Unconscious bias and higher education

Equality Challenge Unit

Acknowledgments

This literature review was undertaken by Cornish Consultancy Services. ECU would like to thank Tinu Cornish and Pete Jones for their work on it.

Additionally, ECU would like to thank Ian Clarke, Leeds Metropolitan University and Fiona McClement, University College London for their comments and assistance.

Further information

Claire Herbert
info@ecu.ac.uk

Unconscious bias and higher education

Contents

Overview 1
Unconscious or implicit bias? 1
Format of this report 2
Methodology for the literature review 3

1 Summary of recommendations 4
Look for evidence of where implicit bias may be having an impact 4
Create a culture of equality 5
Ensure transparency 5
Reduce stereotypes 6
Use the role of chair to guide the selection process 8
Prepare individual selectors 8
Ensure the qualities of different candidates are valued in the same way 9

2 Index of psychological studies 10
Existence and impact of implicit bias 10
Mitigating implicit bias 12

3 Psychological theories linked with implicit bias 15
Explicit bias 15
Implicit bias 15
Social categorisation theory 18
Stereotyping and attitudes 19
The link between bias and discrimination 23
The implicit association test (IAT) 25

4 Impact of implicit bias on behaviour and actions 26
Recruitment and selection situations 26
Implicit bias, body language and interactions with others 35
Impact of bias in other situations and sectors 38

© Equality Challenge Unit, September 2013

5 Methods and techniques for reducing implicit bias 40
Discounting commonly held stereotypes 40
Using context to influence implicit responses 50
Changing the way an out-group member is evaluated and categorised 53
Using contact to change the level of threat evoked by an out-group 54
Using motivation to change responses to an out-group 57
Encouraging people to take responsibility for their implicit biases 61

Conclusions 70

Appendix 1 Methodology 71
Aim 71
Type of review 71
Identifying a longlist of literature 72
Identifying a shortlist of literature 74

Appendix 2 Implicit association tests 76
Background to the IAT 76
The IAT procedure 77
Criticisms of the IAT 78
Variables that may have an effect on the IAT 82
Predictive validity of the IAT 85

Bibliography 88

Overview

Interest in unconscious bias has increased over the past few years, with a significant body of supporting evidence that has allowed the concept to become widely recognised both inside and outside the field of psychology.

This report is intended to help higher education institutions (HEIs) understand unconscious bias, and discover how to reduce its impact, with particular reference to staff selection and recruitment.

Unconscious bias is a term used to describe the associations that we hold which, despite being outside our conscious awareness, can have a significant influence on our attitudes and behaviour. Regardless of how fair minded we believe ourselves to be, most people have some degree of unconscious bias. The means that we automatically respond to others (eg people from different racial or ethnic groups) in positive or negative ways. These associations are difficult to override, regardless of whether we recognise them to be wrong, because they are deeply ingrained into our thinking and emotions.

Acknowledging and taking responsibility for unconscious bias is not just a moral imperative, it is also financially and reputationally important. For HEIs, making biased decisions affects the recruitment and selection of staff and students, and the ability of those staff and students to achieve their full potential.

Unconscious or implicit bias?

In the rest of this report, we refer to 'implicit' rather than 'unconscious' bias, in line with most studies in this area. These terms describe broadly similar biases and are often used interchangeably. They do, however, have slightly different meanings.

Unconscious bias refers to a bias that we are unaware of, and which happens outside of our control. It is a bias that happens automatically and is triggered by our brain making quick judgments and assessments of people and situations, influenced by our background, cultural environment and personal experiences.

Implicit bias refers to the same area, but questions the level to which these biases are unconscious especially as we are being made increasingly aware of them. Once we know that biases are not always explicit, we are responsible for them. We all need to recognise and acknowledge our biases and find ways to mitigate their impact on our behaviour and decisions.

Format of this report

This report is based on a literature review of psychological experiments exploring unconscious/implicit bias. Accompanying the literature are some of the policy implications and recommendations for HEIs that can be drawn out of the findings.

The purpose of the review is to look at the impact of implicit bias in higher education settings, and specifically on recruitment and selection. However, the review also includes studies in other situations and sectors where they are transferable and relevant.

1 Summary of recommendations

Section 1 includes Equality Challenge Unit's (ECU) recommendations based on the literature review.

2 Index of psychological studies

The experiments included in the literature review are indexed for ease of reference.

3 Psychological theories linked with implicit bias

Section 3 explains some of the key psychological theories used throughout the report. Readers familiar with psychology or implicit bias may wish to go straight to section 4.

4 Impact of implicit bias on behaviour and actions

Section 4 summarises research looking at the impact of implicit bias in decision-making situations, including judges' legal decisions and doctors' diagnoses and treatment decisions, as well as recruitment and selection decisions in higher education. The purpose of this section is to understand the consequences of implicit bias in decision-making situations to get some idea about how it might have an impact in a higher education setting.

5 Methods and techniques for reducing implicit bias

Section 5 explores studies that have tried to interfere with the process of implicit bias to prevent an impact on decision making. The studies are accompanied by policy implications and have informed the conclusions and recommendations.

The majority of research into implicit bias has been conducted in laboratories, rather than real-life settings. Section 5 aims to look at how these experiments could relate to real-world situations and be transferred into practice within HEIs.

Methodology for the literature review

There is a large amount of literature on implicit bias. The primary aim of this review is to use existing literature, research and publications to explore the ways in which black and minority ethnic (BME) staff may be affected by implicit bias in recruitment and selection, and what institutions can do to lessen the impact.

The review assumes that implicit bias exists, and that if BME people are affected by implicit bias in recruitment and selection then other minority groups will also be affected. No attempt has been made to ascertain whether BME people are affected to a greater or lesser extent than any other group of people.

This review uses a mixed-methods approach. The aim has been to review the existing evidence surrounding implicit bias and to identify evaluated techniques for reducing its impact on decision making so that HEIs can implement evidence-based initiatives.

The full methodology is available at appendix 1.

1 Summary of recommendations

Recognition of implicit bias must not replace an acknowledgment that explicit bias and discrimination exist and continues to be an issue in the higher education sector.

While the methods for measuring and predicting implicit biases may not be perfect, there is little doubt that it does exist and also has the potential to affect our actions and decision making.

Greenwald *et al* (2009) (see p85) highlight that explicit measure of socially-sensitive bias may not be very accurate. Regardless of how fair-minded we think we are, we are all likely to benefit from implicit bias interventions.

People and institutions not only have a moral responsibility for their implicit biases, but a business responsibility; institutions need to be efficient and effective, and decisions and actions need to be taken based on evidence and fact, rather than stereotypes and hunches.

While implicit bias is likely to be relevant to many areas of an institution's work, for example appraisals and grievances, Research Excellence Framework submissions, student admissions and course evaluations, the report's recommendations are based on the findings from the literature review that focus on recruitment and selection.

Look for evidence of where implicit bias may be having an impact

To determine where bias might exist, a first step is to examine existing data and analyse where bias may have had an impact.

Many institutions will already have done this, but monitoring the numbers of people applying, being shortlisted, and being selected at interview by protected characteristic will enable trends to be uncovered. This could apply to both external recruitment and internal promotions. Where there are significant trends, for example where there is a high proportion of black applicants, but a lower proportion of black applicants shortlisted and a lower proportion still of black applicants offered the job, further exploration should be carried out.

Wherever possible, we recommend the data is analysed by specific ethnic group, rather than 'black and minority ethnic'. Where the numbers are small, analysis can be conducted by aggregating data for all posts at a particular grade, or for a set time period (eg, every five years).

Create a culture of equality

Findings from the researchers Ziegert and Hanges (2005) and Carlsson and Rooth (2007) show that having written equality and diversity policies is insufficient; policies need to be put into practice. Saying an institution is committed to equality and diversity is not the same as an institution demonstrating it is committed to equality and diversity.

If managers make it clear that they are committed to equality, for example by attending equality events, debating and discussing the issues, and holding themselves and others to account for lack of progress, then the rest of the institution is likely to follow.

Ensure transparency

Paperwork

Ensure that all shortlisting exercises and interviews are properly documented in a standard and consistent manner to show why people were shortlisted and recruited, and how they were more suitable for the post compared with other applicants. Also consider the language used to justify such decisions; are they valid decisions, based on the selection criteria, or are they expressed as subjective opinions, such as an applicant's ability to 'fit in with the team'. Such phrases may be triggered by 'loyalty' rather than 'equality' (see Zogmaister *et al*, 2008), which may benefit the interviewers' in-group (see p18).

Consider auditing the paperwork from previous recruitment campaigns to help ensure that a proper process is in place. Having robust paperwork on recruitment decisions will also help to protect an institution from legal challenge.

Legal challenge

An unsuccessful applicant challenging an appointment decision may request the notes made during an interview process. In 2009, in response to a Freedom of Information Act request from an unsuccessful job applicant, the information commissioner directed Leicester City Council to provide anonymised application forms from other applicants. HEIs need to be prepared to demonstrate through their paperwork how fair and transparent decisions have been made in recruitment and selection decisions.

Anonymous shortlisting

Wherever possible, HEIs should consider anonymous shortlisting of candidates. All of the CV studies highlight the level of conscious and unconscious bias that can influence shortlisting decisions when irrelevant information is included on application forms. HEIs need to consider:

= **how jobs are advertised**
Is it fair, do head hunters use transparent processes that promote equality, does everyone have an equal chance of applying?

= **how people initially apply for vacancies**
Is there a standard application form, or do people submit a CV? Having an application form ensures that everyone is submitting the same types of information in the same order, so they can be compared fairly.

= **whether shortlisting can be done anonymously**
Particularly for professional and support positions, human resources (HR) processes could be adapted to remove information such as name, school, university, all monitoring data, and anything else that is irrelevant to the application.

= **how qualifications are recognised**
Is the shortlisting panel prepared to consider overseas and alternative qualifications in a fair and transparent way?

Reduce stereotypes

Institutions should look for situations where they can promote counter-stereotypical images of underrepresented groups. This is important for staff and students. If people are always presented with an image of certain groups of people in certain roles then the association becomes automatic and influences our view of that group of people and that role. For example, 'vice-chancellor' should not be synonymous with 'male', in the same way that 'secretary' should not be synonymous with 'female'.

Good practice

- Ensure a diversity of guest speakers and lecturers are invited for special events and conferences. Collect and analyse monitoring data on guest speakers and lecturers and address any imbalance that is noticed.
- Conduct an image audit of the institution. Consider images within board and conference rooms, paintings on walls, images in marketing materials and websites. Where images are of board and senate members, or previous chancellors and vice-chancellors, consider where they would be encountered by visitors and what visitors may interpret from them. At the same time, it is important to portray an accurate image of the institution, and so a balance is needed. If an institution looks incredibly diverse in its prospectus, but in reality is not, then prospective students and staff may feel misled.
- Consider the events and seminars organised throughout the year and how they could contribute to mitigating generally held stereotypes about different groups of people. This should include embedding diversity within non-specific events as well as organising bespoke events such as black history month, international women's day, etc. For example, the institution may organise tours of the local area for new staff and students; tours could include the black history and prominent female history of the area, rather than the black and female history being included in a separate tour.
- Consider the content of the curriculum and how it could be diversified to ensure a balanced view point. Again, this should be embedded into existing courses, rather than being added as an additional element.

Training for staff

Alerting staff to their implicit biases and asking them to tackle them is not an easy task. The studies in this report show that it is possible to make the situation worse and for individuals to become defensive and/or avoid interacting with people in case they are biased towards them.

Institutions could consider introducing training in a supportive, unthreatening environment to give staff the chance to think about their biases in a constructive way. This is especially important for people who may be undertaking important decisions such as in recruitment or admissions.

Use the role of chair to guide the selection process

Chairs of selection panels face a particular challenge. They need to frame a context and support mindsets that will minimise categorisation and implicit bias, but also do this in a way that does not make majority panel members feel they are being put under undue pressure to conform.

Chairs may want to plan a preamble that highlights diversity as a shared, freely chosen value. They may want to discuss the role of association in creating implicit bias – where minorities are underrepresented in the post being recruited for, panel members may be influenced by that stereotype.

When considering questions of fit, whether to the team or the organisation, the chair may want to ensure these discussions do not highlight the need for loyalty and therefore increase the potential for implicit bias (Zogmaister *et al*, 2008).

When comparing candidates from majority and minority groups the chair may want to direct the panel to consider the ways in which they are similar (de-categorisation) (Hall *et al*, 2009) before looking at the ways in which they differ.

Prepare individual selectors

For the individual selector a key consideration is that associations can produce bias regardless of their endorsements of those biases. If they are selecting for a post in which a group is underrepresented they can still be implicitly affected by those biases, regardless of their commitment. This was demonstrated by studies showing that female psychologists (Steinpreis *et al*, 1999) and women in STEMM (science, technology, engineering, mathematics and medicine) departments (Moss-Racusin *et al*, 2012) are just as likely to discriminate against female candidates as their male counterparts. Before selecting for such posts, selectors may want to familiarise themselves with examples; for example, reading about famous female scientists/leaders and set an implantation intention (eg 'think female think leader').

Throughout the selection process, the selector may want to be mindful of how consideration of issues such as 'fit' may focus loyalty towards the in-group, and make a conscious effort to highlight the values of equality and multiculturalism. The individual selector will also want to remind themselves of an appropriate counter-stereotypic implantation intention.

Ensure the qualities of different candidates are valued in the same way

The same qualities can be perceived very differently depending on whether they are demonstrated by a member of the majority or the minority group (Eagly and Carli, 2007). Uhlmann and Cohen (2005) explored this in a recruitment context and found that merit was redefined to fit the profile of the candidate of the preferred (in this case) gender:

'In three studies, participants assigned male and female applicants to gender-stereotypical jobs. However, they did not view male and female applicants as having different strengths and weaknesses. Instead, they redefined the criteria for success at the job as requiring the specific credentials that a candidate of the desired gender happened to have.'

Rather than assume that candidates of a particular gender did or did not have particular necessary attributes for a job, they instead moulded the key necessary attributes for the job depending on the attributes demonstrated by candidates of the 'preferred' gender.

This emphasises the need for anonymous shortlisting and clear job descriptions with weightings applied to the various essential qualities necessary to perform the job well. Before any candidates are shortlisted or interviewed, the panel should be clear about what they are looking for in relation to skills, knowledge and experience, and these should not change once candidates are assessed.

This is especially important in promotion situations. Where employees are clear about what they need to demonstrate to be promoted then they can develop those skills and gain relevant experience. Where the attributes needed change, bias can creep in and prevent some groups of people from being promoted at the same rate as others.

2 Index of psychological studies

Existence and impact of implicit bias

Organisational culture	
Ziegert and Hanges (2005) p26	Explores the importance of an 'equality-friendly' organisational culture and how implicit biases can manifest in decisions and actions where the culture is racially biased.
Implicit bias in shortlisting CVs	
Wood *et al* (2009) p28	Commissioned by the Department for Work and Pensions, the research found that applicants with typically white British names are more likely to be shortlisted for jobs than those with names associated with minority ethnic backgrounds.
Steinpreis *et al* (1999) p29	Found that both male and female academic psychologists were more likely to want to employ a male early career researcher than an equally qualified female early career researcher. This evened out in later career decisions once an academic had more experience on their CV.
Moss-Racusin *et al* (2012) p30	Staff in a science faculty rated male applicants for a laboratory manager role as more competent than equally qualified female candidates. They also chose a higher starting salary for male candidates.
Carlsson and Rooth (2007) p30	Demonstrated the reduced likelihood of being shortlisted for jobs in Sweden with a Middle Eastern name as opposed to a Swedish name.
Carlsson and Rooth (2008) p32	Demonstrated the added discrimination of having foreign qualifications as well as a foreign name in shortlisting decisions.
Rooth (2010) p33	Explored whether the discrimination identified in Carlsson and Rooth (2007, 2008) was implicit or explicit, and found a correlation between implicit racial bias and shortlisting decisions.
Agerström *et al* (2007) p34	Found that both a group of students and a group of employers demonstrated implicit preferences for native Swedish men over Arab-Muslim men and looked at the impact of that on hiring preferences.

Implicit bias in body language	
McConnell and Leibold (2001) p35	Explored differences in the way research participants behaved and responded to a white v black researcher and how this correlated to their implicit and explicit bias scores. Also included is the debate around the reliability of the study by Blanton *et al* (2009). Psychologists on both sides of the debate found that participants behaved differently to the two different researchers.
Shelton *et al* (2005) p38	Explored how friendly research participants found each other during a conversation on a set topic. They found black participants were more likely to view their white partner favourably where their partner showed higher levels of racial bias through an IAT.
Richeson and Shelton (2005) p39	Explored whether black and white research participants were able to identify differences in the body language of white people when they interacted with people from different ethnic groups. They found that black participants were more able to identify biased body language in specific contexts. They hypothesised that this might be because their own personal experiences of discrimination made them more aware of the differences.
Implicit bias in medicine	
Green *et al* (2007) p38	Explored the implicit bias in doctors and found they were more likely to prescribe effective drugs to white rather than black patients.
Implicit bias in social policy	
Pérez (2010) p39	Found that implicit attitudes significantly influence individual preferences for social policy, in this case illegal and legal immigration policy.

Mitigating implicit bias

Discounting commonly held stereotypes	
Dasgupta and Greenwald (2001) p41	Explored the impact of positive exemplars of black people and negative exemplars of white people on individuals' level of implicit bias. Pro-white bias was significantly reduced.
Plant *et al* (2009) p41	Explored the positive impact President Obama has had in reducing people's implicit racial bias.
Dasgupta and Asgari (2004) p42	Found that female participants' gender biases were significantly affected when they were exposed to women in non-stereotypical contexts.
Olson and Fazio (2006) p45	Participants watched a random sequence of images without knowing the purpose of the study. Throughout, images representing black people with good and white people with bad characteristics were interspersed and reduced their implicit bias.
Ramasubramanian (2007) p46	Found that receiving training to view the media critically and receiving counter-stereotypical news articles reduced participants' bias towards an African-American news story.
Park *et al* (2007) p46	Found that participants exposed to literature highlighting the positives of Arab-Muslim culture reduced their implicit bias against Arab Americans.
Crisp and Nicel (2004) p47	Participants responded with the word 'yes' whenever they saw counter-stereotypical stimuli of an out-group member in an attempt to reduce bias.
Kawakami *et al* (2005) p47	Participants matched female faces with non-stereotypic words, and their gender bias reduced as a consequence.
Gawronski *et al* (2008) pp47–48	Participants pressed 'yes' every time they saw a face and word combination that was counter-stereotypic, and 'no' every time they saw a stereotypic combination. They found the 'no' condition actually increased bias.
Impact of context on implicit bias	
Stewart *et al* (2010) p50	Participants trained to use situational rather than dispositional explanations for stereotypical behaviour showed reduced levels of negative stereotyping and automatic bias.
Barden *et al* (2004) p51	Explored the impact of social role and context on implicit bias, for example different reactions to a black face in a ghetto background than a black face outside a church.
Zogmaister *et al* (2008) p53	Considered how priming interviewers with 'loyalty' sentences may increase implicit bias in interviews, whereas priming with 'equality' sentences may decrease bias.

