

Fazio, Russel H., & Towles-Schwen, T. (1999). The MODE model of attitude-behavior processes. In *Dual-process theories in social psychology* (pp. 97–116). New York, NY, US: Guilford Press.

Fazio, Russell H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*(6), 1013–1027.

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

Feist, G. J. (2008). *The Psychology of Science and the Origins of the Scientific Mind*. Yale University Press.

Festinger, L. (1962). *A Theory of Cognitive Dissonance*. Stanford University Press.

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. SAGE.

Fine, C. (2010). Delusions of Gender: How Our Minds, Society, and Neurosexism Create Difference. New York: W.W. Norton & Co.

Finson, K. D. (2002). Drawing a Scientist: What We Do and Do Not Know After Fifty Years of Drawings. *School Science and Mathematics*, *102*(7), 335–345.

Fischbein, S. (1990). Biosocial Influences on Sex Differences for Ability and Achievement Test Results as Well as Marks at School. *Intelligence*, *14*(1), 127–139.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. Journal of Personality and Social Psychology, 82(6), 878–902.

Fiske, S. T., & Dupree, C. (2014). Gaining trust as well as respect in communicating to motivated audiences about science topics. Proceedings of the National Academy of Sciences, 111(Supplement 4), 13593–13597.

Gu, B., & Li, Y. (1995). Sex Ratio at Birth and Son Preference in China. Asia Journal, 2(1), 87–108.

Fleischhauer, M., Strobel, A., Enge, S., & Strobel, A. (2013). Assessing Implicit Cognitive Motivation: Developing and Testing an Implicit Association Test to

Measure Need for Cognition: An indirect measure to assess need for cognition. *European Journal of Personality*, *27*(1), 15–29.

Focquaert, F., Steven, M. S., Wolford, G. L., Colden, A., & Gazzaniga, M. S. (2007). Empathizing and systemizing cognitive traits in the sciences and humanities. *Personality and Individual Differences*, *43*(3), 619–625.

Friese, M., Hofmann, W., & Wänke, M. (2008). When impulses take over: moderated predictive validity of explicit and implicit attitude measures in predicting food choice and consumption behaviour. *The British Journal of Social Psychology*, *47*(3), 397–419.

Frome, P. M., Alfeld, C. J., Eccles, J. S., & Barber, B. L. (2006). Why don't they want a male-dominated job? An investigation of young women who changed their occupational aspirations. *Educational Research and Evaluation*, *12*(4), 359–372.

Frost, B. C., Ko, C.-H. E., & James, L. R. (2007). Implicit and explicit personality: a test of a channeling hypothesis for aggressive behavior. *The Journal of Applied Psychology*, *92*(5), 1299–1319.

Furr, R. M., & Bacharach, V. R. (2008). *Psychometrics: An Introduction*. SAGE Publications.

Galdi, S., Arcuri, L., & Gawronski, B. (2008). Automatic mental associations predict future choices of undecided decision-makers. *Science*, *321*(5892), 1100–1102.

Galdi, S., Cadinu, M., & Tomasetto, C. (2014). The Roots of Stereotype Threat: When Automatic Associations Disrupt Girls' Math Performance. *Child Development*, *85*(1), 250–263.

Gast, A., & Rothermund, K. (2010). When old and frail is not the same: Dissociating category and stimulus effects in four implicit attitude measurement methods. *Quarterly Journal of Experimental Psychology*, *63*(3), 479–498.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731.

Gawronski, B., & Bodenhausen, G. V. (2014). Implicit and Explicit Evaluation: A Brief Review of the Associative-Propositional Evaluation Model: APE Model. *Social and Personality Psychology Compass*, *8*(8), 448–462.

Gawronski, B., & De Houwer, J. (2013). Implicit Measures in Social and Personality Psychology. In H. T. Reis & C. M. Judd (Eds.), *Handbook of Research Methods in*

Social and Personality Psychology (2nd ed., pp. 283–310). New York: Cambridge University Press.

Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What Do Implicit Measures Tell Us? Scrutinizing the Validity of Three Common Assumptions. *Perspectives on Psychological Science*, *2*(2), 181–193.

Gawronski, B., Peters, K. R., Brochu, P. M., & Strack, F. (2008). Understanding the relations between different forms of racial prejudice: a cognitive consistency perspective. *Personality & Social Psychology Bulletin*, *34*(5), 648–665.

Gawronski, B., & Strack, F. (2004). On the propositional nature of cognitive consistency: Dissonance changes explicit, but not implicit attitudes. *Journal of Experimental Social Psychology*, *40*(4), 535–542.

Gawronski, B., Walther, E., & Blank, H. (2005). Cognitive consistency and the formation of interpersonal attitudes: cognitive balance affects the encoding of social information. *Journal of Experimental Social Psychology*, *41*(6), 618–626.

Geer, J. H., Estupinan, L. A., & Manguno-Mire, G. M. (2000). Empathy, social skills, and other relevant cognitive processes in rapists and child molesters. *Aggression and Violent Behavior*, *5*(1), 99–126.

Glass, J. L., Sassler, S., Levitte, Y., & Michelmore, K. M. (2013). What's So Special about STEM? A Comparison of Women's Retention in STEM and Professional Occupations. *Social Forces; a Scientific Medium of Social Study and Interpretation*, *92*(2), 723–756.

Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2014). Measures of Implicit Attitudes May Conceal Differences in Implicit Associations: The Case of Antiaging Bias. *Social Psychological and Personality Science*, *5*(3), 271–278.

Good, J. J., Woodzicka, J. A., & Wingfield, L. C. (2010). The effects of gender stereotypic and counter-stereotypic textbook images on science performance. *The Journal of Social Psychology*, *150*(2), 132–147.

Govan, C. L., & Williams, K. D. (2004). Changing the affective valence of the stimulus items influences the IAT by re-defining the category labels. *Journal of Experimental Social Psychology*, *40*(3), 357–365.

Gray, M., & James, A. (2007). Connecting Gender and Economic Competitiveness: Lessons from Cambridge's High-Tech Regional Economy. *Environment and Planning A*, *39*(2), 417–436.

Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I., & Banaji, M. R. (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of General Internal Medicine*, *22*(9), 1231–1238.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes. *Psychological Review*, *102*(1), 4–27.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *109*(1), 3–25.

Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, *79*(6), 1022.

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. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift Für Experimentelle Psychologie*, *48*(2), 85–93.

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., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*(1), 17–41.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin*, *85*(4), 845–857.

Harton, H. C., & Lyons, P. C. (2003). Gender, empathy, and the choice of the psychology major. *Teaching of Psychology*, *30*(1), 19–24.

Hayes, A. F., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods*, *39*(4), 709–722.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. Psychology Press.

Higher Education Statistics Agency. (2018, January). Higher Education Student

Statistics: UK, 2016/17 - Student numbers and characteristics. HESA.

Hilgard, J., Bartholow, B. D., Dickter, C. L., & Blanton, H. (2014). Characterizing switching and congruency effects in the Implicit Association Test as reactive and proactive cognitive control. *Social Cognitive and Affective Neuroscience*, nsu060.

Hilton, J. L., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, *47*(1), 237–271.

Hirsch, E. M. (2007). The Role of Empathy in Medicine: A Medical Student's Perspective. *Virtual Mentor*, *9*(6), 423–427.

Hofmann, W., & Friese, M. (2008). Impulses got the better of me: alcohol moderates the influence of implicit attitudes toward food cues on eating behavior. *Journal of Abnormal Psychology*, *117*(2), 420–427.

Hofmann, W., Friese, M., & Strack, F. (2009). Impulse and Self-Control from a Dual-Systems Perspective. *Perspectives on Psychological Science*, *4*(2), 162–176.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A Meta-Analysis on the Correlation Between the Implicit Association Test and Explicit Self-Report Measures. *Personality and Social Psychology Bulletin*, *31*(10), 1369–1385.

Hogg, M. A., & Hains, S. C. (1996). Intergroup relations and group solidarity: Effects of group identification and social beliefs on depersonalized attraction. *Journal of Personality and Social Psychology*, *70*(2), 295–309.

Hojat, M., DeSantis, J., & Gonnella, J. S. (2017). Patient Perceptions of Clinician's Empathy. *Journal of Patient Experience*, *4*(2), 78–83.