Changing out-group evaluation and categorisation	
Hall *et al* (2009) p53	Aimed to highlight the similarities people had with their out-groups, rather than their differences.
Lebrecht *et al* (2009) p54	Encouraged participants to first and foremost see people as people, rather than to put people into categories.
Aberson (2004) p55	Participants who reported having close friends within the target underrepresented group exhibited less implicit bias towards that group.
Shook and Fazio (2008) p55	Using college students sharing dormitory rooms, researchers found that interracial room sharing led to improvements in automatically activated racial attitudes compared with same-race room sharing.
Turner and Crisp (2010) p56	Imagined contact, whereby participants imagined having a positive encounter with an out-group member reduced their implicit bias towards that group.
Motivation to control bias	
Allen *et al* (2010) p52	Explored whether reductions in implicit bias were caused by positive context or by the individuals' inhibition of their biases.
Devine *et al* (2002) p57	Primed participants with varying levels of motivation to control their bias.
Legault *et al* (2011) p58	Found a reduction in participants' explicit bias where they were primed with a brochure designed to facilitate their internal motivation to control prejudice.
Richeson (2004) and Correll *et al* (2008) pp59–60	Aimed to look at the difference in impact of colour-blind v multicultural approaches. The findings varied.
Stewart and Payne (2008) p62	Participants were primed with either 'safe', 'accurate' or 'fast'. Their responses were then measured when shown images of black and white faces with guns. Those primed with 'think safe' showed a reduction in racial bias.
Mendoza *et al* (2010) p62	This study found that where individuals primed themselves with 'if I see a person, then I will ignore their race' they found a reduction in their bias.
Vorauer (2012) p63	Participants who took a race IAT and then immediately interacted with someone from a minority ethnic group were perceived as more unfriendly than those who had not.

Morris and Ashburn-Nardo (2009) and Henry-Darwish and Sanford (2012) p64	Found that taking an IAT in a supportive setting with feedback had a positive effect.
Castillo and Brossart (2007) p65	Taking a multicultural counselling course had a positive impact on implicit attitudes compared with taking a regular introductory counselling course. However, the positive contact with minority ethnic trainers may have had an impact.
Rudman *et al* (2001) p65	Students enrolling on a bias and conflict seminar found they had reduced anti-black bias, but the results may be affected by contact with black lecturers and students.
Carnes and Devine (2012) and Devine *et al* (2012) p66	Used bias-reduction training programmes in a university setting.

3 Psychological theories linked with implicit bias

Explicit bias

Explicit attitudes are the views and opinions that we are consciously aware of, for example our attitude towards a particular political party. We may not necessarily be open and honest about our explicit attitudes with other people, but we are aware of them.

Explicit attitudes are typically measured by self-report questionnaires in which people answer a series of questions about their views and attitudes towards something (eg a group of people sharing a particular characteristic). When completing explicit-bias tests, people tend not to answer truthfully but do so in ways that they consider socially desirable; people may want to portray a particular image of themselves or seek to protect an egalitarian self image (Devine, 1989).

Implicit bias

Implicit attitudes, on the other hand, are those views and opinions that we may not be aware of. They are evaluations that are automatically triggered when we encounter different people or situations, and commonly function without a person's full awareness or control (Greenwald and Banaji, 1995).

As people are unable to control and manage their implicit attitudes and biases and cannot easily hide them in the same way as their explicit biases, they are seen as a more reliable measure of actual prejudice. Measures of implicit attitudes are often more reliable in predicting behaviour (Dovidio *et al*, 1997).

Despite levels of explicit prejudice falling (Abrams and Houston, 2006), actual discrimination remains a continuing problem for many sections of society. Understanding implicit bias – what causes it, how it impacts decision making and what can be done to moderate it – is important if there is to be a narrowing of the gap between the ideals we aspire to and reality.

Where do implicit biases come from?

Implicit biases are a form of bias pertaining to the mental processes of perception, memory, judgment and reasoning, also known as *cognitive* bias. Cognitive biases arise because our human decision-making processes are not just factual or objective, but are influenced by a variety of factors including:

- **information-processing short cuts** – technically referred to as heuristics that could include instances where we might use our intuition, or common sense based on what we think we know (see also, **Social categorisation theory** on p18)
- **motivational and emotional factors**, for example from our own personal experiences
- **social influences**, such as the media and stereotypes

(Schwarz, 2000)

Other cognitive biases and thinking errors

The errors we make in judging people are similar to other cognitive errors and they have common causes. Many of us will have learned how perceptual illusions shed light on how perception and cognition work (Eagleman, 2001).

Figure 1

Hidden dog illusion
Our perception imposes meaning on a random collection of dots

Successive studies (see Kahneman, 2011 for a review) have shown that there are distinctive patterns in these thinking errors we make. By understanding these we shed some light on the nature of thinking and human bias.

In their book *The invisible gorilla*, Chabris and Simons (2010) highlight how our everyday intuitions about the world are flawed. They describe an experiment in which subjects were asked to watch a short film of two teams passing a basketball. The participants are asked to count the number of passes made by one of the teams. Participants were so absorbed in the task they did not notice a woman wearing a gorilla suit crossing the basketball court and thumping her chest before moving on. She does this relatively slowly and is on screen for nine seconds (see **www.theinvisiblegorilla.com/videos.html**).

'Amazingly roughly half the subjects do not see the gorilla'
(Chabris and Simons, 2010: p6)

This cognitive error, known as inattentional blindness, arises because of a lack of attention to an unexpected object.

Research like this provides the foundations for our understanding of how we make judgments and decisions about other people. In his book *Thinking, fast and slow*, Kahneman (2011) suggests that our mental processes can be conceptualised as operating through two systems:

1. fast, automatic, frequent, emotional, stereotypic, subconscious
   'System 1 operates automatically and quickly, with little or no effort and no sense of voluntary control.'
2. slow, effortful, infrequent, logical, calculating, conscious
   'System 2 allocates attention to the effortful mental activities that demand it, including complex computations.'

One of the reasons people may not notice implicit bias in themselves is that other 'system 1' thinking errors or cognitive biases act to make the implicit bias seem reasonable and justifiable (Reskin, 2000). For example, 'confirmatory bias' leads us to seek out and prioritise information that confirms what we already thought, and to be blind to evidence that contradicts this (Dougherty *et al*, 1994).

Social categorisation theory

Biases arise partly because of the way our brains rapidly categorise other people for expedient and efficient decision making. When people interact the vast amount of information available to them about each other is cognitively overwhelming: we simply cannot process everything about each new person we encounter.

As a result, information about people and objects is compressed, grouped and placed into easy-to-use categories. This enables us to make rapid judgments about new people and situations without having to process everything in great detail. Social psychologists call this *social categorisation* whereby we routinely and rapidly sort people into groups rather than think of each as unique (Tajfel and Turner, 1979).

The advantage of categorisation is that it allows us to save time and effort when processing information about others, thereby allowing us to pay attention to other tasks or more novel information with our limited processing resources. However, social categorisation also brings with it a second powerful process in the form of stereotyping and bias. This can be particularly damaging when we are rapidly judging people categorised as part of our 'out-group' compared with those categorised as part of our 'in-group'.

In-group v out-group

Our in-group members are likely to be those we have categorised as having a shared identity, characteristic or interest to us:

> 'Human social cognition has adapted to ancestral environments defined by simple 'us' versus 'them' category boundaries, enabling a clear way of distinguishing friend from foe. This preference for representing categorical differences is reflected in contemporary social psychological research.
>
> Extensive evidence shows how categorical thinking enables others to be identified, in a matter of milliseconds, as in-group or out-group members.'
>
> (Crisp and Meleady, 2012: p858)

Stereotyping and attitudes

The word 'stereotype' was first used in 1922 by the journalist Walter Lippman to describe the cognitive and behavioural patterns of humans. Until that time it had been a term used in the printing industry to describe a process whereby fixed casts are produced. Today, the term stereotype is used by social scientists and the public alike to describe the construction of a conventional, formulaic and oversimplified conception, opinion or image of a social group. This impression about a group or person is often based on limited knowledge of that group or person.

Researchers examining stereotypes believe that the two constructs of stereotype and attitude are closely related. An attitude is an expression of favour or disfavour towards a person, place, thing or event. Although attitudes can be both positive and negative, the term prejudice is typically used to refer to what is considered an unfair negative attitude toward a social group or a member of that group.

In a study in which participants were unknowingly influenced, they responded differently with the experimenter depending on the prompt: participants primed with the concept of rudeness interrupted the experimenter more quickly and frequently than participants primed with polite-related stimuli.

In a second experiment, participants shown an elderly stereotype walked more slowly down the hallway when leaving the experiment than control participants (Bargh *et al*, 1996). These studies suggest a correlation between how people think/feel and how they act/behave.

Although implicit biases operate in the unconscious mind, they can have significant real-world effects. Empirical research shows that automatic bias plays an important role in producing discriminatory behaviour and judgments, and that measures of implicit bias are significant predictors of the level of discriminatory behaviours and judgments.

Stereotype content

Bias and stereotyping are not just based on negative perceptions; some apparently positive stereotypes can be used to justify the exclusion or oppression of certain groups in society. According to the 'stereotype content model' (Fiske *et al*, 2002), stereotypes are proposed to vary along two dimensions:

1. Warmth
   'People infer warm (or cold) intent from respectively cooperative or competitive structural relationships between individuals or groups. That is, those groups who cooperate appear warm and trustworthy; those who compete appear cold and untrustworthy, even exploitative.'
2. Competence
   'Knowing a stranger's intentions solves only part of the dilemma, because one must know the Other's capability to enact those intentions. An incompetent foe poses less threat and an incompetent friend offers less benefit than their more competent counterparts. People infer this competence (capability, skill) from apparent status (prestige, economic success).'

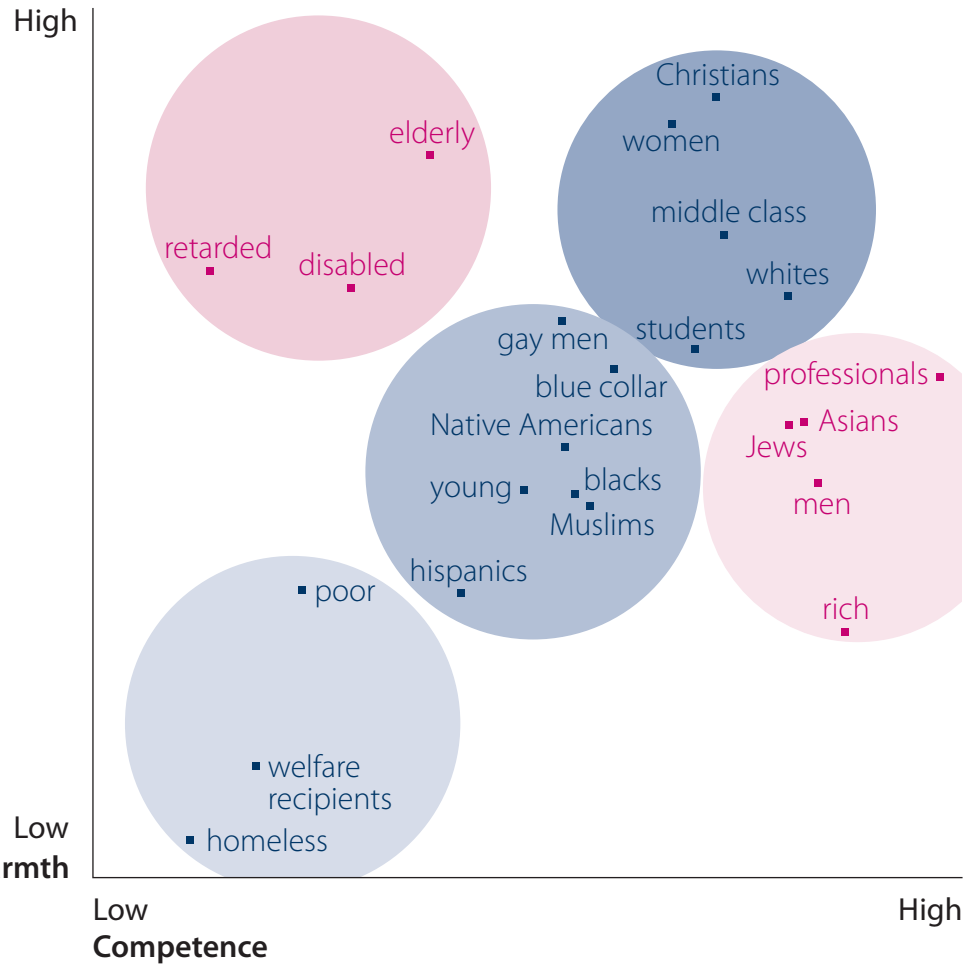
   (Fiske, 2012: p34)

Examples based on US research include older people being seen as warm but not competent and rich people as competent but not warm (see overleaf).

While the specific contents of stereotypes will vary across cultures, the dimensions (warmth and competence) have been found to be similar (Cuddy *et al*, 2009).

Figure 2 represents the results of a study conducted by Fiske *et al* (2002) exploring societal stereotypes of different groups. Following on from the findings of a previous study, college students in the USA were asked to complete a questionnaire measuring society's (rather than their own individual) warmth towards different groups and their assessment of that group's competence, status and competition. It should be noted that this is based on an American study and therefore some of the terminology is different to that commonly used in the UK.

Figure 2



(Fiske *et al*, 2002: p892)

Belief in a stereotype

A person does not have to endorse a stereotype to still be affected by it. For example, recent research found that female academics in science, technology, engineering, mathematics and medicine departments were just as likely to discriminate against female candidates for employment as their male colleagues (Moss-Racusin *et al*, 2012).

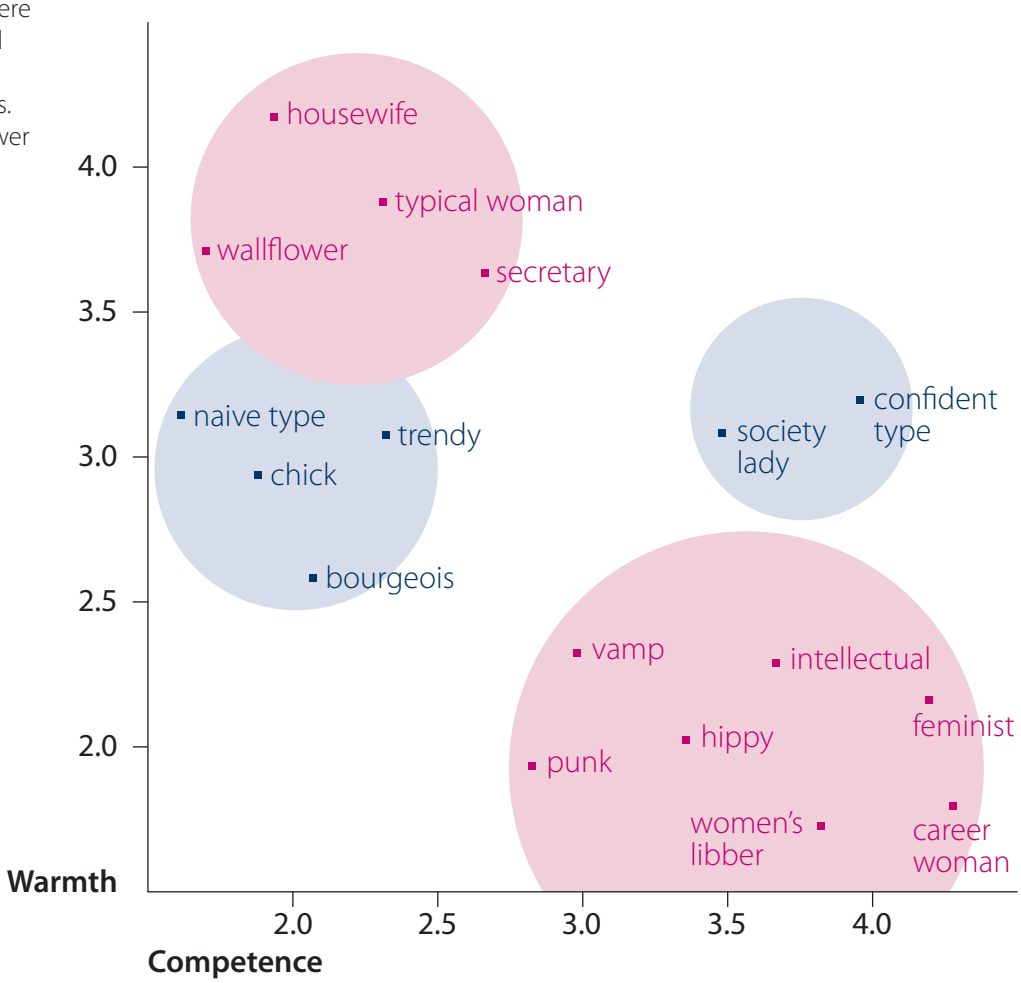
Stereotype subtypes

Perceptions of stereotype subtypes can also vary within groups. Research on gender stereotypes, for example, demonstrates that women can be viewed along a spectrum that varies from warm, nurturing and needing protection to hard and competitive 'Queen bees' provoking either benevolent or hostile forms of prejudice in turn (Glick and Fiske, 1996).

Benevolent prejudice still results in discrimination due to the restricted range of opportunities offered to groups that are evaluated as warm but low in competence. However, minorities can face a 'double bind' if they demonstrate behaviours not typically associated with their group, which can be responded to negatively. For example, studies of gender and leadership have found that not only are women considered less favourably than men for leadership positions, but if they do display the qualities considered important for leadership, these qualities are considered less favourably than when demonstrated by a man (Eagly and Karau, 2002).

Figure 3 represents the findings from a study by Eckes (2002) where students in Germany were asked to rate female sub-groups on competence and warmth ratings. The students were asked to answer how society typically views the groups, rather than how they viewed them themselves.

Figure 3



(Eckes, 2002: p104)

The link between bias and discrimination

According to social identity theory (Tajfel and Turner, 1979), people's sense of self is partly informed by differences in status between groups. They will try to sustain a positive in-group identity by achieving a distinctive and respected position for their in-groups. In his studies of boys at summer camps, Sherif (1966) showed that any two groups could be created and then turned into hostile enemies simply by making them compete. He demonstrated that if the only way for a group to gain is for the other to lose, then hostility, negative stereotypes and bias will follow. Sherif also showed that intergroup relations could be improved by setting goals where neither group could succeed without the other's help or contribution.

Realistic conflict theory

Boys between the ages of 11 and 12 thought they were attending a summer camp, though the camp staff were in fact the researchers. The boys were carefully selected to be similar in socioeconomic status and to have good physical and mental health. They were strangers to each other before they met. The boys were assigned to one of two groups and given time in each group for natural leaders to emerge and to engage in activities that built group cohesiveness. The two groups were then brought together and had to compete in various activities to gain prizes.

'The end result of the series of competitive contests and reciprocally frustrating encounters between the Eagles and Rattlers was that neither group wanted to have anything whatsoever to do with the other under any circumstances. On previous days, the now familiar invectives and names had been hurled back and forth ('stinkers', 'braggers', 'sissies' and many considerably worse), derogation of the out-group had been expressed in word and deed (eg holding noses when in their vicinity). Now both groups objected even to eating in the same mess hall at the same time.'

(Sherif, 1966: pp82–83)

However, groups do not have to be in competition for bias to arise. Simply putting people into 'minimal' groups (Tajfel, 1970) is enough.

Minimal group paradigm is a method for investigating the minimal conditions required for discrimination to occur between groups. Experiments using this approach have demonstrated that even arbitrary and virtually meaningless distinctions between groups (eg the colour of their shirts), and where the 'reward' is simply the allocation of points, can trigger a tendency to favour one's own group at the expense of others.

What this means is that bias is likely to be generated simply if, when comparing themselves with others, people see themselves as belonging to a social category ('in-group') rather than simply as individuals (Tajfel and Turner, 1979). The implications are that since we all belong to a variety of social categories (gender, age, race, class, role, interests and so on) each person has a complex set of in-groups and out-groups, the relevance of which are in constant flux. As a consequence, although we are all capable of being biased, a person's particular biases will arise out of a set of relationships unique to them. Bias therefore needs to be seen as a process rather than a characteristic of a particular person. Bias can be directed to a wide variety of groups, can change in intensity over time, and can be expressed in a variety of ways (Abrams, 2010).

Bias control

The extent to which people are motivated to control their prejudices and what the source of that motivation is differs. We have a bias control mechanism in the brain which operates automatically and which tries to prevent our biases becoming behaviour (Amodio *et al*, 2004). To trigger this mechanism our brain needs to see a mismatch between our wider goals (eg our desire to be or to be seen as fair, or not to get fired) and our instinctive biases.

In society today it is generally unacceptable to behave in a prejudiced manner and there are various sanctions against doing so. People may therefore be motivated to respond without prejudice in order to avoid social sanctions. In other words, they are externally motivated to respond without prejudice. It is also possible, however, to be motivated to respond without prejudice

because of personally important non-prejudiced standards (values and beliefs). In other words, to be internally motivated to respond without prejudice (Devine *et al*, 2002).

People whose motivation is self-determined (ie the internally motivated) more effectively control implicit and explicit prejudice across situations and strive for positive interracial interactions. In contrast, people who respond without prejudice to avoid social sanction (ie are primarily externally motivated) more consistently fail at regulating difficult to control implicit and explicit prejudice and respond with anxiety and avoidance in interracial interactions (Plant *et al*, 2010). Implicit attitudes are therefore malleable and subject to a range of influences ranging from self and social cues, the context, characteristics of the stereotype trigger, as well as individual differences in bias control (Blair, 2002).

The implicit association test (IAT)

The IAT is a metric often used within social psychology to measure the strength of a person's automatic association between mental representations of social groups (concepts) and positive or negative ideas (constructs).

The subsequent sections of this report examine studies exploring the existence of implicit bias, how implicit bias affects behaviour and decisions, and most importantly, how we can mitigate the impact of implicit bias. The studies included are those that measure implicit bias using an IAT and consequently it is important to understand the mechanisms of the IAT and the debate around its validity and limitations.

The IAT is not universally accepted as a valid and reliable measure of a person's implicit bias. However, 'at present, it is the dominant method for assessing implicit associations' (Rudman, 2008: p426).

See appendix 2 for the full background and debate around the IAT.

4 Impact of implicit bias on behaviour and actions

Although there have been hundreds of academic studies published around the IAT methodology, there are far fewer about how IAT scores may predict behaviours. Those that do exist are almost exclusively in university/laboratory settings using undergraduate students as subjects. This is the *modus operandi* of academic research but although such studies can be useful in explaining the processes, practitioners will want to see the evidence in real-world settings.

Although all academic papers have been reviewed regardless of the setting, most discussion is around those papers involving IAT that have something to say about how we might behave when assessing or evaluating others, particularly when that is in a staff-selection context. This section includes key empirical and meta-analysis studies, but excludes opinion or review papers where no new data was either gathered or analysed.

Recruitment and selection situations

Impact of workplace culture on recruitment decisions

Ziegert and Hanges' 2005 study recruited 103 undergraduate students (83 white and 20 from ethnic groups other than black) who had previously completed a range of tests (including the race-based IAT) and questionnaires as part of their college induction. Participants were asked to play the role of a manager and complete an in-tray exercise including the evaluation of a number of job applications. Participants were randomly assigned to either a 'climate for equality' or a 'climate for racial bias' group. This distinction was achieved by placing a memo from the fictitious president of the company in the in-tray bundle. In both groups, the memo instructed participants to take into account a job applicant's education and experience in making their evaluations of each candidate. However, for participants in the 'climate for racial bias' group, the president's memo also said:

> 'Given that the vast majority of our workforce is white, it is essential we put a white person in the VP position. I do not want to jeopardise the fine relationship we have with our people in the units. Betty (the outgoing vice president) worked long and hard to get those folks to trust us, and I do not want her replacement to have to overcome any personal barriers.'
>
> (Ziegert and Hanges, 2005: p558)

The researchers had predicted that participants in the climate for racial bias group would exhibit more discrimination than those in the climate for equality group, and that was the case. There were differences between the individual participants, but the organisational culture did have a significant impact on decisions:

'The relationship between discrimination and implicit racist attitudes was almost nonexistent [in the climate for equality] … the climate for racial bias manipulation sent a signal about social norms for discrimination that resulted in implicit racial attitudes being related to disparate ratings of black and white applications.'

(Ziegert and Hanges, 2005: p559)

'When individuals were given a business justification for racial discrimination their implicit racist attitudes were positively related to their discriminatory behaviour.'

(Ziegert and Hanges, 2005: p561).

Policy implications

While this study is not based in a real-world setting, it highlights the need for managers to promote an inclusive work environment.

These findings are not without their critics. The most obvious concern, not unusual in studies of this nature, is that the scenario is not realistic. The memo would not happen in the real world, not least because it would be unlawful. Blanton *et al* (2009) also argued that despite assertions that the job applications used were 'equivalent', the evidence suggested this was not the case as there were significant differences in the way two of the six applications had been scored (one black and one white application scored significantly better than the other two from that ethnic group).

Blanton *et al* also suggested that the inclusion of the IAT added only small amounts of prediction when compared with an assumption that everyone was moderately biased. Ziegert and Hanges (2009) responded and argued their case for the methodology used, and suggested Blanton *et al* (2009) had been selective in the evidence they had reported.

Shortlisting CVs (without the use of IAT)

In several studies, researchers have responded to real job advertisements by sending out identical or comparatively identical CVs in response to the job advertisement. This aims to place the research in real-world settings. The researchers then manipulate key information about the applicant to see what impact this has on outcomes. For example, they give the applicant a common white, Asian or black name, or change an applicant's photograph. They then wait for responses and compare the rate at which different social groups are shortlisted. The sifting process is therefore carried out by real recruiters for real jobs.

The following studies all adopt this technique to explore the impact of social characteristics on the success of job applicants, but without the use of an IAT to measure implicit bias.

Impact of ethnicity on shortlisting decisions

Wood *et al* (2009) conducted a study instigated and published by the Department for Work and Pensions in the UK.

Formally advertised job vacancies were identified in seven British cities for nine occupations: IT support, IT technician, accountant, accounts clerk, HR manager, teaching assistant, care assistant, sales assistant and office assistant. These roles were chosen to represent more and less technical skills levels.

A set of three applications was developed in response to these advertisements that were closely matched in terms of their education, skills and work history.

Ethnic identity was conveyed using names found to be widely associated with the ethnic groups included in the study (black African, black Caribbean, Chinese, Indian, Pakistani/Bangladeshi, white). These names were randomly assigned to each application (one of the three was white, with the other two from different minority ethnic groups).

Responses from employers were monitored, with the key positive outcome being a call-back for an interview. Discrimination was measured as differential treatment at an aggregate level between the ethnic groups in the study. The fact that applications were sent for the same vacancies provided the control.

'Of the 987 applications with a white name, 10.7 per cent received a positive response. This compared with 6.2 per cent of the 1974 applications with an ethnic minority name – a net difference of 4.6 percentage points. Put another way, 16 applications from ethnic minority applicants had to be sent for a successful outcome in our test compared with nine white. That is, 74 per cent more applications from ethnic minority candidates needed to be sent for the same level of success.'

(Wood *et al*, 2009: p11)

Impact of gender on early career researchers

Steinpreis *et al* (1999) sent out the CV of an early career researcher seeking employment to 238 academic psychologists in the USA. The gender of the applicant was manipulated (male/female) but the CVs were otherwise the same. The outcomes showed that both male and female academics were more likely to want to employ a male job applicant than a female job applicant with an identical record. Similarly, both sexes reported that the male job applicant had more adequate teaching, research and service experience compared with the female job applicant with an identical record.

The academics were also asked to examine an impressive CV from a real scientist with the same gender manipulation. In contrast to the less experienced CV, when male and female academics examined the highly competitive CV of the real scientist who had achieved early tenure, they were equally likely to employ the male and female candidates and there was no difference in their ratings of their teaching, research and service experience.

Policy implications

While this finding is based only on one study and in a specific field, it is interesting that gender inequality was more likely to occur in decisions relating to early career academics. This has implications for women beginning their careers, or at least their science careers, and also affects pipeline issues. It potentially also suggests that qualifications and experience may elevate some academics above the potential for discrimination.

Impact of gender on being shortlisted as a laboratory manager

Similar results were seen in Moss-Racusin *et al* (2012). A randomised double-blind study (n=127), within science faculties rated the application materials of a student who was randomly assigned either a male or female name for a position as a laboratory manager. Selectors rated the male applicant as significantly more competent and hireable than the female applicant. They also chose a higher starting salary and offered more career mentoring to the male applicant. The gender of the selector did not affect responses.

Policy implications

This study demonstrates that the gender of the shortlister made no difference to the decisions being made. While it is important to consider the diversity of the people conducting recruitment and selection processes to ensure a diversity of experiences and opinions, it is important to remember that they are all equally susceptible to biased behaviour.

Impact of ethnicity on skilled, semi-skilled and unskilled job applications

Carlsson and Rooth (2007) explored the impact of racial discrimination in employment in Sweden. Previous studies suggested that racial discrimination was worst against people from a Middle Eastern background, and that this group also had higher rates of unemployment than native Swedes, so they set out to investigate whether the two were connected.

In a nine month period between May 2005 and February 2006 they sent 3104 job applications to 1552 vacancies for skilled, semi-skilled and unskilled jobs in the Stockholm and Gothenburg areas of Sweden. Each application was created with equally matched experience and skills and then randomly assigned either a typically native Swedish name, or a Middle Eastern name.

In addition to sending the applications, they investigated the workplace of the vacancy to assess the size of the organisation, the structure, the diversity and turnover of the workforce, the proportion of immigrant employees and proportion of

immigrants living in the local area. They also used follow-up interviews with the employers to ascertain more information, including the gender and ethnicity of the person conducting the shortlisting.

Their aggregated results found that the interview call-back rate was 29 per cent with a Swedish name and only 20 per cent with a Middle Eastern name. Interestingly, the relative call-back for Swedish-sounding names was higher in lower skilled jobs than higher skilled jobs despite the fact that lower skilled occupations have a higher proportion of immigrant employees.

The study found that applicants with a Middle Eastern name were less likely to be called back for interview where:

- a man, rather than a woman was conducting the shortlisting
- the organisation had fewer than 20 employees
- the organisation was located in an area with a relatively high number of immigrants

Furthermore, where the organisation had a relatively high turnover of staff, those with a Middle Eastern name were slightly more likely to be called back than those with a Swedish name. The other characteristics did not identify any statistically significant trends.

The researchers explain these results to some extent by suggesting that larger organisations and those with higher levels of staff turnover may have robust recruitment processes in place which reduce the possibility of bias. Furthermore, larger organisations may invite more people for interview overall.

Policy implications

All of the CV studies demonstrate the potential benefits of using anonymous shortlisting wherever possible. Discrimination can still occur at the interview stage, but at least underrepresented groups will have a more equal chance of reaching it. The research also suggests that having robust, systematic recruitment processes may help to reduce discrimination.

Impact of foreign qualifications on shortlisting decisions

Carlsson and Rooth (2008) took their experiment a stage further to include the origin of an applicant's qualifications, as well as the impact of the origin of their name.

They created three types of job application for 11 different occupations, each with comparable experience, skills and qualifications. The first was assigned a Swedish name, the second a Middle Eastern name, and the third was assigned a Middle Eastern name, and their qualifications were changed to having been obtained in the Middle East.

The researchers controlled for age and ensured the Middle Eastern applications had information to assure the shortlister that they were proficient in Swedish, and also had equivalent work experience to the applicants with the Swedish name.

They encountered a number of issues with the analysis of the data, but generally found that the difference in call-back rate for applicants with a Swedish name against those with a Middle Eastern name was 23 per cent explained by foreign qualifications, and 77 per cent explained by their name.

Policy implications

This study highlights the need for institutions to consider their own understanding of and potential bias towards qualifications obtained overseas. This may impact on international staff applying for vacancies in UK higher education, as well as overseas students applying for undergraduate and postgraduate courses.

In addition, while this study demonstrates the impact of qualifications obtained overseas, it is possible the same effect might occur based on the UK institution the applicant attended, or subjects and subject routes that are equivalent to traditional routes of entry to higher education.

Is the discrimination implicit or explicit?

Rooth (2010) went a stage further and attempted to determine whether the discrimination identified in the previous two studies was explicit or implicit. In this study, researchers contacted all of the recruiters who had been involved in the 2007 and 2008 studies to involve them in this third project.

They located the firm's recruiter and more specifically, the individual responsible for the shortlisting in the two previous studies. They did not say that they had submitted fake applications but that they had been interested in following those particular recruitment campaigns. They then invited the recruiters to take part in their study looking at general recruitment practices.

Approximately 50 per cent of the recruiters were not located, or stated that the recruitment decision was not theirs, or not theirs exclusively, or were not interested in participating. The recruiters undertook an IAT focused on attitudes to Arab-Muslim men and an explicit attitude questionnaire. In the end only 193 employers, or 26 per cent of those in the first experiment and 158 or 19 per cent of employers in the second experiment completed the IAT and questionnaire.

The researchers concluded:

'We find strong and consistent negative correlations between the IAT score and the probability that the firm/recruiter invited the applicant with an Arab-Muslim-sounding name for an interview. The results are identical in the two experiments and show that the probability to invite job applicants with names such as Mohammed or Ali decreases by five percentage points when the recruiter has at least a moderate negative implicit association toward Arab-Muslim men in Sweden.'

(Rooth, 2010: p529)

Policy implications

The researchers concluded that the study demonstrates that implicit bias can have an impact on real-life situations and on recruitment and selection decisions.

Hiring preferences with an IAT

Agerström *et al* (2007) used an IAT in a recruitment situation. Agerström developed two IATs that measured preferences towards native Swedish men v Arab-Muslim men.

Like Rooth (2010), Agerström *et al* (2007) added to the IAT a number of explicit measures:

- a feeling thermometer asked the participants to rate their positive or negative feelings on a ten-point scale (1=very negative feelings, 10=very positive feelings) toward Arab-Muslim and Swedish men
- a hiring-preference task that asked the participants to choose which groups they prefer when hiring people (as Nosek *et al*, 2005)

The first experiment carried out by Agerström *et al* (2007) was with 87 undergraduate students in Sweden who more quickly associated native Swedish men with 'good words' than Arab-Muslim men. The mean difference was 155 milliseconds (showing that participants more easily associate Arab-Muslim men with low-performing attributes and Swedish men with high-performing attributes, than they associate Arab-Muslim men with high-performing attributes and Swedish men with low-performing attributes).

The hiring-preference task showed a significant correlation with the Arab-Muslim stereotype IAT, but the IATs did not correlate significantly with the other explicit measures.

Agerström *et al* (2007) repeated the experiment with 158 employers (80 men, 78 women) responsible for the recruitment of staff at various companies based in Stockholm and Gothenburg, Sweden. The employers all had more than ten employees and were randomly selected from job advertisements.

Again the employers completed the IATs and explicit measures, and the same patterns were identified but with a much larger difference: the employers matched 'Muslim men' with 'negative' and 'Swedish men' with 'positive' quicker than the reverse, with a mean difference of 350 milliseconds.

Overall, 148 (94 per cent) of the employers showed at least a slight preference towards Swedish men, while only two employers showed the reverse pattern. In the hiring-preference rating task, a total of 52 employers (33 per cent) explicitly stated that they

slightly or strongly prefer Swedish men to Arab-Muslim men when hiring staff. The Arab-Muslim attitude IAT showed a significant correlation with the feeling thermometer but no significant correlation was found with the hiring-preference rating task.

Implicit bias, body language and interactions with others

One of the key early papers, and one often cited in the discussion on IAT prediction of behaviour, is McConnell and Leibold (2001), involving 41 white undergraduate students:

'Participants arrived at the laboratory for an experiment on 'word perception' and were greeted by a white female experimenter. Unbeknown to the participant, a hidden video camera was positioned to record the participants' and experimenters' full bodies and their entire range of movements during scripted social interactions. A hidden unidirectional microphone recorded their discussions. They were directed to a rolling desk chair initially positioned 120 cm away from the experimenter's chair, allowing participants to establish a preferred distance from the experimenter.'
(McConnell and Leibold, 2001: p437)

The experimenter engaged them in polite scripted conversation about the experiment including a joke. After the completion of a 'confidential' questionnaire and the recording of experimenter notes on their interaction, the white experimenter led them to the IAT test, indicating that she was leaving shortly. In the ten minutes the participants took to complete the IAT, a black female experimenter replaced the white experimenter. Once again, the participant was directed to a chair positioned 120 cm from the experimenter's chair, allowing the participant to establish a preferred seating distance. Once again the experimenter engaged them in polite scripted conversation about the experiment including a joke at the same point as the white experimenter.

Trained judges who were unaware of participants' explicit attributes then coded 16 behaviours seen in participants using a scale from 1 (none) to 9 (very much). They looked for:

'friendliness during the interaction, the abruptness or curtness of the participant's responses to questions, the participant's general comfort level, how much the participant laughed at the experimenter's joke, and the amount of participant's eye contact with the experimenter.

> On 5-point scales, they assessed the participant's forward body lean toward the experimenter (vs. leaning away), the extent to which the participant's body faced the experimenter (vs. facing away), the openness of the participant's arms (vs. crossed arms), and the expressiveness of the participant's arms (vs. not moving at all). Judges also calculated the distance between the experimenter and the participant's chair at the end of the interaction to gauge social distance. Judges also recorded the participant's speaking time, number of smiles, number of speech errors, number of speech hesitations (eg 'um'), number of fidgeting body movements (eg swinging feet and shifting positions), and the number of extemporaneous social comments made by the participant.'
>
> (McConnell and Leibold, 2001: p438)

Along with the IAT scores and various explicit measures of bias, McConnell and Leibold created a score of the difference between the judges' behavioural ratings between the white and the black experimenter. The IAT scores correlated with the explicit measures and inversely with experimenter ratings, but importantly also inversely with the judges' ratings: higher IAT scores (ie where participants demonstrated less favourable attitudes to the black experimenter) correlated with less positive behaviours.

Blanton *et al* (2009) re-examined the methodology, analysis and conclusions of the study and considered that McConnell and Leibold had wrongly handled outliers (ie data that is very different from other data collected) where test takers returned extreme scores. McConnell and Leibold (2001) had used the original Harvard IAT algorithm (see appendix 2 for more information on the IAT algorithm), which did not take into account the effects of age and processing speed. Blanton *et al* argued that an outlier had been included in the analysis that should have been excluded. When the outlier was excluded the correlation between IAT scores and behaviour became statistically significant.

The Blanton *et al* re-analysis revealed a pattern of behaviour consistent with a pro-black behavioural bias, rather than the anti-black bias suggested in the original study. 90 per cent (37 of 41) of test takers in the McConnell and Leibold (2001) data had positive IAT scores, suggesting an anti-black bias. Blanton *et al*

argued that if 90 per cent of respondents were biased against black people and the IAT metric is not arbitrary (see Blanton and Jaccard, 2006), one would expect a large proportion of the sample to discriminate against the black experimenter. They felt this was not the case and in their analysis over 70 per cent of the sample acted more positively toward the black experimenter. In the untransformed data, it appeared that those with higher IAT scores were the least behaviourally biased in the sample.

Additionally:

- they noted that the way the measures were sequenced may have contaminated the study: having sat through an explicit questionnaire with the white experimenter, Blanton *et al* argued that it is unlikely test takers were oblivious to the subject matter the researchers were investigating
- they argued that appearing racist in the IAT may have subsequently affected interactions with the black experimenter
- they had concerns over the inter-judge reliability

McConnell and Leibold (2009) responded, putting a counter view to the critique including a rationale for including the outlier, and asserted that the judges were extensively trained and had spent four months painstakingly coding the videos.

They regard the point of the 2001 study not as whether one set of ratings (eg those for black experimenters) is greater than another set (eg those for white experimenters), but the relative difference between these sets (ie how friendly participants were toward members of one race relative to another) that is predicted by the IAT, which itself is a relativistic measure.

Policy implications

The conclusions of this study are evidently open for debate. However one consistent finding is that the white and the black experimenter were treated differently by the research participants, it is just unclear how they were treated differently and why.

While more research is needed in this area, it is worth considering the impact of body language in an interview situation. It is possible that interviewers may implicitly encourage or discourage participants, and the level of encouragement may be impacted by the biased opinions held by the recruiter.