Hönekopp, J. (2012). Digit Ratio 2D:4D in Relation to Autism Spectrum Disorders, Empathizing, and Systemizing: A Quantitative Review: Digit ratio 2D:4D and autism. *Autism Research*, *5*(4), 221–230.

Houben, K., & Wiers, R. W. (2008). Implicitly positive about alcohol? Implicit positive associations predict drinking behavior. *Addictive Behaviors*, *33*(8), 979–986.

Houwer, J., & Bruycker, E. (2007). The implicit association test outperforms the extrinsic affective Simon task as an implicit measure of inter-individual differences in attitudes. *British Journal of Social Psychology*, *46*(2), 401–421.

Houwer, J. D., Custers, R., & Clercq, A. D. (2006). Do smokers have a negative implicit attitude toward smoking? *Cognition and Emotion*, *20*(8), 1274–1284.

Howard, J. W., & Rothbart, M. (1980). Social categorization and memory for in-group and out-group behavior. *Journal of Personality and Social Psychology*, *38*(2), 301–310.

Hughes, M., & Tuch, S. A. (2003). Gender Differences in Whites' Racial Attitudes: Are Women's Attitudes Really More Favorable? *Social Psychology Quarterly*, *66*(4), 384–401.

Hunt, J. (2010). *Why Do Women Leave Science and Engineering?* (Working Paper No. 15853). National Bureau of Economic Research.

Ickes, W., Gesn, P. R., & Graham, T. (2000). Gender differences in empathic accuracy: Differential ability or differential motivation? *Personal Relationships*, *7*(1), 95–109.

Jajodia, A., & Earleywine, M. (2003). Measuring alcohol expectancies with the implicit association test. *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors*, *17*(2), 126–133.

Jellison, W. A., McConnell, A. R., & Gabriel, S. (2004). Implicit and Explicit Measures of Sexual Orientation Attitudes: In Group Preferences and Related Behaviors and Beliefs among Gay and Straight Men. *Personality and Social Psychology Bulletin*, *30*(5), 629–642.

Jong, P. de, Hout, M. van den, Rietbroek, H., & Huijding, J. (2003). Dissociations between implicit and explicit attitudes toward phobic stimuli. *Cognition and Emotion*, *17*(4), 521–545.

Jordan, C. H., Spencer, S. J., Zanna, M. P., Hoshino-Browne, E., & Correll, J. (2003). Secure and defensive high self-esteem. *Journal of Personality and Social Psychology*, *85*(5), 969–978.

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology*, *33*(1), 1–27.

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A Decade of System Justification Theory: Accumulated Evidence of Conscious and Unconscious Bolstering of the Status Quo. *Political Psychology*, *25*, 881–919.

Kaiser, S., Walther, S., Nennig, E., Kronmüller, K., Mundt, C., Weisbrod, M., … Vogeley, K. (2008). Gender-specific strategy use and neural correlates in a spatial perspective taking task. *Neuropsychologia*, *46*(10), 2524–2531.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test.

Journal of Personality and Social Psychology, *81*(5), 774–788.

Karpinski, A., & Steinman, R. B. (2006). The Single Category Implicit Association Test as a measure of implicit social cognition. *Journal of Personality and Social Psychology*, *91*(1), 16–32.

Kay, D., & Stewart, A. J. (2001). Framing gendered identities. In R. K. Unger (Ed.), *Handbook of the psychology of women and gender* (pp. 84–97). New York: Wiley.

Kessels, U. (2005). Fitting into the stereotype: How gender-stereotyped perceptions of prototypic peers relate to liking for school subjects. *European Journal of Psychology of Education*, *20*(3), 309–323.

Khorashad, B. S., Baron-Cohen, S., Roshan, G. M., Kazemian, M., Khazai, L., Aghili, Z., … Afkhamizadeh, M. (2015). The "Reading the Mind in the Eyes" Test: Investigation of Psychometric Properties and Test–Retest Reliability of the Persian Version. *Journal of Autism and Developmental Disorders*, *45*(9), 2651–2666.

Klauer, K. C. (2005). Task-Set Inertia, Attitude Accessibility, and Compatibility-Order Effects: New Evidence for a Task-Set Switching Account of the Implicit Association Test Effect. *Personality and Social Psychology Bulletin*, *31*(2), 208–217.

Klauer, Karl Christoph, Schmitz, F., Teige-Mocigemba, S., & Voss, A. (2010). Understanding the role of executive control in the Implicit Association Test: Why flexible people have small IAT effects. *The Quarterly Journal of Experimental Psychology*, *63*(3), 595–619.

Klonis, S. C., Plant, E. A., & Devine, P. G. (2005). Internal and External Motivation to Respond Without Sexism. Personality and Social Psychology Bulletin, 31(9), 1237–1249.

Ko, S. J., Judd, C. M., & Blair, I. V. (2006). What the Voice Reveals: Within- and Between-Category Stereotyping on the Basis of Voice. Personality and Social Psychology Bulletin, 32(6), 806–819.

Ko, S. J., Muller, D., Judd, C. M., & Stapel, D. A. (2008). Sneaking in through the back door: How category-based stereotype suppression leads to rebound in feature-based effects. Journal of Experimental Social Psychology, 44(3), 833–839.

Koballa, T. R., Krajkovich, J. G., & Smith, J. K. (1983). Comments on "the development of the image of science and scientists scale". *Journal of Research in Science Teaching*, *20*(6), 595–597.

Koenig, A. M., & Eagly, A. H. (2005). Stereotype Threat in Men on a Test of Social

Sensitivity. *Sex Roles*, *52*(7–8), 489–496.

Krajkovich, J. G., & Smith, J. K. (1982). The development of the image of science and scientists scale. *Journal of Research in Science Teaching*, *19*(1), 39–44.

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological Research Online: Report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. *American Psychologist*, *59*(2), 105–117.

Krueger, R. A., & Casey, M. A. (2009). *Focus Groups: A Practical Guide for Applied Research*. SAGE Publications.

Kurtz-Costes, B., Rowley, S. J., Harris-Britt, A., & Woods, T. A. (2008). Gender Stereotypes about Mathematics and Science and Self-Perceptions of Ability in Late Childhood and Early Adolescence. *Merrill-Palmer Quarterly*, *54*(3), 386–409.

LaFrance, M., Hecht, M. A., & Paluck, E. L. (2003). The contingent smile: a meta-analysis of sex differences in smiling. *Psychological Bulletin*, *129*(2), 305–334.

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the implicit association test: IV. *Implicit Measures of Attitudes*, 59–102.

Lane, K. A., Goh, J. X., & Driver-Linn, E. (2012). Implicit science stereotypes mediate the relationship between gender and academic participation. *Sex Roles*, *66*(3–4), 220–234.

Lawson, J., Baron-Cohen, S., & Wheelwright, S. (2004). Empathising and systemising in adults with and without Asperger Syndrome. *Journal of Autism and Developmental Disorders*, *34*(3), 301–310.

Leaper, C., Anderson, K. J., & Sanders, P. (1998). Moderators of gender effects on parents' talk to their children: a meta-analysis. *Developmental Psychology*, *34*(1), 3–27.

Leon, A. J. (2014). *The Good, The Geek, and the Ugly: A Critical Examination of the Portrayal of Geek Women in Popular Media*. Lehigh University.

Lippa, R. A. (2005). Subdomains of Gender-Related Occupational Interests: Do They Form a Cohesive Bipolar M-F Dimension? *Journal of Personality*, *73*(3), 693–730.

Lippmann, W. (1922). Public Opinion. Harcourt, Brace.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd edition). New York: Wiley.

Lorre, C., Prady, B., Collier, M., Belyeu, F. O., Cohen, R., Goetsch, D., Molaro, S., ... Warner Home Video (Firm). (2008). The big bang theory: The complete first season. Burbank, CA: Distributed by Warner Home Video.

Losh, Susan C., Wilke, R., & Pop, M. (2008). Some Methodological Issues with 'Draw a Scientist Tests' among Young Children. *International Journal of Science Education*, *30*(6), 773–792.

Losh, Susan Carol. (2010). Stereotypes about scientists over time among US adults: 1983 and 2001. *Public Understanding of Science*, *19*(3), 372–382.

Lubinski, D., & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science*, *1*(4), 316–345.

Maccoby, E. E., & Jacklin, C. N. (1974). *The Psychology of Sex Differences*. Stanford University Press.

Manning, J. T., Reimers, S., Baron-Cohen, S., Wheelwright, S., & Fink, B. (2010). Sexually dimorphic traits (digit ratio, body height, systemizing–empathizing scores) and gender segregation between occupations: Evidence from the BBC internet study. *Personality and Individual Differences*, *49*(5), 511–515.

Manson, C., & Winterbottom, M. (2012). Examining the association between empathising, systemising, degree subject and gender. *Educational Studies*, *38*(1), 73–88.

Mason, M. A., & Goulden, M. (2004). Marriage and Baby Blues: Redefining Gender Equity in the Academy. *The Annals of the American Academy of Political and Social Science*, *596*, 86–103.

McConahay, J. B. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In John F. Dovidio & Samuel L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91–125). San Diego, CA, US: Academic Press.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes. *Journal of Experimental Social Psychology*, *37*(5), 435–442.

Mccullough, M. E., Emmons, R. A., & Tsang, J.-A. (2002). The grateful disposition: a conceptual and empirical topography. *Journal of Personality and Social Psychology*, *82*(1), 112–127.

McCullough, M. E., Worthington, E. L., & Rachal, K. C. (1997). Interpersonal forgiving in close relationships. *Journal of Personality and Social Psychology*, *73*(2), 321–336.

McFarland, S. G., & Crouch, Z. (2002). A Cognitive Skill Confound on the Implicit Association Test. *Social Cognition*, *20*(6), 483–510.

Mead, M., & Métraux, R. (1957). Image of the Scientist among High-School Students A Pilot Study. *Science*, *126*(3270), 384–390.

Meltzoff, A. (2013). Origins of Social Cognition. *Navigating the Social World: What Infants, Children, and Other Species Can Teach Us*, 139.

Mestre, M. V., Samper, P., Frías, M. D., & Tur, A. M. (2009). Are women more empathetic than men? A longitudinal study in adolescence. *The Spanish Journal of Psychology*, *12*(1), 76–83.

Michalska, K. J., Kinzler, K. D., & Decety, J. (2013). Age-related sex differences in explicit measures of empathy do not predict brain responses across childhood and adolescence. *Developmental Cognitive Neuroscience*, *3*, 22–32.

Mierke, J., & Klauer, K. C. (2003). Method-Specific Variance in the Implicit Association Test. *Journal of Personality and Social Psychology*, *85*(6), 1180–1192.

Miller, A. (1981). A Survey of Introductory Psychology Subject Pool Practices among Leading Universities. *Teaching of Psychology*, *8*(4), 211–213.

Miller, D. I., Nolla, K. M., Eagly, A. H., & Uttal, D. H. (2018). The Development of Children's Gender-Science Stereotypes: A Meta-analysis of 5 Decades of U.S. Draw-A-Scientist Studies. *Child Development*, *00*(0), 1–13.

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, *132*(3), 455–469.

Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: on the development and operation of cues for control. *Journal of Personality and Social Psychology*, *83*(5), 1029–1050.

Moore, D. J. (2004). Affect intensity, gender and the expression of emotion in response to advertising appeals. *Advances in Consumer Research*, *31*, 29–30.

Moors, A., & De Houwer, J. (2006). Automaticity: A Theoretical and Conceptual

Analysis. *Psychological Bulletin*, *132*(2), 297–326.

Morgan, D. L. (1997). *The Focus Group Guidebook*. SAGE Publications.

Morgan, R., Kirby, C., & Stamenkovic, A. (2016). *The UK STEM Education Landscape: A report for the Lloyd's Register Foundation from the Royal Academy of Engineering Education and Skills Committee*.

MORI, I. (2014). Public Attitudes to Science 2014: Main Report. *London, Ipsos MORI Social Research Institute*, 192.

Mujtaba, T., & Reiss, M. (2014). A survey of psychological, motivational, family and perceptions of physics education factors that explain 15 year-old students' aspirations to study post-compulsory physics in English schools. *International Journal of Science and Mathematics Education*, *12*, 371–393.

Mummendey, A., Otten, S., Berger, U., & Kessler, T. (2000). Positive-Negative Asymmetry in Social Discrimination: Valence of Evaluation and Salience of Categorization. *Personality and Social Psychology Bulletin*, *26*(10), 1258–1270.

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: how situational cues affect women in math, science, and engineering settings. *Psychological Science*, *18*(10), 879–885.

Nelson, S. C. (2011). Feeling Thermometer. In P. J. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods* (pp. 275–277). Thousand Oaks, CA, US: Sage Publications, Inc.

Nettle, D. (2007). Empathizing and systemizing: What are they, and what do they contribute to our understanding of psychological sex differences? *British Journal of Psychology*, *98*(2), 237–255.

Newton, L. D., & Newton, D. P. (1998). Primary children's conceptions of science and the scientist: is the impact of a National Curriculum breaking down the stereotype? *International Journal of Science Education*, *20*(9), 1137–1149.

Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002a). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, *6*(1), 101–115.

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition*, 625–664.

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

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and Using the Implicit Association Test: II. Method Variables and Construct Validity. *Personality and Social Psychology Bulletin*, *31*(2), 166–180.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. *Automatic Processes in Social Thinking and Behavior*, 265–292.

Nosek, B. A., & Smyth, F. L. (2007). A Multitrait-Multimethod Validation of the Implicit Association Test. *Experimental Psychology*, *54*(1), 14–29.

Nosek, B. A., & Smyth, F. L. (2011). Implicit Social Cognitions Predict Sex Differences in Math Engagement and Achievement. *American Educational Research Journal*, *48*(5), 1125–1156.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., … Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*(1), 36–88.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., … Greenwald, A. G. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10593–10597.

Olson, K. (2011). Double-Barreled Question. In P. J. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods* (pp. 209–211). Thousand Oaks, CA, US: Sage Publications, Inc.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science*, *12*(5), 413–417.

Olson, Michael A., & Fazio, R. H. (2004). Reducing the Influence of Extrapersonal Associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology*, *86*(5), 653–667.

Olson, Michael A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality & Social Psychology Bulletin*, *32*(4), 421–433.

Operario, D., & Fiske, S. T. (2004). Stereotypes: Content, Structures, Processes, and Context. In Social cognition (pp. 120–141). Malden: Blackwell Publishing.

Osgood, C. E. (1952). The nature and measurement of meaning. Psychological Bulletin, 49(3), 197–237.

Ottaway, S. A., Hayden, D. C., & Oakes, M. A. (2001). Implicit Attitudes and Racism: Effects of Word Familiarity and Frequency on the Implicit Association Test. *Social Cognition*, *19*(2), 97–144.

Paechter, C. (2010). Tomboys and girly-girls: embodied femininities in primary schools. *Discourse: Studies in the Cultural Politics of Education*, *31*(2), 221–235.

Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology*, *94*(1), 16–31.

Penke, L., Eichstaedt, J., & Asendorpf, J. B. (2006). Single-Attribute Implicit Association Tests (SA-IAT) for the Assessment of Unipolar Constructs. *Experimental Psychology*, *53*(4), 283–291.

Perugini, M. (2005). Predictive models of implicit and explicit attitudes. *British Journal of Social Psychology*, *44*(1), 29–45.

Peters, K. R., & Gawronski, B. (2011). Are We Puppets on a String? Comparing the Impact of Contingency and Validity on Implicit and Explicit Evaluations. *Personality and Social Psychology Bulletin*, *37*(4), 557–569.

Petty, R. E., Tormala, Z. L., Briñol, P., & Jarvis, W. B. G. (2006). Implicit ambivalence from attitude change: An exploration of the PAST model. *Journal of Personality and Social Psychology*, *90*(1), 21–41.

Ratcliff, J. J., Lassiter, G. D., Markman, K. D., & Snyder, C. J. (2006). Gender differences in attitudes toward gay men and lesbians: the role of motivation to respond without prejudice. *Personality & Social Psychology Bulletin*, *32*(10), 1325–1338.

Resmini, M. (2016). The ‘Leaky Pipeline'. *Chemistry – A European Journal*, *22*(11), 3533–3534.

Richeson, J. A., & Nussbaum, R. J. (2004). The impact of multiculturalism versus color-blindness on racial bias. *Journal of Experimental Social Psychology*, *40*(3), 417–423.

Robson, C. (2002). *Real World Research: A Resource for Social Scientists and Practitioner-Researchers*. Wiley.