Likewise, it is possible that the body language of the employee may be affected by the characteristics of the interview panel. This is further evidenced in the study below relating to IAT and the law.

Impact of bias in other situations and sectors

IAT in medicine and clinical psychology

Green *et al* (2007) used IAT data from 220 white medical doctors along with measures of explicit bias around prescribing an effective drug to patients from different ethnic groups. Doctors were presented with short case studies of symptoms and asked to make prescriptions on drug use. They found that although measures of explicit bias did not predict prescribing practices, the IAT did predict discriminatory prescribing – doctors were more likely to prescribe effective drugs to white rather than black patients.

Implicit bias in body language

Shelton *et al* (2005) explored the relationship between implicit racial bias and body language in interactions with someone from a different ethnic group. They recruited 29 black students, and 67 white students from Princeton University, and matched them with someone of the same sex, creating 29 pairings of a white with a black student, and 19 pairings with two white students. The students were asked to take an IAT and were then told they were taking part in research for a graduate student's dissertation on first impressions. They were led to a room and sat opposite their pair and were asked to pick a conversation topic from a bowl, without realising that all of the conversation topics were the same on race relations.

Participants then rated their partners on a variety of areas including their likelihood to be friends with the person afterwards and their engagement in the interaction. The researchers found that generally black participants rated white participants with higher levels of racial bias (according to their IAT) more favourably

than white participants with lower levels of racial bias. They acknowledge various limitations with the research, but some of their suggested explanations for the results included:

- those with higher levels of implicit racial bias may have tried harder to regulate their behaviour and were therefore making a lot of effort to seem friendly and engaged with black partners. In contrast those with lower levels of bias may not have felt the need to over-compensate and therefore seemingly made less effort to interact
- those with low implicit racial bias may actually have been trying so hard to regulate their behaviour that they may have appeared less engaged, or may have 'choked' under the pressure
- those with higher racial bias may have given the appearance of listening and engaging, but may not have been doing so in a meaningful way

Richeson and Shelton (2005) recruited 30 black and 30 white participants to explore their ability to recognise racial bias in people's body language. The participants were shown a 20 second silent video of white people interacting with an unseen other and were asked to assess their body language. The 'other' person in the clip was either a black or a white individual and the interaction was either about something general, or about something race specific.

The researchers found that the black participants were able to identify biased body language in contextually relevant interactions (ie where the people in the video were discussing race-specific issues). They hypothesised that this might be because black participants had more awareness of biased behaviour as they are from a minority ethnic group.

Implicit bias in social policy

IAT has also been shown to predict attitude towards wider social policy. Pérez (2010) used a representative web survey of 350 adults, examining implicit attitude towards Latino immigrants. Pérez found that implicit attitudes significantly influence individual preferences for illegal and legal immigration policy. This was regardless of political ideology, socioeconomic concerns, and measures of intolerance toward immigrants.

5 Methods and techniques for reducing implicit bias

Research into techniques for reducing implicit bias can be categorised along the following lines:

- discounting commonly held stereotypes
- using context to influence implicit responses
- changing the way an out-group member is evaluated and categorised
- using contact to change the level of threat evoked by an out-group
- using motivation to change responses to an out-group
- encouraging people to take responsibility for their implicit biases

This section provides an exploration of the key studies for each of these approaches.

Most of the research on interventions has been carried out in the laboratory. This has its advantages in that it is possible to isolate a specific factor to be studied, but it raises issues of whether the intervention can transfer to the real world.

Discounting commonly held stereotypes

This aims to undermine the categorisation process by challenging and preventing commonly held ideas and stereotypes about particular groups.

Stereotype disconfirmation using positive examples

Stereotypes consist of a set of attributes that are assumed to be characteristic of a particular group. The stereotype disconfirmation method attempts to undermine this stereotype by providing exposure to positive examples (role models) that exhibit characteristics that run counter to the stereotype.

For example, women are stereotypically not seen as leaders and managers. Counter-stereotypic female examples would be women who demonstrate characteristics of being strong leaders and who hold significant positions. In organisations in which a group is underrepresented in a particular role or level of the organisation, having positive exposure and contact with the minority example will have an impact on levels of implicit bias.

Positive images of black people and negative images of white people

Dasgupta and Greenwald (2001) studied the effects on implicit and explicit bias of being presented with famous black examples (eg actor Denzel Washington) and infamous white examples (eg Hitler), or famous white examples (eg actor Tom Hanks) and infamous black examples (eg boxer Mike Tyson). The control group were given pictures of flowers and insects.

In all three trials, participants were familiarised with the examples through the use of a general knowledge quiz in which underneath the picture of the example were two pieces of information, one correct and one incorrect. Participants had to indicate which piece of information was correct. Pro-white implicit bias was significantly reduced immediately after exposure to positive black examples. This effect persisted when measured again after 24 hours.

The 'Obama effect'

President Obama is one of the most famous positive black role models and Plant *et al* (2009) set out to explore his impact. His positive media exposure and subsequent election may have had a nationwide impact in the USA and was thought to be responsible for the significant drop in implicit bias in 2008 among student populations whose implicit and explicit attitudes had been regularly sampled since the late 1990s (Greenwald *et al*, 1998). As recently as 2006, the majority of white people in the USA tested (75–85 per cent) were found to usually show anti-black bias.

The researchers explain:

'In stark contrast, we failed to observe anti-black bias among 74 white participants' average scores on either the evaluative IAT … or the stereotyping IAT … Further, 45 per cent of participants showed negative D-scores (indicating pro-black bias) on the evaluative race IAT compared with previous work that found only 7 per cent of white participants responded with negative D-scores in a sample of over 85,000 (Banaji, 2005).'

(Plant *et al*, 2009: p961)

Their research suggests that the possible reason for this drop in bias could be:

'that participants had positive black examples come to mind or anticipated that other people have positive examples come to mind when they thought of black people and this was associated with low levels of racial prejudice.'

(Plant *et al*, 2009: p961)

Seeing women as leaders in higher education

Gender stereotypes are beliefs about differences between women and men. These beliefs are shared, by men and women alike, and are present in all cultures that have been studied (Desert and Leyens, 2006). Traits such as being competitive and assertive are associated with men and collaborative traits such as being emotional, supportive and concerned about others are associated with women.

In a rare combination of laboratory and natural field experiment, Dasgupta and Asgari (2004) convincingly demonstrated that exposure to female leaders reduced women's automatic gender stereotypic beliefs. Previous research suggested that reading biographies about famous and high-achieving women temporarily reduces women's gender stereotypic bias. Dasgupta and Asgari extended those findings in two contrasting, naturally occurring environments: a women's college and a mixed college.

There are significantly more women in leadership positions and science and mathematics faculties in women's colleges. In 1993, 45.5 per cent of mathematics and science faculty staff in US women's colleges were female, compared with 11.4 per cent within mixed colleges (Sebrechts, 1993). Using matched samples or after controlling for potential variables (eg social class, SAT scores and geographical region), studies have found that:

= compared with female students at mixed colleges, students at women's colleges are more likely to change majors from female-dominated disciplines to neutral or male-dominated disciplines (Solnick, 1995)

= women's college graduates are more likely to attain high-end professional positions and incomes three to ten years after graduation (Riordan, 1990, 1994)

= women's college graduates are more likely to choose professions in which women are underrepresented, such as medicine (Tidball, 1985), and other sciences (Tidball and Kistiakowsky, 1976) compared with female graduates of coeducational institutions

In the study by Dasgupta and Asgari (2004), two groups of female students were followed, one in each environment, where they were exposed to the women within their college, for example, female deans, professors and science lecturers.

After a term at college, women who attended a mixed college had greater implicit gender bias than when they started college. In contrast, women who attended an all women's college had less implicit gender bias than when they started.

To rule out the concern that women who chose a same-sex college would already differ on gender attitudes, the researchers measured attitudes of both groups a few months into their first year and again in their second year.

'All participants completed an implicit association test … to assess the extent to which they automatically associated women with leadership qualities relative to supportive qualities (abbreviated as the gender-IAT).'

(Dasgupta and Asgari, 2004: p646)

After a year, women in the single-sex college expressed no gender stereotypes, while those at the mixed college expressed strong gender stereotypes.

Results showed that the amount of exposure to female faculty members significantly predicted implicit gender beliefs in year two of college. In other words, the more female students were taught by female faculty members the more they implicitly associated women with leadership qualities.

The other factor that significantly affected gender beliefs was the proportion of mathematics and science courses taken. In the women's college this would have resulted in a higher level of exposure to counter-stereotypic female examples in comparison to the mixed college due to the significantly higher

level of female faculty members in these subjects in the women's college. In the mixed college if a female student did more science and mathematics courses then they would end up with less exposure to counter-stereotypic female examples because of the underrepresentation of women in these roles.

Policy implications

These studies highlight the importance for staff and students to see people in roles that they would not stereotypically be seen in. Ten years ago people may have never paired the US president with a black man. It is important that people are able to pair, for example, BME people with senior roles and in disciplines in which they are underrepresented in the same way they might pair white with vice-chancellor.

Institutions should be mindful of:

- opportunities for developing the pipeline and ensuring a more diverse workforce is able to fill middle and senior manager roles
- the diversity of guest lecturers they invite to their institution
- the diversity of keynote speakers at conferences and seminars
- the diversity of imaging generally around the learning/working environment, for example, posters, paintings, room and building names, marketing materials and the website

Using repeated pairings to reverse commonly held views

Evaluative conditioning aims to produce change by using associative conditioning (repeated pairing of two stimuli together) of which the participants are not consciously aware. Since people develop negative racial attitudes by being exposed to repeated pairings of images of different social groups and negative events, for example, black people and crime, the researchers in this category argue that the reverse ought to be true (Olson and Fazio, 2001, 2002).

Our literature review found three studies of this type. Included under this category are studies that found that surreptitiously inducing participant's to smile by clenching a pencil in their mouths (Ito *et al*, 2006), moving a joy stick towards a picture of the out-group member (Kawakami *et al*, 2007), and seeing pleasant

objects (ice cream, a baby) (Olson and Fazio, 2006) while viewing out-group members, all resulted in a reduction of implicit bias.

In the Olson and Fazio (2006) study participants were neither instructed nor motivated to categorise black people as members of a specific category. In fact, the interest in race was obscured. Participants were told that the experiment was about 'attention and surveillance' and that they would see a stream of randomly assembled images on the computer screen. Their task was to press a button whenever a pre-specified item appeared. Throughout that task, black with good images and white with bad images were interspersed in a random manner, among a variety of other images, a total of 24 times each.

The fact that it only took 24 exposures to achieve the reduction in implicit bias is a key benefit of this approach. In comparison to many bias-reduction interventions that require considerable energy on behalf of participants, change occurs though a relatively effortless learning mechanism (Olson and Fazio, 2006).

Policy implications

The relative ease of creating positive associations that reduce implicit bias should alert us to the ease with which the opposite can occur in the real world. If people are constantly exposed to pairings of minority groups with negative attitudes and events then negatively implicit attitudes are surely being established.

Many organisations find that the ethnic composition of, for example their cleaning staff, is very different to that of the senior management team. If people are making those associations on a daily basis then they are likely to start to associate some ethnic groups with cleaning, and other ethnic groups with senior management. If this then affects their decision-making processes in recruitment situations and how they view different candidates, it can exacerbate this trend.

These studies highlight the importance of presenting on a regular and repeated basis positive images, stories and news items about minority groups in institutions as a strategy for minimising the impact of implicit bias.

Reviewing the media critically

In Ramasubramanian (2007), randomly assigned subjects either watched a video that encouraged them to question media coverage and be critical of what was presented, or a control video on journalistic writing. Participants then either read a stereotypic or a counter-stereotypic news story about African Americans or Asian Indians.

Receiving the first video or receiving the counter-stereotypic news story was not enough individually to reduce implicit bias, but implicit attitudes were reduced for those participants that received both methods. Interestingly the effect was only significant for the African-American news story. One reason for this could be that the counter-stereotypic example in the Asian trials – a successful Asian businessman – is of itself another stereotype, albeit a benevolent one, and therefore may not have been stereotype disconfirming.

Reading an essay on the positives of Arab-Muslim culture

In Park *et al* (2007), one third of research participants were exposed to positive social information about Arab Americans through participants being asked to:

> 'read an essay on multiculturalism in which Arab-Muslim culture was introduced as an example (positive information condition). In this text, many positive aspects of Arab-Muslim culture were mentioned (eg abstinence, sharing wealth with the poor, strong neighbourhood, and contribution to the world in mathematics, physics, and architecture, etc).'
>
> (Park *et al*, 2007: p41)

As a contrast, one third of the participants were exposed to an article on terrorism and one third of the control group to an article on health. Only in the positive social information condition was there a reduction in implicit bias against Arab Americans.

Creating an affiliation to an out-group

Three of the four studies in this category used tasks that involved the participants responding with the word 'yes' whenever the out-group prime (this can either be a picture or a word) was matched with a stereotype-incongruent word as another method of evaluative conditioning.

> 'The term in-group means a group to which someone belongs, and the term out-group means a group to which someone does not belong. This difference in affiliation has profound and robust effects on people's evaluations of members of the different groups. In-groups appear to have an inherent, and automatic, positivity associated with them, whereas out-groups have an inherent negativity. In other words, people appear to think of their own group in positive terms and of the other group in negative terms, at even preconscious levels (ie without even realizing it).'
>
> (Crisp and Nicel, 2004: p247)

The purpose of the following three studies was to see whether these evaluations can be altered. Each study used a different method:

- matching female faces with non-stereotypic words like 'strong' or 'powerful' (Kawakami *et al*, 2005)
- getting participants to respond by pressing 'yes' every time they saw a face and word combination that was incongruent with the prevailing gender and racial stereotypes of the out-group (Gawronski *et al*, 2008)

Participants could be required to respond to the correct pairing as many as 50 times.

Linking 'yes' to counter-stereotypical face and word combinations

In Gawronski *et al* (2008), research participants were informed that:

'The following task is concerned with the cultural stereotype of men and women. As you probably know, men are often considered as strong whereas women are often considered as weak. This, however, is a cultural stereotype that may or may not be true. In the following task, you will be presented with male and female names. In addition, you will be presented with words relating to strongness [sic] and weakness that will appear on the screen shortly after the names. Your task is to respond 'YES!' each time you see a combination that is INCONSISTENT with the cultural stereotype of men and women. Specifically, you are asked to respond 'YES!' with the space bar each time you see a FEMALE name and a word relating to 'STRONGNESS' or a MALE name and a word relating to 'WEAKNESS.' Please attend particularly to combinations that are INCONSISTENT with the cultural stereotype of men and women! For combinations that are consistent with the cultural stereotype of men and women, you do not have to do anything. Again, please respond 'YES!' with the space bar each time you see a combination that is INCONSISTENT with the cultural stereotype of men and women. Please try to respond as quickly as possible!'

(Gawronski *et al*, 2008: p3)

Reduce the positive or reduce the negative?

Two of the studies also compared the efficacy of responding 'no' when the in-group prime was matched with a stereotype-congruent word, but the results were inconsistent.

The Crisp and Nicel (2004) study found more reduction of implicit bias in the in-group paired with a negative trait condition. Here the aim is to reduce the automatic association of the in-group with positivity. This was found to be more effective in reducing implicit in-group preferences than the trials that trained participants to associate the out-group with positivity.

The Gawronski *et al* (2008) study found that the in-group negation condition actually strengthened implicit bias. However, this was a different technique in that participants were simply pressing the NO key when they saw stereotype-congruent pairings. Crisp and Nicel (2004) argue 'relative primacy of negative information in

person perception is a well-established phenomenon' (p252). It is possible that simply saying 'no' was not negative enough to have the impact on reducing bias in favour of the in-group that the negative pairings of the Crisp and Nicel study had.

Impact of counter-stereotypic conditioning

In the Kawakami *et al* (2005) study mentioned above, where female faces were matched with non-stereotypic words like 'strong' or 'powerful, participants had reduced levels of implicit bias following the task. However, when participants subsequently carried out a CV selection task there was no difference in the level of discrimination against female candidates in comparison to the control. Unless, that is, they carried out a filler task. Filler tasks – in this case doing a number of arithmetical sums – are designed to distract from conscious awareness of the purpose of the subsequent task. In a third trial, a task designed to minimise conscious control during the CV selection reversed the usual pattern of bias with more women being chosen.

Potential damage of 'heavy-handed' interventions

These results suggest that where participants believe that they have been unduly influenced they will moderate their response in a direction opposite to that of the influence (Wegener and Petty, 1997; Wilson and Brekke, 1994). As the training technique to reduce bias used by Kawakami *et al* was obvious, participants resisted this influence and strategically modified their responses unless they were distracted by either the filler task or the conscious control distraction task. These results suggest that 'heavy-handed' attempts to change bias could have unintended negative consequences and that participants may actively resist such attempts.

Policy implications

As mentioned above, difficulties arise from actively promoting images and tasks designed to develop negative perceptions of an in-group, and even promoting positive imagery of an out-group needs to be done sensitively and not in a heavy handed way. However, it is important to consider the implications of positive in-group bias, especially when recruiting, developing and assessing staff and students.

It is not enough to remove negative bias – unfair positive bias is also an issue. Initiatives explored in the rest of this section may be of help. In addition, institutions could make use of the positive action provisions within the Equality Act 2010:

'It is not unlawful to recruit or promote a candidate from an underrepresented group who is of equal merit to another candidate, if the employer reasonably thinks the candidate has a protected characteristic that is underrepresented in the workplace; or that people with that characteristic suffer a disadvantage connected to that characteristic.'

Using context to influence implicit responses

Stewart *et al* (2010) trained participants to select situational (luck, the rest of the team, etc) rather than dispositional explanations (behavioural, character or cultural flaws) for stereotype-confirming black behaviour, for example lateness explained by transport problems rather than laziness. One of the ways that out-groups are perceived differently to in-group members is the attribution made about the reasons for a person's behaviour.

The ultimate attribution error

Stewart *et al* looked to explore ways to reduce the ultimate attribution error (UAE) (Pettigrew, 1979). The UAE suggests that where individuals demonstrate behaviours, or find themselves in situations that are linked to a negative stereotype about them (eg women being bad at mathetmatics), they are judged by others using dispositional explanations (eg women being innately, genetically, bad at mathematics) rather than situational explanations (eg women culturally being discouraged from pursuing mathematics and/or more encouraged to pursue 'female' subjects).

The researchers recruited 72 white undergraduates and split them into a training group and a control group. The training cohort were asked to assign situational (rather than dispositional) explanations to behaviours pertaining to negative stereotypes about black men (eg being lazy and promiscuous). One example was a picture of a black man with a line reading 'arrived at work an hour late'. The participants then had to choose either 'The power went out and reset his alarm' or 'He is a particularly irresponsible person' as the explanation.

They found those trained to look for situational reasons showed reduced levels of stereotyping and implicit bias towards negative stereotypes of black men in the subsequent tasks in the research, compare to the control group.

Policy implications

In real organisational settings, training people to understand the UAE and how to counteract this form of bias could be a highly relevant and practical method of reducing implicit bias.

Framing the reasons for a candidate either being offered a job or not within a situational context, rather than justifying it through their personal traits and attributes may be a useful method for reducing bias.

Unfortunately, this study did not measure how long the effect persisted following the training and whether it had any impact on selection decisions, but that is something that institutions could evaluate themselves if trialling such training.

Impact of context and social role on bias

A possible explanation for the mechanism by which positive examples influence implicit bias is that they create a context in which the meaning of race shifts. For example, the 'good' black police chief in US TV shows would be expected to elicit less anti-black bias than the black gangster. In three studies that examined the impact of social role as a contextual construct on differential implications of race (Barden *et al*, 2004), levels of implicit bias varied depending on the type of social role the minority prime was in: for example, athlete/student; prisoner/church-goer/factory worker; prisoner/lawyer. To take one example, placing a black face in a ghetto background (implying gangster) would produce anti-black bias. If this same face is placed in front of a picture of a church (implying church member) then no bias would be activated. All three studies produced significant reversals of racial bias as a function of implied role.