Rommes, E., Overbeek, G., Scholte, R., Engels, R., & Kemp, R. D. (2007). 'I'm not interested in computers': Gender-based occupational choices of adolescents. *Information, Communication & Society*, *10*(3), 299–319.

Rubin, M., & Hewstone, M. (1998). Social Identity Theory's Self-Esteem Hypothesis: A Review and Some Suggestions for Clarification. *Personality and Social Psychology Review*, *2*(1), 40–62.

Rudman, Laurie. A., & Kilianski, S. E. (2000). Implicit and Explicit Attitudes Toward Female Authority. *Personality and Social Psychology Bulletin*, *26*(11), 1315–1328.

Rudman, Laurie. A., Feinberg, J., & Fairchild, K. (2002). Minority Members' Implicit Attitudes: Automatic Ingroup Bias as A Function of Group Status. *Social Cognition*, *20*(4), 294–320.

Rudman, Laurie A., & Glick, P. (2001). Prescriptive Gender Stereotypes and Backlash Toward Agentic Women. *Journal of Social Issues*, *57*(4), 743–762.

Rudman, Laurie A., & Phelan, J. E. (2010). The Effect of Priming Gender Roles on Women's Implicit Gender Beliefs and Career Aspirations. *Social Psychology*, *41*(3), 192–202.

Rueckert, L. (2011). Are Gender Differences in Empathy Due to Differences in Emotional Reactivity? *Psychology*, *02*(06), 574–578.

Russell, T. A., Tchanturia, K., Rahman, Q., & Schmidt, U. (2007). Sex differences in theory of mind: A male advantage on Happé's "cartoon" task. *Cognition & Emotion*, *21*(7), 1554–1564.

Rutland, A., Cameron, L., Milne, A., & McGeorge, P. (2005). Social norms and self-presentation: children's implicit and explicit intergroup attitudes. *Child Development*, *76*(2), 451–466.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: a systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91*(6), 995–1008.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*(2), 336–356.

Schmukle, S. C. (2005). Unreliability of the dot probe task. *European Journal of Personality*, *19*(7), 595–605.

Schmukle, S. C., & Egloff, B. (2005). A Latent State-Trait Analysis of Implicit and

Explicit Personality Measures. *European Journal of Psychological Assessment*, *21*(2), 100–107.

Schnabel, K., Asendorpf, J. B., & Greenwald, A. G. (2008). Understanding and using the implicit association test: V. measuring semantic aspects of trait self-concepts. *European Journal of Personality*, *22*(8), 695–706.

Schneider, J. S. (2010). Impact of Undergraduates' Stereotypes of Scientists on their Intentions to Pursue a Career in Science. Retrieved from

Settles, I. H., Jellison, W. A., & Pratt-Hyatt, J. S. (2009). Identification with multiple social groups: The moderating role of identity change over time among women-scientists. *Journal of Research in Personality*, *43*(5), 856–867.

Sherman, S. J., Rose, J. S., Koch, K., Presson, C. C., & Chassin, L. (2003). Implicit and Explicit Attitudes Toward Cigarette Smoking: the Effects of Context and Motivation. *Journal of Social and Clinical Psychology*, *22*(1), 13–39.

Sidanius, J., Pratto, F., van Laar, C., & Levin, S. (1999). Social Dominance Theory: Its Agenda and Method. *Political Psychology*, *25*(6,). Retrieved from

Sinclair, S., Dunn, E., & Lowery, B. (2005). The relationship between parental racial attitudes and children's implicit prejudice. *Journal of Experimental Social Psychology*, *41*, 283–289.

Sjøberg, S., & Schreiner, C. (2010). The ROSE project: An overview and key findings. *Oslo: University of Oslo*, 1–31.

Sjøberg, S., & Schreiner, C. (2012). Results and Perspectives from the Rose Project. In D. Jorde & D. Justin, *Science Education Research and Practice in Europe.* (Vol. 5). Rotterdam: SensePublishers.

Smeding, A. (2012). Women in Science, Technology, Engineering, and Mathematics (STEM): An Investigation of Their Implicit Gender Stereotypes and Stereotypes' Connectedness to Math Performance. *Sex Roles*, *67*(11–12), 617–629.

Smith, B. K., & Jensen, E. A. (2016). Critical review of the United Kingdom's "gold standard" survey of public attitudes to science. *Public Understanding of Science*, *25*(2), 154–170.

Smith, E. R., & DeCoster, J. (2000). Dual-Process Models in Social and Cognitive Psychology: Conceptual Integration and Links to Underlying Memory Systems. *Personality and Social Psychology Review*, *4*(2), 108–131.

Smyth, F. L., & Nosek, B. A. (2015). On the gender–science stereotypes held by scientists: explicit accord with gender-ratios, implicit accord with scientific identity. *Frontiers in Psychology*, *6*, 415.

Spelke, E. S. (2005). Sex Differences in Intrinsic Aptitude for Mathematics and Science? A Critical Review. *American Psychologist*, *60*(9), 950–958.

Steffens, M. C. (2004). Is the implicit association test immune to faking? *Experimental Psychology*, *51*(3), 165–179.

Steffens, M., & Plewe, I. (2001). Items' Cross-Category Associations as a Confounding Factor in the Implicit Association Test. *Zeitschrift Für Experimentelle Psychologie : Organ Der Deutschen Gesellschaft Für Psychologie*, *48*, 123–134.

Steinke, J., Lapinski, M. K., Crocker, N., Zietsman-Thomas, A., Williams, Y., Evergreen, S. H., & Kuchibhotla, S. (2007). Assessing Media Influences on Middle School–Aged Children's Perceptions of Women in Science Using the Draw-A-Scientist Test (DAST). *Science Communication*, *29*(1), 35–64.

Stieger, S., & Reips, U.-D. (2010). What are participants doing while filling in an online questionnaire: A paradata collection tool and an empirical study. *Computers in Human Behavior*, *26*(6), 1488–1495.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*(3), 220–247.

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin*, *135*(6), 859–884.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178.

Tajfel, H., & Turner, J. (2001). An integrative theory of intergroup conflict. In M. A. Hogg & D. Abrams (Eds.), *Intergroup relations: Essential readings* (pp. 94–109). New York, NY, US: Psychology Press.

Teachman, B. A., Gapinski, K. D., Brownell, K. D., Rawlins, M., & Jeyaram, S. (2003). Demonstrations of implicit anti-fat bias: The impact of providing causal information and evoking empathy. *Health Psychology*, *22*(1), 68–78.

Teatero, M. L., & Netley, C. (2013). A Critical Review of the Research on the Extreme Male Brain Theory and Digit Ratio (2D:4D). *Journal of Autism and Developmental Disorders*, *43*(11), 2664–2676.

Teige-Mocigemba, S., Klauer, K. C., & Sherman, J. W. (2010). A practical guide to Implicit Association Task and related tasks. In B. Gawronski & B. K. Payne, *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 117–139). New York, US: Guilford Press.

The Ethics Committee of the British Psychological Society. (2009). Code of ethics and conduct.

Thomas, M. D., Henley, T. B., Snell, C. M., Thomas, M. D., Henley, T. B., & Snell, C. M. (2006). The Draw a Scientist Test: A Different Population and a Somewhat Different Story. *College Student Journal*, *40*(1), 140–148.

Thomson, N. D., Wurtzburg, S. J., & Centifanti, L. C. M. (2015). Empathy or science? Empathy explains physical science enrollment for men and women. *Learning and Individual Differences*, *40*, 115–120.

Tintori, A., & Palomba, R. (2017). *Turn on the light on science*. London: Ubiquity Press.

Titchener, E. B. (1909). The Psychology of Feeling and Attention. *Journal of Philosophy, Psychology and Scientific Methods*, *6*(3), 64–77.

Todd, A. R., Bodenhausen, G. V., Richeson, J. A., & Galinsky, A. D. (2011). Perspective taking combats automatic expressions of racial bias. *Journal of Personality and Social Psychology*, *100*(6), 1027–1042.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Cambridge, MA, US: Basil Blackwell.

UCAS. (2017). End of cycle report 2017: Patterns by applicant characteristics.

Valla, J. M., & Ceci, S. J. (2011). Can sex differences in science be tied to the long reach of prenatal hormones? Brain organization theory, digit ratio (2D/4D), and sex differences in preferences and cognition. *Perspectives on Psychological Science*, *6*(2), 134–146.