Alternative explanation

However, an alternative explanation for these results is not that the church, athlete and factory worker settings represented positive contexts, but that they simply activated racial stereotype subtypes (Devine and Baker, 1991). According to this theory there is not just one stereotype about a category, but a number of stereotypes. For example, wise and fair black police chief, caring black nurse, etc. Some of these subtypes will activate negative bias and some will not. Rather than the changed context reducing bias, all that may have been achieved in the studies in this category was the replacement of one stereotype for another, albeit ones that did not provoke an aversive reaction.

Bias inhibition or bias control?

Allen *et al* (2010) set out to explore whether positive context acts to inhibit automatic biases or to activate bias-suppression mechanisms. In this study, participants completed a black-white IAT with primes presented in positive or negative contexts similar to the studies above, along with a measure of motivation to control prejudice.

Although participants showed less implicit bias in positive v negative contexts, this effect was stronger among participants who were more highly motivated to control prejudice. These results suggest that the reduction of implicit bias in positive contexts was due to successful inhibition of biases rather than reduced activation of biases.

Contexts that prime loyalty v equality

In the laboratory, a common method for indirectly evaluating how comfortable participants feel about a future discussion with an out-group member is to ask them to arrange two chairs while they are waiting for their discussion partner to arrive. In reality, the discussion is a fiction and the experimenter always lets the participant know after a few minutes have passed that their discussion partner has had to 'cancel' at the last minute. How closely participants arrange the two chairs is an indication of how comfortable they feel about the forthcoming interaction.

Priming for equality rather than priming for loyalty

This study focused on three in-group/out-group differentiations: national, local and religious. Two measures of implicit attitudes were used: the IAT (Greenwald *et al*, 1998), and the go/no-go association task (Nosek and Banaji, 2001). Intergroup goals were primed by asking participants to unscramble sentences (Srull and Wyer, 1979) in one of three different conditions (equality, loyalty and control). An example of a loyalty sentence is 'M helps members of his team'. An example of an equality sentence is 'C fights for equality between people'.

In each of the different measures of intergroup bias (IAT, go/no-go and seating distance), a significant linear trend was identified, indicating that the level of intergroup bias expressed after equality activation was minimal, compared with an intermediate level in the control condition and a high level after activation of loyalty. Interestingly, the loyalty sentences made reference to supporting and helping members of the in-group, such as helping one's team.

Policy implications

This study raises the question of whether typical interview questions that ask about being a 'team player' may be priming 'loyalty', thereby increasing the possibility of the activation of implicit biases. Institutions should consider the questions that they ask candidates in application forms and in interview situations and consider how the wording and phrasing could be interpreted.

Changing the way an out-group member is evaluated and categorised

Reducing differentiation

The next two studies aim to reduce implicit bias by using interventions that reducing the perception of differences between the in-group and the out-group. Using a task that had been successfully used to reduce explicit prejudice (asking participants to list characteristics that are shared between the in-group and the out-group), Hall *et al* (2009) found a significant reduction in implicit bias in comparison to a control group.

However, in a second trial where they made group identity noticeable or important, they failed to find an effect. This mirrors research on explicit prejudice in which attempts to blur intergroup boundaries can lead to an increase in explicit prejudice for those that are highly committed to their in-group identity. For example, people who are highly committed to a Scottish identity could resist being subsumed under an over-arching British identity (Crisp and Beck, 2005). Results such as these suggest that careful consideration needs to be exercised when the bias-reduction strategy of 'the melting pot' is used.

Individualise

An alternative de-categorisation strategy is one in which one attempts to 'see everyone as people'. People see other-race faces as more similar than their own-race faces, a bias termed the other-race effect (Meissner and Brigham, 2001). In a study with a strong design but with a small sample, 20 white participants were trained to individualise black faces. Following training, as white participants become better able to individualise black faces, their implicit racial biases were correspondingly reduced (Lebrecht *et al*, 2009).

Viewing others according to their personal rather than stereotypic characteristics was also one of the methods used in the successful bias-reduction training programme outlined below (Devine *et al*, 2012).

Using contact to change the level of threat evoked by an out-group

The study of the impact of contact (the quantity and quality of the relationships between in-group and out-group members) has a long heritage in social psychology. Allport's (1954) 'contact hypothesis' states that contact with out-group members will lead to more positive attitudes about the out-group when people:

- have equal status
- have common goals
- are in a cooperative or interdependent setting
- have support from authorities

Contact that meets these criteria will improve intergroup attitudes more than contact that does not (Pettigrew, 1998). In our review we were only able to locate four studies that examined the impact of contact on implicit bias as opposed to explicit bias.

Impact of interracial friendship on implicit bias

Two of the studies in this category explored the impact of interracial friendships. Aberson (2004) asked participants to complete a self-report measure of their friendships with minority groups and completed an IAT. They found that participants with close friends who were members of the target group (black or Latino) exhibited less implicit prejudice than participants without close friends from the target group.

Shook and Fazio (2008) took advantage of the random allocation of dormitory rooms to new college students, to use a real-life situation in the USA. A computer randomly assigned white students to either a white or a black roommate. Owing to a shortage of accommodation on campus, students were unlikely to be able to change their accommodation in their first term. Students were tested for implicit and explicit bias, and completed self-report measures of satisfaction with their roommate and levels of intergroup anxiety in their first two and last two weeks of their first quarter on campus.

Although participants in interracial rooms reported less satisfaction and less involvement with their roommates than participants in same-race rooms did, their automatically activated racial attitudes and intergroup anxiety improved over time. This was not the case among students in same-race rooms. So even where it was not desired, contact with out-group members who had the same status, in a situation of interdependence and where they would have shared at least some common goals (studying for a degree, enjoying college life and so on) appears to have a positive impact. In real-world situations though, it may be difficult to achieve all the conditions that increase the impact of contact.

Imagined contact

Research by Turner and Crisp (2010) suggests that the benefits of contact can be reaped simply by imagining positive contact. In two well-designed studies, participants were asked to imagine talking to either an elderly stranger (study 1 with 25 participants) or a Muslim stranger (study 2 with 40 participants). Establishing a positive connection for two minutes resulted in participants having more positive implicit attitudes towards their target group than those in the control groups.

Explanation for the positive impact of contact

Understanding why contact has the impact it does would be helpful in designing interventions based on contact. Contact could possibly have the impact it does because it:

= reduces intergroup anxiety

= disconfirms stereotypes

= individuates the minority group member

Or it could be any combination of the above.

Karpinski and Hilton (2001) developed an environmental association model interpretation of implicit attitude formation. According to this theory, negative implicit attitudes are a product of exposure to mostly negative representations of African Americans. Quality contact with black people establishes positive associations which in turn lead to a reduction of negative implicit associations. Taking this research into consideration with the research on the impact of positive out-group examples suggests that appropriate representation of minority groups should be a key concern for organisations.

Policy implications

Quality of contact appears to be more important than quantity of contact (Aberson, 2004) and exposure may not need to be sustained for long periods to be effective. This makes the earlier suggestions regarding visiting lecturers and the promotion of positive images even more important.

It also has implications for mentoring. Having friendship potential is seen as a key factor which contributes to contact successfully affecting prejudice. Mentoring is seen as an important method for supporting minorities in organisations (Kalev *et al*, 2006), however, because of the social distance involved, the usually senior majority group mentor and the junior minority group mentee in traditional mentoring schemes may have less impact on the implicit attitudes of the majority group senior mentor (Fiske, 1993).

Reverse or reciprocal mentoring schemes (Harvey *et al*, 2009), where senior staff are mentored by talented minorities, may create more equality in the relationship, enabling greater friendship potential and, in turn, a greater impact on implicit attitudes.

However, such models rely on the majority mentor being open to the learning experience and potentially having to challenge their own views and beliefs. Where this is attempted, it needs to be undertaken without leading to the person 'freezing' or seeming cold as a result of feeling uncomfortable (as identified in some of the studies below).

Using motivation to change responses to an out-group

Some people are more internally motivated to control bias – it is something that accords with their values and sense of self. Others are more externally motivated, for these people bias reduction is not a personal value, but they are concerned to conform to social norms on the issue; they do not want to appear to be biased.

There is a considerable body of research that reveals how different types and levels of motivation to control bias impact behaviour (Devine *et al*, 2002). However, research on motivation to control bias has focused on individual differences and only the study reported below was designed to explore the impact of motivation-based bias-reduction interventions on implicit attitudes. This study also assessed whether certain popular practices might, in fact, increase bias.

Conditioning and measuring motivation

Legault *et al* (2011) randomly assigned participants to one of three groups which were assigned different information brochures about prejudice before their motivation to reduce their prejudice was measured:

= an autonomy-brochure condition, where:
'the value of non-prejudice was emphasised. Participants' inner motivation for prejudice reduction was encouraged by emphasizing choice and explaining why prejudice reduction is important and worthwhile'

= a controlling-brochure condition, where:
'participants were urged to combat prejudice and to comply with social norms of non-prejudice'

= a no-brochure condition, where:
'participants read only introductory information about the definition of prejudice'

(Legault *et al*, 2011: p1473)

Autonomy was designed to facilitate an internal motivation to control prejudice and the controlling condition was designed to facilitate an external motivation to control prejudice.

The study found that participants in the autonomy-brochure condition demonstrated considerably less bias than those in the no-brochure or the controlling-brochure condition. However, it only examined explicit bias.

In a subsequent study, the brochures were replaced with a more subtle method in which participants completed a questionnaire where the primes were either autonomous or controlling:

'Autonomy-prime condition:
I enjoy relating to people of different groups.
Being non-prejudiced is important to me.
I can freely decide to be a non-prejudiced person.
I value diversity.
It's fun to meet people from other cultures.

Controlling-prime condition:
It is socially unacceptable to discriminate based on cultural background.
People should be unprejudiced.
I would be ashamed of myself if I discriminated against someone because they were black.
People in my social circle disapprove of prejudice.'

(Legault *et al*, 2011: p1475)

Implicit bias was assessed using an IAT. Those participants primed with autonomy displayed no preference for white over black, while those in the controlling condition displayed considerably more bias than those in the no-prime condition.

Policy implications

Approaches to diversity training that emphasise controlling tactics and undue pressure may, in fact, be counterproductive; it is important that people feel empowered and given the autonomy to control and manage their own biases.

Additionally, consider what motivates people to want to control their biases. Mitigating implicit bias in decision-making is not just a 'nice to have', but something that potentially increases efficiency and quality.

Ideology

Richeson (2004) examined the effectiveness of adopting a multicultural v a colour-blind approach for the reduction of implicit bias. According to Richeson:

'colour-blind perspectives advocate reducing, eliminating, and ignoring category memberships, whereas multiculturalism advocates considering, and sometimes emphasizing and celebrating, category memberships.'

(Richeson, 2004: p417)

To determine the impact of either ideology, participants were randomly assigned to either a colour-blind or a multicultural condition in which they read a one-page statement of the respective ideology and then completed two tasks designed to encourage participants' agreement with the ideology. The participants then carried out an IAT followed by various measures of explicit bias.

Although there was a robust pro-white bias in participants' responses in both groups, participants exposed to the colour-blind ideology demonstrated a larger pro-white bias compared with participants in the multicultural ideology condition. However, in neither control group was a way found to determine how effective this intervention would be in comparison to just doing nothing; both interventions could be making bias worse by making race salient, just that one is making things slightly less worse than the other.

Subsequently, a study that attempted to replicate the findings under conditions of low and high conflict, and which did include a control group, found no significant difference between the two ideologies and the control in either situation (Correll *et al*, 2008). However, this study did use a different type of measure of implicit attitudes that could have been less effective than the one used in the Richeson (2004) study.

Policy implications

Whether to target initiatives at particular groups of people could become a whole study on its own. There are debates around ensuring groups are not stigmatised or problematised, and mechanisms for reducing the potential for backlash also need to be considered.

However, it is interesting that in this study, a 'colour-blind' approach was not as effective. Institutions may want to consider this when developing initiatives aimed at tackling underrepresentation and advancing equality.

Encouraging people to take responsibility for their implicit biases

The methods for reducing implicit bias reviewed so far have been strategies in which the participant is essentially passive – they are not consciously involved in enacting the strategy designed to reduce implicit bias. In fact, researchers in many of the experiments we have reviewed go to considerable lengths to hide the purpose of the study through the use of cover stories and filler tasks.

The following studies on implementation intentions and training review research in which the participant is consciously involved. A key concern with methods of conscious control of implicit attitudes is that calling attention to race can increase implicit bias, even if the purpose of making race salient was to get the participant to avoid the influence of race (Payne *et al*, 2002). Methods of conscious control are important. These put the person in a position to choose to use a method of bias control when they know they are in a situation either where they know they have strong bias, or where the stakes are high like recruitment and promotion situations.

If-then action plans

Implementation intentions are if-then action plans, for example, 'if I crave a snack then I will eat an apple'. They have been found to be more effective than simple goals, for example, 'I want to eat more healthily'. They allow people to 'automate' goal pursuit by strengthening the association between the environmental clue (craving) and the intention (eating an apple) making both more accessible. Furthermore, once the intention has been set they require little effort and motivation to employ (Brandstatter *et al*, 2001).

Self-priming

Both of the studies in this section used the so called 'shooter task' (Correll *et al*, 2002) to assess the impact of their intervention.

The shooter task

In the shooter task participants have to distinguish whether a figure is carrying a gun or a neutral object like a phone. Typically, participants will overestimate the number of times the object is a gun for black figures and underestimate for white figures. The focus of the participant's attention is on correctly identifying the object, and the fast response times mean that they are unaware of the influence race has on their

response. Any difference between the responses to black and white figures is therefore a measure of implicit bias.

In the Stewart and Payne (2008) study, 146 participants were divided into three groups and each group was given a different implementation intention when seeing a black figure:

- 'think safe'
- 'think accurate'
- 'think fast'

Think 'safe'

The automatic bias to respond 'gun' in response to a black face was significantly higher than in response to a white face in the 'think fast' and 'think accurate' trials, but not in the 'think safe' trails. This result was obtained without the 'think safe' participants having to slow down their response times, indicating that the reduction in stereotyping was facilitated by a change in automatic stereotyping and not controlled thinking. Importantly, the reduction in implicit bias was obtained with little practice and was generalised to novel faces.

Further self-priming

In a refinement of the Stewart and Payne study, Mendoza *et al* (2010) obtained similar results with an instruction that was reduced to:

'If I see a person, then I will ignore his race!'

This was repeated three times internally. The importance of this result is that not only did the counter-stereotypic intention reduce bias, but that it did so fairly effortlessly.

Policy implications

As described below, training people in implicit bias can be most effective when people are empowered with strategies they can choose to use to mitigate the impact of the bias after the training.

Encouraging people to remind themselves to think of candidates as people and to ignore their social characteristics is a simple and potentially effective thing to do.

Taking an IAT

Three studies were identified that examined the effectiveness of taking an IAT.

Vorauer (2012) explored the impact of taking an IAT on a subsequent interracial interaction. In previous research, Vorauer had identified that people with a high motivation to control prejudice would 'choke under the pressure' of appearing to be unbiased in an interracial interaction and make them appear guarded and unfriendly (Vorauer and Turpie, 2004). As a consequence, she wanted to challenge the findings of other researchers in the field who argued that being aware that one might be prejudiced might lead to self-correcting behaviour (Monteith and Mark, 2005). According to Monteith and Mark, when people who are highly motivated to control prejudice identify that they have not performed as they should, they use the 'mistake' as a cue to perform differently in a subsequent situation.

Vourauer (2012) suggests that there are four possible responses a participant might have on becoming aware of their own prejudices.

= **Increased awareness prompting increased self-control**
People learn to control prejudiced responses through self-regulatory outcomes that follow from awareness of failures to control stereotyping. Being aware leads to efforts to improve future interracial reactions (Monteith *et al*, 2002).

= **Enhanced caution in interracial interactions**
People respond to the possibility of being biased by 'adopting a cautious, prevention-oriented approach to intergroup exchanges that involves limited self-disclosure and engagement; fear of saying or doing the wrong thing may lead to inhibited behaviour that comes across as unfriendly and aloof'.

= **Reduced feelings of efficacy to handle interracial interactions**
People's 'focus on their limited control over their responses may undermine feelings of efficacy during intergroup interaction – feelings that are important for positive intergroup experiences'.

= **Increased stereotyping**
The 'mere task of categorizing people into one of two clearly separate groups, especially when one of them is the in-group, may detract from the complex, creative and inclusive thinking that has been shown to reduce stereotype activation and prompt more positive intergroup behaviour'.

Subjects were randomly assigned to a race IAT, a race-irrelevant IAT, or a control that took no IAT. Subjects who took the race IAT were not given feedback on their results. Immediately following their tests the white subject either interacted with an aboriginal (Canadian) or a white subject. In both conditions participants then discussed a list of topics in a set order for 15 minutes.

Following the interaction the discussion partners completed various self-report measures of their perception of the interaction. The significant finding is that aboriginal partners felt that their white discussion partners viewed them less positively if the white partner had taken the race IAT.

Getting IAT feedback in a supportive setting

Morris and Ashburn-Nardo (2009) explored the impact of taking an IAT, and then participating in a supportive undergraduate classroom setting. After taking the IAT and being given feedback, measures of participants' feelings and emotions were taken before and after a classroom discussion. A more positive than negative effect was found on both occasions.

An unpublished study that explored the impact of being given feedback where participants had high or low implicit bias on selection of an obese v average weight candidate, found no significant difference between either condition or the control. However, the design of the study was very transparent so it is hard to rule out that participants compensated to prove that were not biased (Henry-Darwish and Sanford, 2012).

Policy implications

Handled and conducted in the wrong way, people may react negatively to an IAT. It is important that they do not inadvertently freeze, avoid, feel threatened, or even antagonistic towards interactions with out-groups as a consequence of taking an IAT.

It is probably not unreasonable to conclude that most people taking bias testing will want the results kept to themselves, but even with that proviso it is important the bias testing is part of a supportive programme in which people learn effective bias-mitigation strategies.

Impact of a multicultural counselling course

The impact of taking a multicultural counselling course on implicit attitudes was compared with just taking an introductory counselling course (Castillo and Brossart, 2007). Data was collected from ten classes (five of each course) over a period of three years, giving a total sample of 87, of which 84 completed the final IAT. Only the participants on the multicultural counselling course revealed a reduction in implicit bias. However, the multicultural training course was delivered by two Latina assistant professors and the control group were taught by two white professors. This means that despite the careful design of most aspects of the study the reduction in implicit attitudes could have been due to exposure to the positive examples and not the course content.

Impact of a bias and conflict seminar

A study where students enrolled in a bias and conflict seminar was found to have significantly reduced implicit and explicit anti-black biases, compared with control students. However, the researchers failed to control for the positive effect of contact or impact of a positive minority example provided by the black lecturer and black students on the course (Rudman *et al*, 2001).

Training

The success of bias reduction through training has been difficult to demonstrate because of poor design and ineffective evaluation of the impact of such programmes (see Paluck and Green, 2009 for a systematic review). Most training studies do not measure implicit bias. The studies that we found that do measure implicit bias are similarly varied in their findings and can suffer from methodological difficulties in isolating the effect being studied from other confounding variables.

However, the two studies below include useful training elements.

Longitudinal impact of a faculty bias reduction programme

A bias literacy programme for a university faculty used a methodology in which implicit bias was treated as a 'remediable habit' (Carnes and Devine, 2012). Although participants were offered the opportunity to complete an IAT before the programme there was no compulsion to do so and no measurement of implicit attitudes before and after the programme.

Evaluation was through a satisfaction survey; 74 per cent found it very useful. Four to six months after the workshop, 24 of the participants were interviewed and provided evidence of the impact on explicit bias. Qualitative information from the interviews was decoded using a procedure in which interviewee statements were assigned to stages of change:

- one third of the interviewees, principally men in science, technology, engineering and mathematics departments, denied there was any need for change
- 23 of the 24 participants described the main benefit as 'increased awareness' of bias
- 75 per cent of those interviewed also expressed actual attitudinal change or plans to change (although the proportion changed and the proportion planning change is not stated)

There were no independent measures of attitudinal change either within the people who took the programme or in the organisation.

Breaking the habit

Devine *et al* (2012) developed a multi-method, longitudinal bias-reduction intervention. The study used a control group and also utilised before and after measures of implicit and explicit attitudes. Devine considers bias reduction as akin to 'breaking the habit' and that therefore overcoming bias is a lengthy process that requires considerable effort in the pursuit of a non-biased goal (Devine, 1989). It requires learning about the situations that trigger bias and how to replace biased responses with responses that reflect the person's non-biased goals.

All participants in the Devine *et al* (2012) study took an IAT and received feedback on their bias scores. Feedback in Devine's model is thought to provide situational awareness of bias, which in turn provokes guilt (in those who are internally motivated to control bias), which in turn motivates commitment to using strategies to reduce bias (Monteith and Mark, 2005).

The experimental group undertook the training intervention and the control group just received the feedback on their level of bias. Additional measures of explicit bias and motivation to respond without bias were also taken.

The training section provided participants with five strategies. The programme explained the strategies in straightforward language with concrete examples of everyday situations in which they could be used. Participants were then asked to generate situations in which they could use each strategy.

'Stereotype replacement

This strategy involves replacing stereotypical responses with non-stereotypical responses. Using this strategy to address personal stereotyping required the recognition that a response is based on stereotypes, labelling the response as stereotypical, and reflecting on why the response occurred. Next one had to consider how the biased response could be avoided in the future and replaced with an unbiased response (Monteith, 1993).

Counter-stereotypic imaging

This strategy involves imagining, in detail, counter-stereotypic individuals (Blair *et al*, 2001). These people can be abstract (eg smart black people), famous (eg Barack Obama), or non-famous (eg a personal friend). The strategy makes positive examples salient and accessible when challenging a stereotype's validity.

Individuation

This strategy relies on preventing stereotypic inferences by obtaining specific information about group members (Brewer, 1988; Fiske and Neuberg, 1990). Using this strategy helps people evaluate members of the group based on personal, rather than group-based, attributes.

Perspective taking

This strategy involves taking the perspective of a member of a stereotyped group. Perspective taking increases psychological closeness to the stigmatised group, which ameliorates automatic group-based evaluations (Galinsky and Moskowitz, 2000).

Increasing opportunities for contact

This strategy involves seeking opportunities to encounter and engage in positive interactions with out-group members. Increased contact can ameliorate implicit bias through a wide variety of mechanisms, including altering the cognitive representations of the group or by directly improving evaluations of the group (Pettigrew, 1998; Pettigrew and Tropp, 2006).

Participants were told that although none of the strategies are difficult to implement, each requires some effort. In addition, the programme emphasised how the strategies are mutually reinforcing. For example, contact with counter-stereotypic people provides grist for counter-stereotypic imaging as well as providing opportunities for individuation, perspective taking and stereotype replacement. Similarly, perspective taking can enhance stereotype replacement and individuation by encouraging people to see the world through the eyes of a stigmatised person. As a set, the strategies were offered as a powerful toolkit for breaking the prejudice habit. The programme also stressed that practising the strategies would help them to reduce implicit bias and, hence, break the prejudice habit.'

(Devine *et al*, 2012: pp1270–1271)

Results showed that participants' level of implicit bias was reduced as a result of the intervention and this effect persisted over the eight weeks of the programme. In fact, the effects of the intervention on implicit race bias at four and eight weeks were not different from each other, indicating that the reduction in implicit race bias persisted at the same level throughout the eight-week interval.

Policy implications

The studies show that it is not enough to simply alert people to the existence of bias and/or to alert them to their own particular biases; people need to be given strategies for addressing their biases which make them feel empowered and autonomous, rather than guilty and controlled.

Strategies need to prime an implementation intention, which may increase feelings of personal control, to ensure people know what they are going to do when a situation arises, rather than just telling themselves that they will not be biased in a particular situation.

The Devine *et al* (2012) studies show that training can work, and need not be particularly complicated, or lengthy. However, a requirement appears to be that participants feel that they have a degree of autonomy in their choice and use of these strategies.

A key part of the Devine *et al* approach was that participants were offered a range of techniques that they could choose to use. They were, however, encouraged to report on their use of their strategies over the course of the programme and had to return for two further sessions of testing and questionnaire completing. Undoubtedly this degree of accountability would have encouraged participants to practise their bias-mitigation strategies, a luxury that may not exist so easily in a real-world setting.

Conclusions

While there may be debate over how to identify and measure implicit bias and its impact, the evidence suggests that we need to address its influence. Institutions need to ensure their policies and procedures are bias-proofed wherever possible and individuals need to consider how their own biases might impact on their behaviour and decisions.

There is no easy answer for how to mitigate the impact of implicit bias, but ECU is committed to working with the sector to develop strategies for tackling this and creating a more inclusive environment in which everyone is able to reach their full potential.

Appendix 1 Methodology

Aim

There is a large amount of literature on implicit bias. The primary aim of this review was to use existing literature, research and publications to ascertain if and how BME applicants (both internal and external) may be affected by implicit bias in recruitment and selection, and if they are, what institutions can do to mitigate that impact.

The review assumes that implicit bias exists, and that if BME people are affected by implicit bias in recruitment and selection then other groups of people will also, for example, women. No attempt was made to ascertain whether BME people are affected to a greater or lesser extent than any other group of people.

Type of review

Consideration was given to different methods for conducting literature reviews and it was felt that a full systematic review would be not only impossible within reasonable timeframes, but not wholly necessary or helpful. Equally, it was felt that some limitations of a narrative review process should be avoided.

Due to the nature of the subject area and the specific area of interest, it was felt to be prudent to take a mixed-methods approach to the review to ensure a robust and transparent review was conducted, without stifling the scope of literature.

The review was systematic in so far as:

- it focused on answering two main questions:
  'Are BME staff impacted by implicit bias in recruitment and selection in UK higher education? If they are, what can institutions do to try to mitigate that impact?'
- a team agreed the search terms, initial inclusion and exclusion criteria, the databases that were used to search for literature and any additional criteria that were applied to search results
- a record was kept of search results per term and database source
- where the literature contained primary research (which was the majority of the literature), time was spent reviewing the robustness of the research methodology and conclusions

However, elements of narrative review were used in that:

- some assumptions were made, namely that implicit bias does exist; the review itself focused on the impact of that bias and whether it does impact on recruitment decision making, and if it does, to what extent
- researchers used agreed search terms to identify potential literature and then used personal judgment in reviewing the title and abstract to decide which literature to longlist. However, this was done with the aim of the review as the central consideration
- the collective longlist was collated and then the team collectively reviewed the list to agree which literature to shortlist for the final review. The team documented the decision-making process and the rationale for excluding/including certain literature
- where researchers were aware of relevant research, or became aware through reviewing the identified literature, articles which had not been picked up by the main searches were also included in the review

Identifying a longlist of literature

The search for literature took place in November 2012. After initial scoping, the search was conducted in the following databases:

- EBSCO
- PsycINFO
- Social Science Research Network (SSRN)
- Education Resources Information Center (ERIC)
- British Library online database search
- ABI/INFORM
- Google Scholar (top 40 records only)

The following terms were searched, where possible using Boolean logic:

- Unconscious OR implicit AND bias AND recruitment
- Unconscious OR implicit AND bias AND selection
- Unconscious OR implicit AND bias AND employment
- Unconscious OR implicit AND bias AND higher education OR HE OR university
- Unconscious OR implicit AND bias AND academia (or where possible to encompass all iterations: acad*)
- Unconscious OR implicit OR bias AND race
- Unconscious OR implicit OR bias AND racism
- Unconscious OR implicit AND bias AND ethnicity (or where possible to encompass all iterations: ethnic*)
- Unconscious discrimination AND recruitment OR selection OR employment
- Implicit discrimination AND recruitment OR selection OR employment
- Unconscious OR implicit AND attitudes AND recruitment
- Unconscious OR implicit AND attitudes AND employment
- Unconscious OR implicit AND attitudes AND Higher education OR HE OR university
- Unconscious OR implicit AND attitudes AND race
- Unconscious OR implicit AND attitudes AND racism
- Unconscious OR implicit AND attitudes AND ethnicity
- Mitigating OR preventing AND unconscious OR implicit AND bias OR attitudes

Identifying a shortlist of literature

The longlist identified 248 items (once duplications had been removed). The longlist was examined independently by the team to explore the abstracts and summaries of each piece to assess its relevance to the aim.

A document or study was considered relevant if it:

= pertained to the implicit bias construct

= referred to IAT methodology/validity

= had involved IAT in any applied context

= had been carried out in any sort of staff or academic setting

Documents or studies also had to have some empirical base for their content. Opinion pieces or reviews without data gathering or analysis were excluded.

A shortlist of 138 studies was divided into four themes:

A implicit bias theory

B IAT methodology

C IAT predictive validity

D implicit-bias interventions

The shortlist studies in themes B, C and D were then individually examined and assessed against an adapted version of the Maryland scale of scientific methods. That scale included whether the study used:

= control groups

= near-randomised sampling

= before and after measurements

= between-group designs

The study sample used was also recorded (source, size and where relevant any demographic data) and the outcome measures employed.

Theme A was considered too broad for a review of this scope, as each theory could be subjected to a review of its own. It was therefore decided to exclude it from the formal review, although the theories included within those studies have formed the basis of section 1.

Where necessary, for papers judged to be highly relevant (eg the use of IAT in an applied selection setting), contact was made with authors for clarification of methods or to access information not cited in the study. Furthermore, studies that had no empirical base of their own were accessed when they had been often cited in references of the studies still under examination. For example, although the paper by Blanton and Jaccard (2006) was largely a review and opinion document, it triggered an exchange of similar review and response papers from the research team responsible for the IAT at Harvard University and it was also cited frequently in key empirical studies (and 306 citations in Google Scholar).

Appendix 2 Implicit association tests

Background to the IAT

The IAT is used to examine positive or negative associations and therefore measure our implicit biases.

Since its inception in 1998 the IAT has become the dominant methodology for measuring implicit bias. However, the principles behind the IAT are not new. A 19th century Dutch doctor, FC Donders, is believed to be the first person recorded as suggesting that the way to tap into unobservable mental processes was through the speed of that processing, and he expended considerable effort developing machinery to do this. However, it was not until the arrival of computers that truly accurate response times could be routinely measured outside a laboratory setting. IAT is believed to bypass two fundamental flaws in traditional explicit measures (such as self-report questionnaires), namely:

- the lack of self-insight
- the unwillingness to report socially undesirable attitudes

Response latency is the time taken between being delivered a stimulus, for example a picture or word of a particular group or characteristic, and the person's response. The modern IAT measures response latency through computerised timing and is designed to tap individual differences in automatic associations between concepts (eg white v Asian people) and attributes (eg good v bad). The IAT requires the participant to classify stimuli (concepts and constructs presented as words or pictures) rapidly and rests on the premise that easier pairings (ie faster responses) are more strongly associated than more difficult pairings (ie slower responses).

The IAT was introduced in recent scientific literature in Greenwald *et al* (1998) after an extended period of development and testing. The IAT is now widely used in social psychological research and also in other settings including political science, consumer, clinical and forensic psychology. As the Greenwald *et al* (1998) IAT was developed by a team at Harvard University it is often referred to as the Harvard IAT or 'Project Implicit' in practitioner publications and discussions. The Harvard IAT has become, as Rooth (2010) described it, the 'standard tool' in social psychology when studying implicit associations (Rooth, 2010: p524).

The IAT is not the only implicit measure, and neither is the IAT a single test. Over the years various derivatives of the Harvard IAT have been developed to address particular research contexts or to respond to technical questions over the Greenwald *et al* (1998) IAT methodology. This section of the review specifically excludes these derivatives of the IAT, for example, the single category IAT, the single block IAT, the personalised IAT, the brief IAT, the go/no-go task, the extrinsic affective Simon task and all derivatives of the approach-avoidance task. Within a relatively short document it would be impossible to even describe, let alone evaluate and compare other implicit measures or the many IAT derivatives. The Harvard IAT has been the subject of hundreds of academic papers since 1998 and has been completed by over 14 million people.

The IAT procedure

The Harvard IAT procedure involves a series of tasks. In each task, the participant is asked to use the assigned 'left' and 'right' keys on the computer keyboard to categorise words or picture stimuli rapidly. Often the test constructs (eg good or bad) are represented by a word and the test concepts (social groups) are represented by pictures (eg black and white faces) although it is possible to use words or pictures for both concept and construct. Items are usually either randomly presented or presented in a partially randomised order; no two tests are exactly alike.

IAT scores are calculated based on response time differences between the compatible and incompatible tasks using a 'D score' algorithm developed by Greenwald *et al* (2003).

D score algorithm paper

To improve the original algorithm a number of new algorithms were examined in terms of their:

- correlations with parallel self-report measures
- resistance to an artefact associated with speed of responding
- internal consistency
- sensitivity to known influences on IAT measures
- resistance to known procedural influences

The best-performing measure incorporates data from the IAT's practice trials, uses a metric that is calibrated by each respondent's latency variability, and includes a latency penalty for errors. This new algorithm strongly outperforms the earlier (conventional) procedure.

The D score algorithm takes into account the baseline response speed of the participant as measured by the practice trials (stages 1 and 2) and uses the standard deviation of the participant in those practice trials. It also controls for the time taken in correcting errors (the IAT requires any incorrect categorising to be corrected with a second key press).

(Greenwald *et al*, 2003)

Criticisms of the IAT

The Harvard IAT is not universally accepted as a valid and reliable measure of a person's implicit biases. Blanton and Jaccard (2006) produced an opinion paper highlighting their concerns around the construction and use of the Harvard IAT, with a focus on the version developed to explore implicit bias towards other ethnic groups (the race IAT). They raised a number of concerns about the way in which the Harvard IAT measured and reported, notably:

- measuring attitude with time
- use of log transformations
- methods of scoring bias
- interpreting results
- linking change to real-world situations

Measuring attitude with time

Blanton and Jaccard argued that the use of a time metric (milliseconds) as a basis for measuring an attitudinal construct was problematic:

'The metric of milliseconds is arbitrary when it is used to measure the magnitude of an attitudinal preference. An attitudinal preference for one group over another is no more an expression of milliseconds than it is an expression on a rating scale.'

(Blanton and Jaccard, 2006: p32)

Furthermore, the absence of a metric-to-metric correspondence raised issues: how much of the attitudinal construct (implicit bias) is indicated by a unit of measurement (time)? The Harvard IAT was seen to be unable to offer this.

Greenwald *et al* (2006) accepted that as a relatively new metric much had yet to be discovered about the IAT, but argued that many of the allegations of meaningless metrics were subsumed into the notion of 'consequential validity' (see Messick, 1995).

Greenwald *et al* (2006) asserted that the well-known conventional associations of values with labels such as small (or weak), medium (or moderate), and large (or strong) devised by Cohen (1977) were appropriate, although no link to behaviour was evidenced for those effect sizes. They pointed to Poehlman *et al* (2005) as a meta-analysis of evidence showing predictive validity of the IAT and presented data from 8529 web-based IATs showing a 0.730 (very strong) correlation between IAT scores and voting intention in the US elections.

Use of log transformations

Regarding the use of log transformations of time data by use of the standard deviation of the test takers' response times, Blanton and Jaccard argued that:

'It is hard to imagine why one would transform response latencies as measured in milliseconds to get a better estimate of time, but there are many reasons why one might want to do this to obtain a better estimate of an attitude.'

(Blanton and Jaccard, 2006: p32)

The Harvard transformation was designed to get a better estimate of time (response latencies) by accounting for the overall processing speed of the test taker.

Methods of scoring bias

The lack of a 'zero point' raised issues around how the test measured no preference in a participant. As the IAT is thought to measure a theoretical construct (implicit bias) that reflects a preference for one social group relative to another, the assumed zero point on the theoretical construct is of no preference. Blanton and Jaccard argued that there was no evidence as to the true zero point on the IAT in terms of attributing no preference.

Normative scoring of the Harvard IAT was a key criticism, ie a test taker's score is measured in relation to other test takers. The Harvard IAT placed a test taker within the sample group, but Blanton and Jaccard argued that this was not evidence of biased behaviour:

'Although it is true that such standardization can convey important and useful information about relative standing, standardization alone cannot convey someone's absolute standing on a psychological dimension of interest, nor does it necessarily calibrate a measure to meaningful external events.'

(Blanton and Jaccard, 2006: p36)

Interpreting results

Blanton and Jaccard took issue with the way score labels were assigned to test results. Harvard had used Cohen's model of small, medium and large effect sizes, and told IAT test takers that they have a 'slight preference' if they have a normed IAT score between 0.20 and 0.50, a 'moderate preference' if they have a normed IAT score between 0.50 and 0.80, and a 'strong preference' if they have a normed IAT score greater than 0.80.

Blanton and Jaccard argued that these assignments were arbitrary and had no evidence of measuring the underlying construct of implicit bias.

Linking change to real-world situations

The inability to report 'clinically significant change' was Blanton and Jaccard's final criticism of metrics in general, including the IAT. They felt that it is not enough to show statistically significant mean changes on an outcome measure that reflects a psychological construct (implicit bias), but it also must be shown that those changes have meaningful consequences for peoples' lives.

The predictability of the IAT is discussed further below.

Debates over the reliability of the IAT

Little in the literature addresses the reliability of the IAT. If a test is to be valid it must first be reliable because if a test score delivered by the same person at different times changes significantly without an intervention we may not be measuring the underlying construct (implicit bias) consistently. Tests have two basic types of reliability; internal and external.

Internal reliability

Internal reliability refers to the way in which the items that make up the test relate to each other. For example, we would expect scores from items in a mathematics test examining multiplication skills to correlate highly with other items examining multiplication. With the IAT each test is unique, but where calculations of internal consistency have been made, they have delivered internal reliability around 0.60. This borderline internal consistency is rarely mentioned in academic papers and does seem to fall at or below the standard required in commercial and research settings which seek a Cronbach's alpha of 0.7.

Cronbach's alpha	Internal consistency
$\alpha \geq 0.9$	Excellent
$0.9 > \alpha \geq 0.8$	Good
$0.8 > \alpha \geq 0.7$	Acceptable
$0.7 > \alpha \geq 0.6$	Questionable
$0.6 > \alpha \geq 0.5$	Poor
$0.5 > \alpha$	Unacceptable

External (test-retest) reliability

Most commercial test publishers expect to see a test-retest reliability of 0.70 or above. Test-retest reliabilities of r = 0.65 and r = 0.69 were reported respectively, by Dasgupta and Greenwald (2001), and by Bosson *et al* (2000).

The Project Implicit team have recently (2013) released, but as yet not published, further research on a shortened version of the IAT called the brief IAT (Sriram and Greenwald, 2009), which uses shorter trials (Nosek *et al*, 2013). In this paper they outline a new 'G' scoring algorithm which appears to show marginally better prediction than the D score algorithm.

Variables that may have an effect on the IAT

Faking on the IAT

The IAT is an indirect measure of the implicit bias construct and as such should be less vulnerable to attempts to distort scores than, for example, self-report questionnaires or interviews. However, as a quite simple time-based metric it is easy to see how a participant may consider trying to fake a result.

Kim (2003) found no faking effects and other studies have found faking effects where the effect was not significant in magnitude (Asendorpf *et al*, 2002; Egloff and Schmukle, 2002; Banse *et al*, 2001). However, participants in those studies had no experience with the IAT, and faking becomes more likely with experience of the test.