Valla, J. M., Ganzel, B. L., Yoder, K. J., Chen, G. M., Lyman, L. T., Sidari, A. P., … Belmonte, M. K. (2010). More than maths and mindreading: Sex differences in empathizing/systemizing covariance. *Autism Research*, *3*(4), 174–184.

Van der Graaff, J., Branje, S., De Wied, M., Hawk, S., Van Lier, P., & Meeus, W. (2014). Perspective taking and empathic concern in adolescence: Gender differences in developmental changes. *Developmental Psychology*, *50*(3), 881–888.

Vescio, T. K., Sechrist, G. B., & Paolucci, M. P. (2003). Perspective taking and prejudice reduction: the mediational role of empathy arousal and situational attributions. *European Journal of Social Psychology*, *33*(4), 455–472.

Von Horn, A., BäCkman, L., Davidsson, T., & Hansen, S. (2010). Empathizing, systemizing and finger length ratio in a Swedish sample. *Scandinavian Journal of Psychology*, *51*(1), 31–37.

Voracek, M., & Dressler, S. G. (2006). Lack of correlation between digit ratio (2D:4D) and Baron-Cohen's "Reading the Mind in the Eyes" test, empathy, systemising, and autism-spectrum quotients in a general population sample. *Personality and Individual Differences*, *41*(8), 1481–1491.

Wager, T. D., Phan, K. L., Liberzon, I., & Taylor, S. F. (2003). Valence, gender, and lateralization of functional brain anatomy in emotion: a meta-analysis of findings from neuroimaging. *NeuroImage*, *19*(3), 513–531.

Wagner, U., Lampen, L., & Syllwasschy, J. (1986). In-group inferiority, social identity and out-group devaluation in a modified minimal group study. *British Journal of Social Psychology*, *25*(1), 15–23.

Wakabayashi, A., Baron-Cohen, S., Uchiyama, T., Yoshida, Y., Kuroda, M., & Wheelwright, S. (2007). Empathizing and Systemizing in Adults with and without Autism Spectrum Conditions: Cross-Cultural Stability. *Journal of Autism and Developmental Disorders*, *37*(10), 1823–1832.

Wakabayashi, A., & Nakazawa, Y. (2010). On relationships between digit ratio (2D:4D) and two fundamental cognitive drives, empathizing and systemizing, in Japanese sample. *Personality and Individual Differences*, *49*(8), 928–931.

Wakabayashi, A., Sasaki, J., & Ogawa, Y. (2012). Sex Differences in Two Fundamental Cognitive Domains: Empathizing and Systemizing in Children and Adults. *Journal of Individual Differences*, *33*(1), 24–34.

Ward, A. (1977). Magician in a White Coat. *Science Activities: Classroom Projects and Curriculum Ideas*, *14*(1), 6–9.

Weisgram, E. S., Dinella, L. M., & Fulcher, M. (2011). The Role of Masculinity/Femininity, Values, and Occupational Value Affordances in Shaping Young Men's and Women's Occupational Choices. *Sex Roles*, *65*(3–4), 243–258.

Wheelwright, S., Baron-Cohen, S., Goldenfeld, N., Delaney, J., Fine, D., Smith, R., … Wakabayashi, A. (2006). Predicting Autism Spectrum Quotient (AQ) from the Systemizing Quotient-Revised (SQ-R) and Empathy Quotient (EQ). *Brain Research*, *1079*(1), 47–56.

Wiers, R. W., Van Woerden, N., Smulders, F. T. Y., & De Jong, P. J. (2002). Implicit and explicit alcohol-related cognitions in heavy and light drinkers. *Journal of Abnormal Psychology*, *111*(4), 648–658.

Wilhelm, M. O., & Bekkers, R. (2010). Helping Behavior, Dispositional Empathic Concern, and the Principle of Care. *Social Psychology Quarterly*, *73*(1), 11–32.

Willer, R., Wimer, C., & Owens, L. A. (2015). What drives the gender gap in charitable giving? Lower empathy leads men to give less to poverty relief. *Social Science Research*, *52*, 83–98.

Williams, J. E., & Best, D. L. (1990). *Measuring sex stereotypes: A multination study (rev. ed.)*. Thousand Oaks, CA, US: Sage Publications, Inc.

Williams, W. M., & Ceci, S. J. (2012). When Scientists Choose Motherhood. *American Scientist*, *100*(2), 138–145.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*(1), 101–126.

Winter, L. (2008). *Attitudes towards technology based on stereotypes, self-categorisation and sense of control*. University of Leicester, Leicester, United Kindom.

WISE. (2015). Women in Science, Technology, Engineering and Mathematics: The Talent Pipeline from Classroom to Boardroom - UK Statistics 2014.

WISE. (2017). Core STEM Graduates 2017. Retrieved 26 June 2018, from https://www.wisecampaign.org.uk/statistics/core-stem-graduates-2017/

Witt, M. G., & Wood, W. (2010). Self-regulation of Gendered Behavior in Everyday Life. *Sex Roles*, *62*(9–10), 635–646.

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, *81*(5), 815–827.

Wood, W., & Eagly, A. H. (2012). Biosocial Construction of Sex Differences and Similarities in Behavior. In *Advances in Experimental Social Psychology* (Vol. 46, pp. 55–123). Elsevier.

Worell, J. (2001). *Encyclopedia of Women and Gender: Sex Similarities and*

Differences and the Impact of Society on Gender. Academic Press.

Wright, B., McKendree, J., Morgan, L., Allgar, V. L., & Brown, A. (2014). Examiner and simulated patient ratings of empathy in medical student final year clinical examination: are they useful? *BMC Medical Education*, *14*, 199.

Wyer. (2003). Intending to stay: Images of scientists, attitudes toward women, and gender as influences on persistence among science and engineering majors. *Journal of Women and Minorities in Science and Engineering.*, *9*(1), 1–16.

Wyer, M., Schneider, J., Nassar-McMillan, S., & Oliver-Hoyo, M. (2010). Capturing stereotypes: Developing a scale to explore US college students' images of science and scientists. *International Journal of Gender, Science and Technology*, *2*(3), 381–415.

Zerbe, W. J., & Paulhus, D. L. (1987). Socially Desirable Responding in Organizational Behavior: A Reconception. *The Academy of Management Review*, *12*(2), 250–264.

Zeyer, A., Çetin-Dindar, A., Md Zain, A. N., Juriševič, M., Devetak, I., & Odermatt, F. (2013). Systemizing: A cross-cultural constant for motivation to learn science. *Journal of Research in Science Teaching*, *50*(9), 1047–1067.

Zitelny, H., Shalom, M., & Bar-Anan, Y. (2017). What Is the Implicit Gender-Science Stereotype? Exploring Correlations Between the Gender-Science IAT and Self-Report Measures. *Social Psychological and Personality Science*, *8*(7), 719

Appendix I Full attribute item categorisation scale

Please rate how much you associate the following traits with *rationality* or *empathy*

	Strongly Rationality	Moderately Rationality	Slightly Rationality	Neither Rationality Nor Empathy	Slightly Empathy	Moderately Empathy	Strongly Empathy
Affection	☐	☐	☐	☐	☐	☐	☐
Analytical	☐	☐	☐	☐	☐	☐	☐
Appreciation	☐	☐	☐	☐	☐	☐	☐
Consistent	☐	☐	☐	☐	☐	☐	☐
Coolness	☐	☐	☐	☐	☐	☐	☐
Comprehensive	☐	☐	☐	☐	☐	☐	☐
Concern	☐	☐	☐	☐	☐	☐	☐
Coherence	☐	☐	☐	☐	☐	☐	☐
Considerate	☐	☐	☐	☐	☐	☐	☐
Deduction	☐	☐	☐	☐	☐	☐	☐
Ethical	☐	☐	☐	☐	☐	☐	☐
Empathetic	☐	☐	☐	☐	☐	☐	☐
Emotion	☐	☐	☐	☐	☐	☐	☐
Feeling	☐	☐	☐	☐	☐	☐	☐
Insight	☐	☐	☐	☐	☐	☐	☐
Induction	☐	☐	☐	☐	☐	☐	☐
Intuitive	☐	☐	☐	☐	☐	☐	☐
Logical	☐	☐	☐	☐	☐	☐	☐
Ordered	☐	☐	☐	☐	☐	☐	☐
Organised	☐	☐	☐	☐	☐	☐	☐
Rational	☐	☐	☐	☐	☐	☐	☐
Reasoning	☐	☐	☐	☐	☐	☐	☐
Soul	☐	☐	☐	☐	☐	☐	☐
Systematic	☐	☐	☐	☐	☐	☐	☐
Sensitive	☐	☐	☐	☐	☐	☐	☐
Sanity	☐	☐	☐	☐	☐	☐	☐
Standardised	☐	☐	☐	☐	☐	☐	☐
Understanding	☐	☐	☐	☐	☐	☐	☐
Warmth	☐	☐	☐	☐	☐	☐	☐
Perspective -taking	☐	☐	☐	☐	☐	☐	☐

Appendix II Participant recruitment email

Email subject: A favour - participant recruitment for a Cambridge PhD project

Dear certain department administrator or student representative:

I am sorry to interrupt you during the busy term time. My name is Yishu Qin, a PhD student in psychology and education from Darwin College, University of Cambridge. I am looking for UK university students to do an online test for my project, looking at what motivates students to choose their majors at higher education level.