Steffens (2004) sought to explore this issue of experience further and employed a personality-based IAT. She gave 48 undergraduate test takers simple instructions to attempt to fake. Reaction times at base rate and at faking still correlated, suggesting that test takers could not completely hide their implicit associations when trying to fake the test. The effect was more apparent in the number of errors made than on reaction times. Coupled with the effect size, these findings suggested a limited ability to fake the IAT. In a second study with 125 undergraduates, half were given information about the criticality of timed responses in calculating scores and instructed to fake and half were not given the hint about timed responses and were just instructed to fake. Steffens found that the IAT is susceptible to faking, but to a limited degree:

> 'The IAT is much less susceptible to faking than questionnaire measures are, even if no selective faking of single dimensions of the questionnaire occurred. However, given limited experience, scores on the IAT, too, are susceptible to faking.'
>
> (Steffens, 2004)

A note should be made at this point about the context of use of the IAT. The motivation to fake is a product of attempts to produce socially attractive results and/or avoid censure. This is most prevalent in high-stakes settings (eg staff selection) or where results are to be seen by others who may judge the test taker based upon their test scores.

Policy implications

If a desire to fake an IAT is driven by an attempt to produce the 'right' results, or the result an employee thinks their employer wants to see, it potentially defeats the purpose of the IAT. Where employers want their employees to consider their own implicit attitudes and how those might impact on their behaviours and actions, the IAT should be delivered in a safe setting. People should feel able to be honest about their test results and not feel afraid to discuss them. Employers may want to think about the confidentiality of test settings and also support for test takers after the test. Rather than the results of the test being the end product, perhaps they could be the starting point of relevant support and bias-reduction strategies.

Malleability of the IAT

Han *et al* (2010) investigated the extent to which the IAT may be influenced by participants being unsure about how to interpret words or stimuli. They suggested that words such as 'pleasant' or 'unpleasant' used in the Harvard IAT are open to multiple interpretations, and therefore may lead to different outcomes.

Using a sample of 53 undergraduates they tried to manipulate IAT scores. They primed the students before taking the IAT in two ways: one group was asked to report on things they liked and the other on things that they felt others liked. This was followed by a race-based IAT. Those who had completed the 'I like' questionnaire prior to IAT testing showed lower scores compared with the 'others like' condition. The mechanism proffered by Han *et al* (2010) is that the words used in the IAT are open to some interpretation and that the priming questionnaire affected the way those interpretations were carried out during testing.

Test familiarity

IAT effects were marginally reduced in size for test takers who had prior experience taking the IAT (Greenwald *et al*, 2003).

Familiarity with stimuli

Greenwald and Nosek (2001) reported that using unfamiliar stimuli (words and pictures) may yield problematic IAT findings. The IAT operates at the level of the categories and when the task stimuli fall into no existing category, the IAT appears not to work as expected. The situation might be similar if items chosen to represent a category in an IAT were difficult to classify.

IAT task order

Greenwald *et al* (1998) noted that the strengths of associations used in the first of the IAT's two combined tasks appeared to be stronger than those used in the second. This finding has since been observed in numerous other studies. If task order affects IAT scores and it is not addressed then scores may not be an accurate reflection of the construct (implicit bias) and may in part be attributable to presentational issues.

Nosek *et al* (2005) showed that this order effect is reduced with an amount of practice on the later tasks of the IAT.

IAT task switching

The Project Implicit team at Harvard regard task-switching effects (the slower adjustment to the change in sorting rules) may be integral to what is measured by the IAT, rather than being an artefactual contributor (ie one resulting because of the investigation) to IAT effects.

Right-hand/left-hand preference

Greenwald and Nosek (2001) reported no effects on IAT scores of test takers' right-hand/left-hand preference.

Cognitive fluency and age

Cognitive fluency and age are both associated with a general slowing of response, which tends to increase IAT scores. The Project Implicit team asserts that the effects of cognitive fluency and age are substantially moderated by using the IAT's improved D scoring algorithm (Greenwald *et al*, 2003).

IAT stimuli items (examples)

Mitchell *et al* (2003) showed that implicit race attitudes were influenced when the pictures/words chosen were selected to represent, for example, either disliked black and liked white or, alternately, liked black and disliked white. Dasgupta and Greenwald (2001) successfully reduced participants' implicit prejudice through brief exposure to negative in-group and positive out-group examples. They found that exposing white-American participants to these examples versus exposure to race-neutral examples significantly reduced their implicit in-group bias. De Houwer *et al* (2001) clearly showed that the IAT tolerated a good deal of variance in an irrelevant attribute, but this highlights the point that IAT content should not include examples. The period over which examples impact IAT scores is unknown.

Predictive validity of the IAT

Meta-analysis of literature

Greenwald *et al* (2009) conducted a meta-analysis of 122 research reports (including 184 independent samples involving 14,900 subjects) to examine the validity of the IAT in predicting behaviour. This meta-analysis examined the studies by a number of criteria including methodological considerations and inter-rater reliability. They also used three judges, two versed in the literature and one 'blind' to the arena.

'This review justifies a recommendation to use IAT and self-report measures jointly as predictors of behavior. Even though the relative predictive validities of the two types of measure varied considerably across domains, each type generally provided a gain in predictive validity relative to using the other alone. The review found that, for socially sensitive topics, the predictive validity of self-report measures was remarkably low and the incremental validity of IAT measures was relatively high. In the studies examined in this review, high social sensitivity of topics was most characteristic of studies of racial and other intergroup behavior. In those topic domains, the predictive validity of IAT measures significantly exceeded the predictive validity of self-report measures.'

(Greenwald *et al*, 2009: p32)

Re-analysis of the studies

Blanton *et al* (2009) requested raw data from a number of authors of published studies showing behavioural prediction of who had used the Harvard IAT, including some of the studies relied upon in the Greenwald *et al* (2009) meta-analysis above. Their aim was to re-examine the data, methods and conclusions and challenge the predictive reliability of the IAT.

One such study, Heider and Skowronski (2007) was designed to test for links between the race IAT and competitive behaviour in a prisoner's dilemma game.

Prisoner's dilemma game

Tanya and Cinque have been arrested for robbing the Hibernia Savings Bank and placed in separate isolation cells. Both care much more about their personal freedom than about the welfare of their accomplice. A clever prosecutor makes the following offer to each. 'You may choose to confess or remain silent. If you confess and your accomplice remains silent I will drop all charges against you and use your testimony to ensure that your accomplice does serious time. Likewise, if your accomplice confesses while you remain silent, they will go free while you do the time. If you both confess I get two convictions, but I'll see to it that you both get early parole. If you both remain silent, I'll have to settle for token sentences on firearms possession charges. If you wish to confess, you must leave a note with the jailer before my return tomorrow morning.'

The 'dilemma' faced by the prisoners here is that, whatever the other does, each is better off confessing than remaining silent. But the outcome obtained when both confess is worse for each than the outcome they would have obtained had both remained silent. A common view is that the puzzle illustrates a conflict between individual and group rationality. A group whose members pursue rational self-interest may all end up worse off than a group whose members act contrary to rational self-interest. More generally, if the payoffs are not assumed to represent self-interest, a group whose members rationally pursue any goals may all meet less success than if they had not rationally pursued their goals individually. A closely related view is that the prisoner's dilemma game uses familiar situations in which it is difficult to get rational, selfish agents to cooperate for their common good.
http://plato.stanford.edu/entries/prisoner-dilemma

Heider and Skowronski (2007) aimed to eliminate the possibility that a predictive relationship between the IAT and behaviour was due to increases in attitudes being activated or brought to mind. In the prisoner's dilemma game, players win points for choosing the same cooperative or competitive strategy as a co-player but the tendency is to favour cooperation with the player's in-group and defection/competition with the out-group.

Heider and Skowronski's study was designed to test for links between the race IAT and friendliness of verbal and non-verbal treatment of confederates. In the re-analysis Blanton and Mitchell suggest that the published results were based on erroneous IAT scores. They suggest that the race IAT did not predict racially biased behaviour.

In an unusual step, in 2011 Heider and Skowronski published a paper reporting that recently discovered coding errors led to inaccurate reporting of multiple regression results in one of the two studies in the original report. These inaccuracies were of particular consequence for conclusions regarding the ability of the IAT to predict participants' non-verbal friendliness toward black partners in the prisoner's dilemma game. In the re-analysis the IAT was not able to show statistical significance as a predictor.

Policy implications

Despite the concerns and issues with the IAT, it remains the dominant method used to research implicit bias. Rudman (2008: p426) summarises her views on the IAT:

'Researchers "vote with their feet," which explains the IAT's wide usage in the United States and the rest of the world. At present, it is the dominant method for assessing implicit associations because of its robust psychometric features, flexibility, and resistance to faking (Nosek *et al*, 2007). The predictive validity of the IAT is also well established, with a meta-analysis showing that the IAT was a reliable predictor of many behaviors (eg consumer choice, academic major, and voter choice) and clearly superior to self-reports when predicting discriminatory behaviors (Greenwald *et al*, in press). In addition, neuroscience research has supported the IAT's validity by showing that it correlates with brain regions associated with emotional conditioning (eg Cunningham *et al*, 2003; Phelps *et al*, 2000).'

The IAT is evidently not a perfect tool. As a method of measurement it is still being developed and even the biggest advocates of IAT acknowledge that work still needs to be undertaken to refine the test more and ensure its validity to a greater degree. That said, even with lower than preferred reliability, as outlined above, the tests do still show some interesting findings. Ultimately it is for people to make up their own mind as to the implications, if any, of what the tests show.

Bibliography

Aberson, CLS **(2004)** 'Implicit bias and contact: the role of interethnic friendships'. *Journal of Social Psychology* 144(3): 335–347.

Aboud, FE **(1988)** *Children and prejudice*. Basil Blackwell, New York.

Abrams, D **(2010)** *Processes of prejudice: theory, evidence and intervention*. Equality and Human Rights Commission, London. **http://equalityhumanrights.com/uploaded_files/research/56_processes_of_prejudice.doc**

Abrams, D & Houston, D **(2006)** *Equality, diversity and prejudice in Britain: results from the 2005 national survey*. Equalities Review. Cabinet Office, London.

Adriaanse, MA, Van Oosten, JMF, De Ridder, DTD, De Wit, JBF & Evers, C **(2011)** 'Planning what not to eat: ironic effects of implementation intentions negating unhealthy habits.' *Personality and Social Psychology Bulletin* 37(1): 69–81.

Agerström, J, Carlsson, R & Rooth, D **(2007)** *Ethnicity and obesity: evidence of implicit work performance stereotypes in Sweden:* working paper 2007:20. Institute for Labour Market Policy, Uppsala.

Agerström, J & Rooth, D **(2011)** 'The role of automatic obesity stereotypes in real hiring discrimination'. *Journal of Applied Psychology* 96(4): 790–805.

Agerström, J & Rooth, D-O **(2009)** 'Implicit prejudice and ethnic minorities: Arab-Muslims in Sweden'. *International Journal of Manpower* 30: 43–55.

Ajzen, I & Fishbein, M **(1977)** 'Attitude-behavior relations: a theoretical analysis and review of empirical research'. *Psychological Bulletin* 84(5): 888–918.

Allen, TJ, Sherman, JW & Klauer, KC **(2010)** 'Social context and the self-regulation of implicit bias'. *Group Processes and Intergroup Relations* 13(2): 137–149.

Allport GW **(1954)** *The nature of prejudice*. Perseus, New York.

Amodio, DM, Harmon-Jones, E, Devine, PG, Curtin, JJ, Hartley, SL & Covert, AE **(2004)** 'Neural signals for the detection of unintentional race bias'. *Psychological Science* 15(2): 88–93.

Apfelbaum, EP, Norton, MI & Sommers, SR **(2012)** 'Racial color blindness: emergence, practice, and implications'. *Current Directions in Psychological Science* 21(3): 205–209.

Asendorpf, JB, 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: 380–393.

Banaji, MR **(2005)** 'Mind bugs: The psychology of ordinary prejudice'. Paper presented at the Midwestern Psychological Association meeting, Chicago, May.

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

Barden, J, Maddux, WW, Petty, RE & Brewer, MB **(2004)** 'Contextual moderation of racial bias: the impact of social roles on controlled and automatically activated attitudes'. *Journal of Personality and Social Psychology* 87(1): 5–22.

Bargh, JA, 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.

Billig, M **(1985)** 'Prejudice, categorization and particularization: from a perceptual to a rhetorical approach'. *European Journal of Social Psychology* 15(1): 79–103.

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

Blair, IV, Ma, JE & Lenton, AP **(2001)** 'Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery'. *Journal of Personality and Social Psychology* 81: 828–841

Blanton, H & Jaccard, J **(2008)** 'Unconscious racism: a concept in pursuit of a measure'. *Annual Review of Sociology* 34: 277–297.

Blanton, H & Jaccard, J **(2006)** 'Arbitrary metrics in psychology'. *American Psychologist* 61: 27–41.

Blanton H, Jaccard J, Gonzales PM, & Christie C **(2006)** 'Decoding the implicit association test: implications for criterion prediction'. *Journal of Experimental Social Psychology* 42(2): 192–212.

Blanton, H, Jaccard, J, Klick, J, Mellers, B, Mitchell, G & Tetlock, P **(2009)** 'Strong claims and weak evidence: reassessing the predictive validity of the IAT'. *Journal of Applied Psychology* 94: 567–582.

Blanton, H & Mitchell, G **(2011)** 'Reassessing the predictive validity of the IAT II: reanalysis of Heider & Skowronski (2007)'. *North American Journal of Psychology* 13(1): 99–106.

Bosson, JK, Swann, WB Jr & Pennebaker, JW **(2000)** 'Stalking the perfect measure of implicit self esteem: the blind men and the elephant revisited?' *Journal of Personality and Social Psychology* 79: 631–643.

Brandstatter, V, Lengfelder, A & Gollwitzer, PM **(2001)** 'Implementation intentions and efficient action initiation'. *Journal of Personality and Social Psychology* 81(5): 946–960.

Brewer, M **(1988)** 'A dual-process model of impression formation'. In Srull, T & Wyer, R (eds.), *Advances in social cognition* (pp1–36). Erlbaum Associates, Hillsdale NJ.

Brief, AP Jr, Dietz, J, Cohen, RR, Pugh, SD & Vaslow, JB **(2000)** 'Just doing business: modern racism and obedience to authority as explanations for employment discrimination'. *Organizational Behavior and Human Decision Processes* 81: 72–97.

Briner, RB, Denyer, D & Rousseau, DM **(2009)** 'Evidence-based management: concept cleanup time?' *Academy of Management Perspectives* 23(4): 19–32.

Carlsson, M, & Rooth, D-O **(2008)** *Is it your foreign name or foreign qualifications? A experimental study of ethnic discrimination in hiring*. Institute for the Study of Labor, Bonn.

Carlsson, M, & Rooth, D-O **(2007)** 'Evidence of ethnic discrimination in the Swedish labor market using experimental data'. *Labor Economics* 14: 716–729.

Carnes, M & Devine, P **(2012)** 'Promoting institutional change through bias literacy'. *Journal of Diversity in Higher Education* 5(2): 63–77.

Castillo, LG & Brossart, DF **(2007)** 'The influence of multicultural training on perceived multicultural counseling competencies and implicit racial prejudice'. *Journal of Multicultural Counseling and Development* 35(4): 243–254.

Chabris, C & Simons, D **(2010)** *The invisible gorilla*. Crown Publishers, New York.

Coe, R **(2002)** 'It's the effect size, stupid. what effect size is and why it is important.' Paper presented at the British Educational Research Association annual conference, Exeter, 12–14 September, 2002.

Cohen, J **(1977)** *Statistical power analysis for the behavioral sciences* (rev. ed.). Erlbaum Associates, Hillsdale, NJ.

Cohen, J **(1988)** *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum Associates, Hillsdale, NJ.

Committee on Maximizing the Potential of Women in Academic Science and Engineering, National Academy of Sciences, National Academy of Engineering & Institute of Medicine **(2006)** 'Section 1: summaries of convocation sessions.' *Biological, social, and organizational components of success for women in academic science and engineering: workshop report*. The National Academies Press, Washington, DC.

Correll, J, Park, B & Allegra Smith, J **(2008)** 'Colorblind and multicultural prejudice reduction strategies in high-conflict situations'. *Group Processes and Intergroup Relations* 11(4): 471–491.

Correll, J, Park, B, Judd, CM & Wittenbrink, B **(2002)** 'The police officer's dilemma: using ethnicity to disambiguate potentially threatening individuals'. *Journal of Personality and Social Psychology* 83(6): 1314–1329.

Crisp, RJ & Beck, SR **(2005)** 'Reducing intergroup bias: the moderating role of in-group identification'. *Group Processes and Intergroup Relations* 8: 173–186.

Crisp, RJ & Meleady, R **(2012)** 'Adapting to a multicultural future'. *Science* 336(6083): 853–855.

Crisp, RJ & Nicel, JK **(2004)** 'Disconfirming intergroup evaluations: asymmetric effects for in-groups and out-groups'. *Journal of Social Psychology* 144(3): 247–271.

Cuddy, AJC, Fiske, ST, Kwan, VSY, Glick, P, Demoulin, S, Leyens, J-P, Bond, MH, Croizet, J-C, Ellemers, N, Sleebos, E, Htun, TT, Yamamoto, M, Kim, H-J, Maio, G, Perry, J, Petkova, K, Todorov, V, Rodríguez-Bailón, R, Morales, E, Moya, M, Palacios, M, Smith, V, Perez, R, Vala, J, & Ziegler, R **(2009)** 'Stereotype content model across cultures: towards universal similarities and some differences'. *British Journal of Social Psychology* 48: 1–33.

Cunningham, WA, Johnson, MK, Gatenby, JC, Gore, JC & Banaji, MR **(2003)** 'Neural components of social evaluation'. *Journal of Personality and Social Psychology* 85(4): 639–649.

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 AG **(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, N, Greenwald, AG & Banaji, MR **(2003)** 'The first ontological challenge to the IAT: attitude or mere familiarity?' *Psychological Inquiry* 14: 238–243.

De Houwer, J, Thomas, S & Baeyens, F **(2001)** 'Associative learning of likes and dislikes: a review of 25 years of research on human evaluative conditioning'. *Psychological Bulletin* 127: 853–869.

Denyer, D, Tranfield, D & Van Aken, JE **(2008)** 'Developing design propositions through research synthesis'. *Organization Studies* 29(3): 393–413.

Desert, M & Leyens, J-P **(2006)** 'Social comparison across cultures: gender stereotypes in high and low power distance cultures'. In Guimond, S (ed.) *Social comparison and social psychology: understanding cognition, intergroup relations and culture* (pp303–317). Cambridge University Press, Cambridge.

Devine, PG **(2001)** 'Implicit prejudice and stereotyping: how automatic are they? introduction to the special section'. *Journal of Personality and Social Psychology* 81: 757–759.

Devine, PG **(1989)** 'Stereotypes and prejudice: their automatic and controlled components'. *Journal of Personality and Social Psychology* 56(1): 5–18.

Devine, PG & Baker, SM **(1991)** 'Measurement of racial stereotype subtyping'. *Personality and Social Psychology Bulletin* 17: 44–50.

Devine, PG, Forscher, PS, Austin, AJ & Cox, WTL **(2012)** 'Long-term reduction in implicit race bias: a prejudice habit-breaking intervention'. *Journal of Experimental Social Psychology* 48(6): 1267–1278.

Devine, PG, Plant, EA, Amodio, DM, Harmon-Jones, E & Vance, SL **(2002)** 'The regulation of explicit and implicit race bias : the role of motivations to respond without prejudice'. *Journal of Personality and Social Psychology*, 82(5): 835– 848.

Donders, FC **(1968)** 'On the speed of mental processes'. In Koster, WG (ed.), Attention and performance II. *Acta Psychologica* 30: 412-431. (Original work published in 1868.)

Dougherty, TW, Turban, DB & Callender, JC **(1994)** 'Confirming first impressions in the employment interview: a field study of interviewer behavior'. *Journal of Applied Psychology* 79(5): 659–665.