I have conducted the pilot survey in the University of Cambridge last year and I am now looking for more participants in London. **I wonder if you could kindly help me to spread my survey link to the students from your department?** For example, is it possible to post my advertisement on your Facebook group or add it to your weekly newsletter?

The study is completely anonymous, and all information collected will be kept strictly confidential. The consent form and the debriefing are there when participants click the link and they can withdraw at any point from the survey.

I have attached the letter of the ethics approval for you to review. This is a very vital part for my PhD thesis and thank you very much for your kind help! I am looking forward to your reply.

Warm Regards
Yishu Qin

PhD student
Psychology and Education
Faculty of Education
University of Cambridge

Darwin College
Cambridge
CB3 9EU
07709 396750

Appendix III Study Advertisement

Advertisement title: Win Amazon vouchers - spare 5 minutes to do an online test to explore your underlying beliefs about personality traits in different professions

My name is Yishu Qin and I am a PhD student from the Faculty of Education, University of Cambridge. I am interested in what motivates students to choose their majors at higher education level. My PhD project (in cooperation with Project Implicit from Harvard University) focuses on how people perceive different majors and whether their perceptions influence their career choices. I was wondering if you could spare 5 minutes to do this online test for my project?

The test itself is as simple as a pairing game and is fun and inspiring. The study is completely anonymous, and the instructions are there when you click on the link. This is a really vital part of my PhD thesis on psychology and education. You will have a chance to win £10, £25 or £100 in amazon vouchers as a reward for participation.

Below is my test link (please use a computer or a laptop, the test requires using keyboard so cannot be run on tablets or phones):

https://implicit.harvard.edu/implicit/Launch?study=/user/nlofaro/contract.qin.empathy/manager.expt.xml&refresh=true (Stage 2 study link)

or

https://implicit.harvard.edu/implicit/Launch?study=/user/nlofaro/contract.qin.empathy2/manager.expt.xml&refresh=true (Stage 3 study link)

For those who have participated in this study, please do not retake the test
If you have any idea to share with me, please contact Yishu Qin: yq228@cam.ac.uk

Appendix IV Introduction webpage in Stage 2 and 3

Thank you for taking part in my PhD project looking at individual differences in attitudes toward or beliefs about scientists. You will complete two tasks: one brief questionnaire and one Implicit Association Task (IAT) in which you will sort words into categories as quickly as possible. We will also ask you to provide some general information about yourself. We would like to compare possible differences among groups in their IAT performance and opinions, at least among those who decide to participate.

You should be able to complete these tasks in less than 10 minutes total. When you finish, you will receive a debrief report and have a chance to win £100 Amazon Voucher.

Data exchanged with this site are protected by SSL encryption, and no personally identifying information is collected. IP addresses are routinely recorded, but are completely confidential.

Before you start, please pay attention to the following:

- Check that you have a good internet connection and speed.
- Find a quiet place where you will not be disturbed.
- Make sure you have enough time to complete the IAT test in one sitting.

If you wish to continue, please click the "continue" button.

Continue

Appendix V Screenshots of the SE-IAT in Stage 2

Instructions to the SE-IAT

Welcome to the Implicit Association Test. You will use the "e" and "i" computer keys to categorize items into groups as fast as you can. These are the four groups and the items that belong to each:

Category	Items
Rationality	Consistent, Coherent, Deductive, Logical, Organised, Rational, Reasoning, Systematic
Empathy	Affectionate, Considerate, Caring, Emotional, Empathetic, Feeling, Thoughtful, Sensitive
Science	Chemistry, Physics, Mathematics, Engineering, Computer Science, Astronomy, Biology, Geology
Liberal Arts	Fine Arts, Linguistics, Philosophy, History, Literature, Sociology, Politics, Music

There are seven parts. The instructions change for each part. Pay attention!

Continue

Block 1 – Target discrimination (20 trials) for practice

Press "E" for
Liberal Arts

Press "I" for
Science

Part 1 of 7

Put a left finger on the **E** key for Liberal Arts.
Put a right finger on the **I** key for Science.
Items will appear one at a time.

If you make a mistake, a red **X** will appear. Press the other key to continue. Go as fast as you can while being accurate.

Press the **space bar** when you are ready to start.

Block 2 – Attribute discrimination (20 trials) for practice

Press "E" for
Rationality

Press "I" for
Empathy

Part 2 of 7

Put a left finger on the **E** key for Rationality.
Put a right finger on the **I** key for Empathy.

If you make a mistake, a red **X** will appear. Press the other key to continue. Go as fast as you can while being accurate.

Press the **space bar** when you are ready to start.

Block 3 – Initial combined task (20 trials) for practice

Press "E" for
Rationality
or
Liberal Arts

Press "I" for
Empathy
or
Science

Part 3 of 7

Use the **E** key for Liberal Arts and for Rationality
Use the **I** key for Science and Empathy
Each item belongs to only one category.

If you make a mistake, a red **X** will appear. Press the other key to continue. Go as fast as you can while being accurate.

Press the **space bar** when you are ready to start.

Block 4 – Initial combined task (40 trials) for testing

Press "E" for
Rationality
or
Liberal Arts

Press "I" for
Empathy
or
Science

Part 4 of 7

This is the same as the previous part.
Use the **E** key for Liberal Arts and for Rationality
Use the **I** key for Science and Empathy
Each item belongs to only one category.

Go as fast as you can while being accurate.

Press the **space bar** when you are ready to start.

Block 5 – Reversed target discrimination (20 trials) for practice

Press "E" for
Science

Press "I" for
Liberal Arts

Part 5 of 7

Watch out, the labels have changed position!
Use the left finger on the **E** key for Science.
Use the right finger on the **I** key for Liberal Arts.

Go as fast as you can while being accurate.

Press the **space bar** when you are ready to start.

Block 6 – Reversed combined task (20 trials) for practice

Press "E" for	Press "I" for
Rationality	Empathy
or	or
Science	Liberal Arts

Part 6 of 7

Use the **E** key for Science and for Rationality
Use the **I** key for Liberal Arts and Empathy
Each item belongs to only one category.

If you make a mistake, a red **X** will appear. Press the other key to continue. Go as fast as you can while being accurate.

Press the **space bar** when you are ready to start.

Block 7 – Reversed combined task (40 trials) for testing

Press "E" for	Press "I" for
Rationality	Empathy
or	or
Science	Liberal Arts

Part 7 of 7

This is the same as the previous part.
Use the **E** key for Science and for Rationality
Use the **I** key for Liberal Arts and Empathy
Each item belongs to only one category.

Go as fast as you can while being accurate.

Press the **space bar** when you are ready to start.

Appendix VI Self-report questionnaires in Stage 2

The SE-IAT questionnaire

1. Which statement best describes your view?
I strongly associate liberal arts with empathy and science with rationality.
I moderately associate liberal arts with empathy and science with rationality.
I slightly associate liberal arts with empathy and science with rationality.
I associate empathy and rationality with science and liberal arts equally.
I slightly associate science with empathy and liberal arts with rationality.
I moderately associate science with empathy and liberal arts with rationality.
I strongly associate science with empathy and liberal arts with rationality.

Please read each statement very carefully and rate how strongly you agree or disagree with the following statements. There are no right or wrong answers, or trick questions.

2. Scientists are less empathetic than liberal arts majors
3. Scientists are more rational than liberal arts majors

The ISSOS questionnaire

4. Scientists have fun with colleagues at work
5. Scientists maintain friendships with colleagues in other departments
6. Scientists do not have a lot of friends
7. Scientists are out of touch with what is happening in the world
8. Scientists have happy relationships
9. Scientists are cooperative
10. Scientists are family oriented
11. Scientists are insecure
12. Scientists are collaborative

The drop-down list to the right of each of the two statements provided seven options: Strongly agree – moderately agree – slightly agree –neither agree nor disagree– slightly disagree – moderately disagree – strongly disagree

Demographic Information

1. Which group from this IAT test do you identify with?
Science/Liberal arts/Other (please specify)
2. Gender: Female/Male
3. Age:_______
4. Ethnicity:_______
5. What is your current position? undergraduate/master/PhD/Postdoc/Other (please specify)
6. Which university you are currently working for, if it is applicable?_______
7. Major field of study or that of your highest degree:_______
   If you have a second major or minor, please indicate the field of study:_______
8. Prior to the current study session, had you ever completed an implicit association test? YES/NO
9. Country/region of primary citizenship:_______
10. Country/region of residence:_______
11. Are you a native English speaker? YES/NO

If not, if you took either or both of the following tests (IELTS and TOEFL), please provide your most accurate recollection of your best OFFICIAL scores:

IELTS:_______ TOEFL:_______

Appendix VII Debriefing webpage in Stage 2 and 3

Thank you for your participation!

How Does the Implicit Association Test (IAT) Work?

The IAT measures the strength of associations between concepts (e.g., Science and Humanities) and evaluations (e.g., Empathy). The main idea is that making a response is easier when closely related items share the same response key.

We would say that one has an implicit association of Empathy with Humanities compared to Empathy with Science if they are faster to categorize words when Empathy and Humanities share a response key relative to when Empathy and Science share a response key. Your automatic association may reflect your underlying attitude towards certain professions.

However, it is important to note that the IAT measures your implicit attitudes toward certain groups. It is designed with the aim to decrease the mental control available to produce the response, reducing the role of conscious intention as well as self-reflective, deliberate processes. You may not consciously endorse the implicit ideas revealed by the IAT. Your explicit attitudes and beliefs can be different from your implicit attitudes. Explicit attitudes and beliefs are ones that are directly expressed or publicly stated. An implicit attitude is one that is powerful enough to operate without conscious control. If you would like to know more about IAT measures in general, please go to the Q&A webpage of Project Implicit for more information.

If you have questions about this study, feel free to contact Ms Yishu Qin (PhD candidate): yq228@cam.ac.uk

If you would like to participate in the raffle for a chance to win £10, £20 and £100 in Amazon vouchers, please clike here.

Appendix VIII An example of the respondent session record

session_id user_id study_name session_date session_status creation_date last_update_date previous_session_id previous_session_schema referrer study_url user_agent

92233xx -1 contract.qin.empathy 09Aug2016 10:07:11 null 09Aug2016 10:07:11 NULL 0 null null /user/nlofaro/contract.qin.empathy/manager.expt.xml Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/52.0.2743.116 Safari/537.36

92233xx -1 contract.qin.empathy 09Aug2016 14:33:21 null 09Aug2016 14:33:21 NULL 0 null null /user/nlofaro/contract.qin.empathy/manager.expt.xml Mozilla/5.0 (Macintosh; Intel Mac OS X 10_11_5) AppleWebKit/601.6.17 (KHTML, like Gecko) Version/9.1.1 Safari/601.6.17

92233xx -1 contract.qin.empathy 09Aug2016 14:33:43 null 09Aug2016 14:33:43 NULL 9223340 r null /user/nlofaro/contract.qin.empathy/manager.expt.xml Mozilla/5.0 (Macintosh; Intel Mac OS X 10_11_5) AppleWebKit/601.6.17 (KHTML, like Gecko) Version/9.1.1 Safari/601.6.17

92233xx -1 contract.qin.empathy 09Aug2016 14:36:02 null 09Aug2016 14:36:02 NULL 0 null null /user/nlofaro/contract.qin.empathy/manager.expt.xml facebookexternalhit/1.1 (+http://www.facebook.com/externalhit_uatext.php)

92233xx -1 contract.qin.empathy 09Aug2016 14:36:23 null 09Aug2016 14:36:23 NULL 0 null null /user/nlofaro/contract.qin.empathy/manager.expt.xml facebookexternalhit/1.1

92233xx -1 contract.qin.empathy 09Aug2016 14:36:23 null 09Aug2016 14:36:23 NULL 0 null null /user/nlofaro/contract.qin.empathy/manager.expt.xml facebookexternalhit/1.1

Appendix IX Participant recruitment email addressing engineering students in Stage 3

Email subject: A favour – Engineering students needed for a Cambridge PhD project

Dear certain department administrator or student representative:

I am sorry to interrupt you during the busy term time. My name is Yishu Qin, a PhD student in Psychology and Education, University of Cambridge. I am looking for UK university students to do an online test for my PhD project, looking at what motivates students to choose their majors at higher education level.

I advertised the study last summer, but the engineering students are underrepresented in my current sample. Therefore, I am writing to gently ask **if you could kindly advertise my study to the students from the Engineering Department in particular**? For example, is it possible to post my advertisement on your Facebook group or add it to your weekly newsletter?

The study is completely anonymous, and all information collected will be kept strictly confidential. The consent form and the debriefing are there when participants click the link and they can withdraw at any point from the survey.

I have attached the letter of the ethics approval for you to review. This is a very vital part for my PhD thesis and thank you very much for your kind help! I am looking forward to your reply.

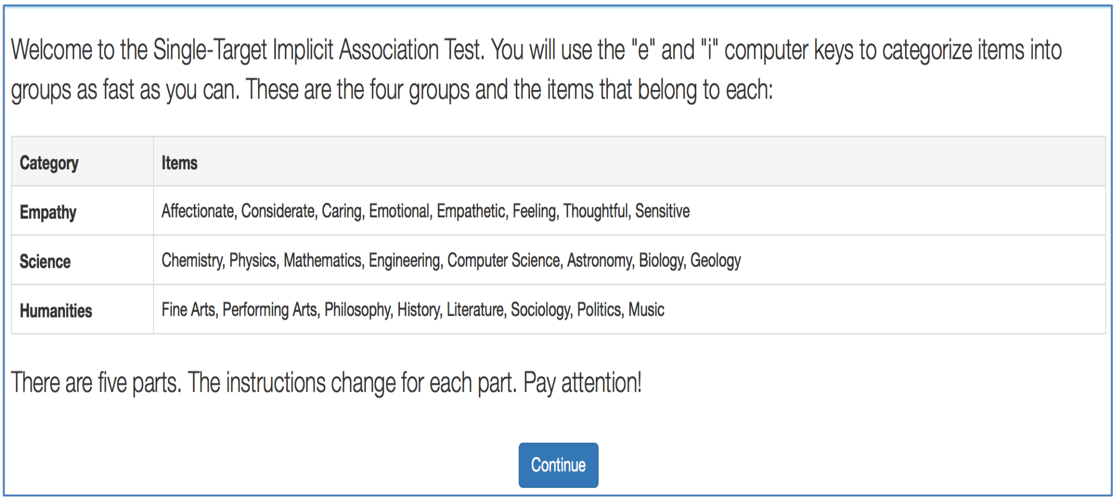
Warm Regards
Yishu Qin

PhD student
Psychology and Education
Faculty of Education
University of Cambridge

Darwin College
Cambridge
CB3 9EU
07709 3967

Appendix X Screenshots of the SSE-IAT in Stage 3

Instructions to the SSE-IAT

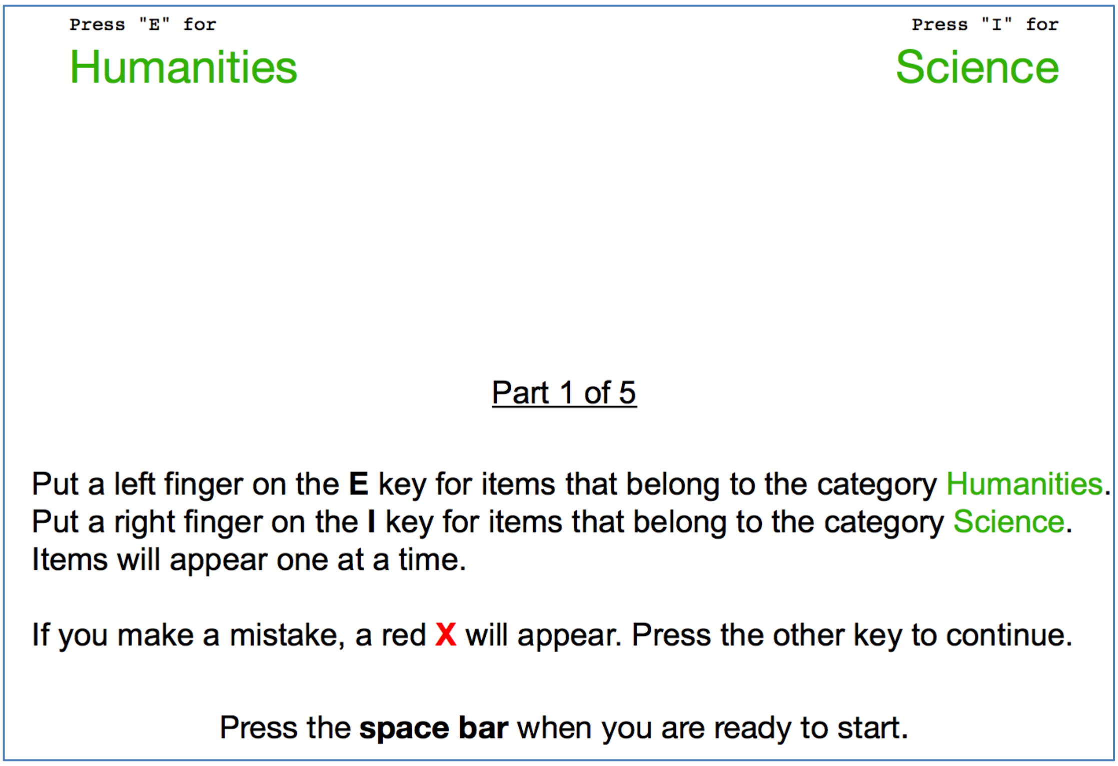

Welcome to the Single-Target Implicit Association Test. You will use the "e" and "i" computer keys to categorize items into groups as fast as you can. These are the four groups and the items that belong to each:

Category	Items
Empathy	Affectionate, Considerate, Caring, Emotional, Empathetic, Feeling, Thoughtful, Sensitive
Science	Chemistry, Physics, Mathematics, Engineering, Computer Science, Astronomy, Biology, Geology
Humanities	Fine Arts, Performing Arts, Philosophy, History, Literature, Sociology, Politics, Music

There are five parts. The instructions change for each part. Pay attention!

Continue

Block 1 – Target discrimination (20 trials) for practive



Block 2 – Initial combined task (40 trials) for practice

Press "E" for
Humanities

Press "I" for
Science
or
Empathy

Part 2 of 5

Put a left finger on the **E** key for items that belong to the category Humanities.
Put a right finger on the **I** key for items that belong to the category Science and for items that belong to the category Empathy.
Items will appear one at a time.

If you make a mistake, a red **X** will appear. Press the other key to continue.

Press the **space bar** when you are ready to start.

Block 3 – Initial combined task (50 trials) for testing

Press "E" for
Humanities

Press "I" for
Science
or
Empathy

Part 3 of 5

Put a left finger on the **E** key for items that belong to the category Humanities.
Put a right finger on the **I** key for items that belong to the category Science and for items that belong to the category Empathy.
Items will appear one at a time.

If you make a mistake, a red **X** will appear. Press the other key to continue.

Press the **space bar** when you are ready to start.

Block 4 – Reversed combined task (40 trials) for practice

Press "E" for
Humanities
or
Empathy

Press "I" for
Science

Part 4 of 5

Put a left finger on the **E** key for items that belong to the category Humanities and for items that belong to the category Empathy.
Put a right finger on the **I** key for items that belong to the category Science.
Items will appear one at a time.

If you make a mistake, a red **X** will appear. Press the other key to continue.

Press the **space bar** when you are ready to start.

Block 5 – Reversed combined task (50 trials) for testing

Press "E" for
Humanities
or
Empathy

Press "I" for
Science

Part 5 of 5

Put a left finger on the **E** key for items that belong to the category Humanities and for items that belong to the category Empathy.
Put a right finger on the **I** key for items that belong to the category Science.
Items will appear one at a time.

If you make a mistake, a red **X** will appear. Press the other key to continue.

Press the **space bar** when you are ready to start.

Appendix XI Self-report questionnaires in Stage 3

The SE-IAT questionnaire

1. Please choose one statement that best describes your view about the relationship between science and empathy.
   I strongly associate science with empathy.
   I moderately associate science with empathy.
   I slightly associate science with empathy.
   I do not associate science with empathy at all.

2. Please choose one statement that best describes your view about the relationship between humanities and empathy.
   I strongly associate humanities with empathy.
   I moderately associate humanities with empathy.
   I slightly associate humanities with empathy.
   I do not associate humanities with empathy at all.

3. Scientists are less empathetic than liberal arts majors

The ISSOS questionnaire

4. Scientists have fun with colleagues at work
5. Scientists maintain friendships with colleagues in other departments
6. Scientists do not have a lot of friends
7. Scientists are out of touch with what is happening in the world
8. Scientists have happy relationships
9. Scientists are cooperative
10. Scientists are family oriented
11. Scientists are insecure
12. Scientists are collaborative

The drop-down list to the right of each of these statements provides seven options: Strongly agree – moderately agree – slightly agree –neither agree nor disagree– slightly disagree – moderately disagree – strongly disagree

The CAS questionnaire

In your future career, how likely is it that you will:

13. Get an advanced degree in science?
14. Have a very successful career in science?
15. Have a lifelong career in science?

The drop-down list to the right of each statements provided seven options:
Very likely – likely – mildly likely – not sure – mildly unlikely – unlikely – very unlikely

Demographic Information

1. Please specify the field you identify with the most:
Options: Arts & Humanities; Engineering & Technology; Life Science & Medicine; Natural Sciences; Social sciences and Management; Other
2. Please specify your gender
Options: Female/Male
3. Please specify your age:_______
4. Ethnicity:
Options: White; Asian/Asian British; Black African/Caribbean/Black British; Other ethnic group
5. Which degree are you currently pursuing?
Options: Bachelor's degree; Master's degree; Doctorate degree; Not currently a university student
6. Which university are you currently affiliated with? _________ If not currently a student, enter you most recent institution: ____________
7. Please indicate the major field of study or that of your highest degree:
If you have a second major or minor, please indicate the field of study:
8. Prior to the current study session, had you ever completed an implicit association test? Options: YES; NO; Don't know
9. What is your country of birth?
10. What is your country of residence?
11. Are you a native English speaker? YES/NO
If not, if you took either or both of the following tests, please provide your most accurate recollection of your best OFFICIAL scores:
IELTS: ________ TOEFL:________

Appendix XII An example of the recorded trial latencies in SSE-IAT

Block number	Block pairing definition	Category name	Trial item	Trial response	Reaction time (ms)	Trial error
4	Humanities, Science+Empathy	Science	Chemistry	Right	572	0
4	Humanities, Science+Empathy	Science	Mathematics	Right	751	0
2	Humanities+Empathy, Science	Science	Biology	Right	690	0
3	Humanities+Empathy, Science	Science	Mathematics	Right	512	0
3	Humanities+Empathy, Science	Science	Engineering	Right	738	0
5	Humanities, Science+Empathy	Humanities	Politics	Left	648	0
5	Humanities, Science+Empathy	Empathy	Thoughtful	Right	1885	1
1	Humanities, Science	Science	Computer Science	Right	813	0
1	Humanities, Science	Science	Chemistry	Right	768	0
1	Humanities, Science	Humanities	Fine arts	Left	594	0

This table shows sample data for individual trials. Note that an item (e.g. mathematics; chemistry; etcetera) receives different reaction times under different blocks. Scores are calculated not on an individual item basis, but based on comparisons between items in compatible and incompatible blocks. Refer to Procedure (Table 5.12, pg.137) for details about blocks.