Dovidio JF **(2002)** 'Implicit and explicit prejudice and interracial interactions'. *Journal of Personality and Social Psychology* 81(1): 62–68.

Dovidio, JF, 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.

Dunton, BC & Fazio, RH **(1997)** An individual difference measure of motivation to control prejudice reactions. *Personality and Social Psychology Bulletin* 23: 316–326.

Eagleman, DM **(2001)** 'Visual illusions and neurobiology'. *Nature Reviews Neuroscience* 2: 920–926.

Eagly, AH & Carli, LL **(2007)** 'Women and the labyrinth of leadership'. *Harvard Business Review* 85(9): 62–71, 146.

Eagly, AH & Karau, SJ **(2002)** 'Role congruity theory of prejudice toward female leaders'. *Psychological Review* 109(3): 573–598.

Eagly, A & Mladinic, A **(1994)** 'Are people prejudiced against women? Some answers from research on attitudes, gender stereotypes, and judgments of competence'. In Stroebe,W & Hewstone, M (eds.) *European Review of Social Psychology* 5: 1–35. Wiley, New York.

Eagly, AH & Mladinic, A **(1989)** 'Gender stereotypes and attitudes towards women and men'. *Personality and Social Psychology Bulletin* 15: 543–548.

Eckes, T **(2002)** 'Paternalistic and envious gender stereotypes: testing predictions from the stereotype content model'. *Sex Roles* 47(3–4): 99–114.

ECU **(2011a)** *Athena SWAN: measuring success*. Equality Challenge Unit, London. **www.athenaswan.org.uk/content/impact-report**

ECU **(2011b)** *The experience of black and minority ethnic staff in higher education in England*. Equality Challenge Unit, London. **www.ecu.ac.uk/publications/experience-of-bme-staff-in-he-final-report**

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

Ely, RJ & Thomas, DA **(2001)** 'Cultural diversity at work: the effects of diversity perspectives on work group processes and outcomes'. *Administrative Science Quarterly* 46(2): 229.

Government Equalities Office **(2010)** *Equality Act 2010: what do I need to know? A quick start guide to using positive action in recruitment and promotion*. **www.gov.uk/government/uploads/system/uploads/attachment_data/file/85014/positive-action-recruitment.pdf**

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

Fiedler, K & Bluemke, M **(2003)** Faking the IAT: aided and unaided response control on the implicit association test. *Basic and Applied Social Psychology* 27(4): 307–316.

Fishbein, M & Ajzen, I **(2010)** *Predicting and changing behavior: the reasoned action approach*. Psychology Press (Taylor & Francis), New York.

Fiske, ST **(2012)** 'Managing ambivalent prejudices: smart-but-cold and warm-but-dumb stereotypes'. *The ANNALS of the American Academy of Political and Social Science* 639(1): 33–48.

Fiske, ST **(2005)** 'Social cognition and the normality of prejudgment'. In Dovidio, JF, Glick, P & Rudman, LA (eds) *On the nature of prejudice: fifty years after Allport* (pp36–53). Blackwell Publishing, Malden, MA.

Fiske, ST **(1993)** 'Controlling other people: the impact of power on stereotyping'. *The American Psychologist* 48(6): 621–628.

Fiske, ST, Cuddy, AJC, 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, ST & Neuberg, SL **(1990)** 'A continuum of impression formation, from category-based to individuating processes: influences of information and motivation on attention and interpretation'. In Zanna, MP (ed.) *Advances in Experimental Social Psychology* 23: 1–74. Academic Press, San Diego, CA.

Fiske, ST & Taylor, SE **(1991)** *Social cognition* (2nd ed.). McGraw-Hill Series In Social Psychology. McGraw-Hill Book Company, New York, NY.

Frijda, NH **(1987)** 'Emotion, cognitive structure, and action tendency'. *Cognition and Emotion* 1(2): 115–143.

Galinsky, AD & Moskowitz, GB **(2000)** 'Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism'. *Journal of Personality and Social Psychology* 78: 708–724.

Gawronski, B, Deutsch, R, Mbirkou, S, Seibt, B & Strack, F **(2008)** 'When 'just say no' is not enough: affirmation versus negation training and the reduction of automatic stereotype activation'. *Journal of Experimental Social Psychology* 44(2): 370–377.

Glick, P & Fiske, ST **(1996)** 'The ambivalent sexism inventory: differentiating and hostile benevolent sexism'. *Journal of Personality and Social Psychology* 70: 491–512.

Green, AR, Carney, DR, Pallin, DJ, Ngo, LH, Raymond, KL, Iezzoni, LI & Banaji, MR **(2007)** 'The presence of implicit bias in physicians and its prediction of thrombolysis decisions for black and white patients'. *Journal of General Internal Medicine* 22: 1231–1238.

Greenwald, AG & Banaji, MR **(1995)** 'Implicit social cognition: attitudes, self-esteem, and stereotypes'. *Psychological Review* 102(1): 4–27.

Greenwald, AG, McGhee, DE & Schwartz, JLK **(1998)** 'Measuring individual differences in implicit cognition: the implicit association test'. *Journal of Personality and Social Psychology* 74: 1464–1480.

Greenwald, A & Nosek, B **(2001)** 'Health of the Implicit Association Test at age 3'. *Zeitschrift für Experimentelle Psychologie* 48: 85–93.

Greenwald, AG, Nosek, BA & Banaji, MR **(2003)** 'Understanding and using the implicit association test: I. An improved scoring algorithm'. *Journal of Personality and Social Psychology* 85: 197–216.

Greenwald, AG, Nosek, BA & Sriram, N **(2006)** 'Consequential validity of the implicit association test: comment on the article by Blanton and Jaccard'. *American Psychologist* 61: 56–61.

Greenwald, A, Poehlman, T, Uhlmann, E & Banaji, M **(2009)** 'Understanding and using the implicit association test: III. Meta-analysis of predictive validity'. *Journal of Personality and Social Psychology* 97: 17–41.

Greenwald, AG, Rudman, LA, Nosek, BA & Zayas, V **(2006)** 'Why so little faith? A reply to Blanton and Jaccard's (2006) skeptical view of testing pure multiplicative theories'. *Psychological Review* 113: 170–180.

Guimond, S, Chatard, A, Martinot, D, Crisp, RJ & Redersdorff, S **(2006)** 'Social comparison, self-stereotyping, and gender differences in self-construals'. *Journal of Personality and Social Psychology* 90(2): 221–242.

Hall, NR, Crisp, RJ & Suen, M **(2009)** Reducing implicit prejudice by blurring intergroup boundaries. *Basic and Applied Social Psychology* 31(3): 244–254.

Han, HA, Czellar, S, Olson, MA & Fazio RH **(2010)** 'Malleability of attitudes or malleability of the IAT?' *Journal of Experimental Social Psychology* 46(2): 286–298.

Harvey, M, McIntyre, N, Thompson Heames, J & Moeller, M **(2009)** 'Mentoring global female managers in the global marketplace: traditional, reverse, and reciprocal mentoring'. *The International Journal of Human Resource Management* 20(6): 1344–1361.

Heider, J & Skowronski, J **(2011)** 'Addendum to Heider and Skowronski (2007): improving the predictive validity of the implicit association test'. *North American Journal of Psychology* 13(1): 17–20.

Heider, J & Skowronski, J **(2007)** 'Improving the predictive validity of the implicit association test'. *North American Journal of Psychology* 9: 53–76.

Henry-Darwish, N & Sanford, C **(2012)** *Awareness of implicit bias: what motivates behavior change?* Liberal and Scholarly Repository. Online resource. **https://dspace.lasrworks.org/handle/10349/1163**

Ito, TA, Chiao, KW, Devine, PG, Lorig, TS & Cacioppo, JT **(2006)** The influence of facial feedback on race bias. *Psychological Science* 17(3): 256–261.

Kahneman, D **(2011)** *Thinking, fast and slow*. Penguin Books Ltd, London.

Kalev, A, Dobbin, F & Kelly, E **(2006)** 'Best practices or best guesses? assessing the efficacy of corporate affirmative action and diversity policies'. *American Sociological Review* 71(4): 589–617.

Karpinski, A & Hilton, JL **(2001)** Attitudes and the implicit association test. *Journal of Personality and Social Psychology* 81(5): 774–788.

Kawakami, K, Dovidio, JF & van Kamp, S **(2005)** 'Kicking the habit: effects of nonstereotypic association training and correction processes on hiring decisions.' *Journal of Experimental Social Psychology* 41(1): 68–75.

Kawakami, K, Phills, CE, Steele, JR & Dovidio, JF **(2007)** '(Close) distance makes the heart grow fonder: improving implicit racial attitudes and interracial interactions through approach behaviors'. *Journal of Personality and Social Psychology* 92(6): 957–971.

Kim, D **(2003)** 'Voluntary controllability of the implicit association test (IAT)'. *Social Psychology Quarterly* 66(1): 83–96.

Lane KA, Banaji MR, Nosek BA, Greenwald AG **(2007)** 'Understanding and using the implicit association test'. In Wittenbrink, B and Schwarz, N (eds) *Implicit measures of attitude* (pp59–102). The Guilford Press, New York.

Lebrecht, S, Pierce, LJ, Tarr, MJ & Tanaka, JW **(2009)** 'Perceptual other-race training reduces implicit racial bias'. *PLoS ONE* 4(1): e4215.

Legault, L, Gutsell, JN & Inzlicht, M **(2011)** 'Ironic effects of antiprejudice messages: how motivational interventions can reduce (but also increase) prejudice'. *Psychological Science* 22(12): 1472–1477.

McConahay, JB **(1986)** 'Modern racism, ambivalence, and the modern racism scale'. In Dovidio, JF & Gaertner, SL (eds) *Prejudice, discrimination and racism* (pp91–126). Academic Press, New York.

McConnell, AR & Leibold, J **(2009)** 'Weak criticisms and selective evidence: reply to Blanton *et al* (2009)'. *Journal of Applied Psychology* 94(3): 583–589.

McConnell AR & Liebold, JM **(2001)** 'Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attitudes'. *Journal of Experimental Social Psychology* 37: 435–442.

McConnell, AR, Leibold, JM & Sherman, SJ **(1997)** 'Within-target illusory correlations and the formation of context-dependent attitudes'. *Journal of Personality and Social Psychology* 73: 675–686.

Meissner, CA & Brigham, JC **(2001)** 'Thirty years of investigating the own-race bias in memory for faces'. *Psychology, Public Policy, and Law* 7: 3–35.

Mendoza, SA, Gollwitzer, PM & Amodio, DM **(2010)** 'Reducing the expression of implicit stereotypes: reflexive control through implementation intentions'. *Personality and Social Psychology Bulletin* 36(4): 512–523.

Messick, S **(1995)** 'Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning'. *American Psychologist* 50: 741–749.

Mitchell, G & Tetlock, PE **(2006)** 'Antidiscrimination law and the perils of mindreading'. *Ohio State Law Journal* 67: 1023–1121.

Mitchell, JA, Nosek, BA & Banaji, MR **(2003)** 'Contextual variations in implicit evaluation'. *Journal of Experimental Psychology: General* 132(3): 455–469.

Monteith, MJ **(1993)** 'Self-regulation of prejudiced responses: implications for progress in prejudice-reduction efforts'. *Journal of Personality and Social Psychology* 65: 469–485.

Monteith, MJ, Ashburn-Nardo, L, Voils, CI & Czopp, AM **(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.

Monteith, MJ & Mark, AY **(2005)** 'Changing one's prejudiced ways: awareness, affect, and self-regulation'. *European Review of Social Psychology* 16: 113–154.

Morris, KA & Ashburn-Nardo, L **(2009)** 'The implicit association test as a class assignment: student affective and attitudinal reactions'. *Teaching of Psychology* 37(1): 63–68.

Moss-Racusin, CA, Dovidio, JF, Brescoll, VL, Graham, M & Handelsman, J **(2012)** 'Science faculty's subtle gender biases favor male students'. *Proceedings of the National Academy of Sciences for the United States of America* 109(41): 16474–16479.

Nesdale, D & Durkin, K **(1998)** 'Stereotypes and attitudes: implicit and explicit processes'. In Speelman, C & Kirsner, K (eds) *Implicit and explicit mental processes* (pp219–232). Erlbaum Associates, Mahwah, NJ.

Nock MK & Banaji, MR **(2007)** 'Prediction of suicide ideation and attempts among adolescents using a brief performance-based test'. *Journal of Consulting and Clinical Psychology* 75(5): 707–715.

Norton, MI, Sommers, SR, Apfelbaum, EP, Pura, N & Ariely, D **(2006)** 'Color blindness and interracial interaction: playing the political correctness game'. *Psychological Science* 17(11): 949–953.

Nosek, B, Bar-Anan, Y, Sriram, N & Greenwald, AG **(2013)** *Understanding and using the brief implicit association test: I. Recommended scoring procedures*. Unpublished research. **http://projectimplicit.net/nosek/papers**

Nosek, BA, Greenwald, AG & Banaji, MR **(2007)** 'The implicit association test at age 7: a methodological and conceptual review'. In Bargh, JA (ed.) *Social psychology and the implicit: the automaticity of higher mental processes* (pp265–292). Psychology Press, New York.

Nosek, BA, Greenwald, AG & Banaji, MR **(2005)** Understanding and using the implicit association test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin* 31: 166–180.

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

Olson MA & Fazio RH **(2008)** 'Discordant evaluations of blacks affect nonverbal behavior'. *Personality and Social Psychology Bulletin* 33: 1214–1224.

Olson, MA & Fazio, RH **(2006)** 'Reducing automatically activated racial prejudice through implicit evaluative conditioning'. *Personality and Social Psychology Bulletin* 32(4): 421–433.

Olson, MA & Fazio, RH **(2002)** 'Implicit acquisition and manifestation of classically conditioned attitudes'. *Social Cognition* 20: 89–104.

Olson, MA & Fazio, RH **(2001)** 'Implicit attitude formation through classical conditioning'. *Psychological Science* 12: 413–417.

Paluck, EL & Green, DP **(2009)** 'Prejudice reduction: what works? A review and assessment of research and practice'. *Annual Review of Psychology* 60: 339–367.

Park, J, Felix, K & Lee, G **(2007)** 'Implicit attitudes toward Arab-Muslims and the moderating effects of social information'. *Basic and Applied Social Psychology* 29(1): 35–45.

Paul, R & Ellen, J **(1993)** 'Beyond formal experimental design: towards an expanded view'. *Personnel Psychology* 46(3): 613–627.

Payne, BK, Lambert, AJ & Jacoby, LL **(2002)** 'Best laid plans: effects of goals on accessibility bias and cognitive control in race-based misperceptions of weapons'. *Journal of Experimental Social Psychology* 38(4): 384–396.

Pérez, E **(2010)** 'Explicit evidence on the import of implicit attitudes: the IAT and immigration policy judgments'. *Political Behavior* 32: 517–545.

Pettigrew, TF **(1998)** 'Intergroup contact theory'. *Annual Review of Psychology* 49: 65–85.

Pettigrew, TF **(1979)** 'The ultimate attribution error: extending Allport's cognitive analysis of prejudice.' *Personality and Social Psychology Bulletin* 5(4): 461–476.

Pettigrew, TF & Tropp, LR **(2006)** 'A meta-analytic test of intergroup contact theory'. *Journal of Personality and Social Psychology* 90: 751–783.

Phelps, E, O'Connor, K, Cunningham, W Funayama, EJ, Gatenby, C, Gore, J, and Banaji, M **(2000)** 'Performance on indirect measures of race evaluation predicts amygdala activation'. *Journal of Cognitive Neuroscience* 12(5): 729–738.

Plant, EA & Devine, PG **(1998)** 'Internal and external motivation to respond without prejudice'. *Journal of Personality and Social Psychology* 75: 811–832.

Plant, EA, Devine, PG, Cox, WTL, Columb, C, Miller, SL, Goplen, J & Peruche, BM **(2009)** 'The Obama effect: decreasing implicit prejudice and stereotyping'. *Journal of Experimental Social Psychology* 45(4): 961–964.

Plant, EA, Devine, PG & Peruche, MB **(2010)** 'Routes to positive interracial interactions: approaching egalitarianism or avoiding prejudice'. *Personality and Social Psychology Bulletin* 36(9): 1135–1147.

Poehlman, TA, Uhlmann, E, Greenwald, AG & Banaji, MR **(2005)** 'Understanding and using the implicit association test: III. Metaanalysis of predictive validity'. *Journal of Personality and Social Psychology* 97(1): 17–41.

Ramasubramanian, S **(2007)** 'Media-based strategies to reduce racial stereotypes activated by news stories'. *Journalism and Mass Communication Quarterly* 84(2): 249–264.

Reskin, B **(2000)** 'The proximate causes of employment discrimination'. *Contemporary Sociology* 29(2): 319–328.

Riach, PA & Rich, J **(2002)** 'Field experiments of discrimination in the market place'. *Economic Journal* 112, F480–F518.

Richeson, J **(2004)** 'The impact of multiculturalism versus color-blindness on racial bias'. *Journal of Experimental Social Psychology* 40(3): 417–423.

Richeson, JA & Shelton, JN **(2005)** 'Thin slices of racial bias'. *Journal of Nonverbal Behavior* 29: 75–86.

Riordan, C **(1990)** *Girls and boys in school: together or separate?* Teachers College Press, New York.

Riordan, C **(1994)** 'The value of attending a women's college'. *Journal of Higher Education* 65: 486–510.

Rooth, D **(2010)** 'Automatic associations and discrimination in hiring: real-world evidence'. *Labour Economics* 17: 523–534.

Rooth, D **(2009)** 'Obesity, attractiveness, and differential treatment in hiring: a field experiment'. *Journal of Human Resources* 44: 710–735.

Rooth, D **(2008)** *Automatically activated stereotypes and differential treatment against the obese in hiring*. Institute for the Study of Labor, Bonn.

Rooth, D **(2007)** *Implicit discrimination in hiring: real world evidence*. Institute for the Study of Labor, Bonn.

Rosenberg, MJ & Hovland, CI **(1960)** 'Cognitive, affective, and behavioral components of attitudes'. In Hovland, CI & Rosenberg, MJ (eds) *Attitude organization and change*. Yale University Press, New Haven.

Rudman, LA **(2008)** 'The validity of the implicit association test is a scientific certainty.' *Industrial and Organizational Psychology* 1(4): 426–429.

Rudman, LA, Ashmore, RD & Gary, ML **(2001)** '"Unlearning" automatic biases: the malleability of implicit prejudice and stereotypes'. *Journal of Personality and Social Psychology* 81(5): 856–868.

Rudman, LA & Glick, P **(2001)** 'Prescriptive gender stereotypes and backlash toward agentic women'. *Journal of Social Issues* 57: 743–762.

Schwarz, N **(2000)** 'Emotion, cognition, and decision making'. *Cognition and Emotion* 14(4): 433–440.

Sebrechts, JS **(1993)** 'Cultivating scientists at women's colleges'. *Initiatives* 55(2): 45–51.

Segrest, S **(2010)** 'Rater error bias training in the employment interview and racioethnicity biased perceptions'. *Journal of Academic Business Ethics* 3: 99–113.

Shelton, N, Richeson, JA, Salvatore, J & Trawalter, S **(2005)** 'Ironic effects of racial bias during interracial interactions'. *Psychological Science* 16: 397–402.

Sherif, M **(1966)** *In common predicament: social psychology of intergroup conflict and cooperation*. Houghton Mifflin, Boston.

Shook, NJ & Fazio, RH **(2008)** 'Interracial roommate relationships: an experimental field test of the contact hypothesis'. *Psychological Science* 19(7): 717–723.

Solnick, SJ **(1995)** 'Changes in women's majors from entrance to graduation at women's and coeducational colleges'. *Industrial and Labor Relations Review* 48: 505–514.

Sriram, N & Greenwald, AG **(2009)** 'The brief implicit association test'. *Experimental Psychology* 56: 283–294.

Srull, TK & Wyer, RS **(1979)** 'The role of category accessibility in the interpretation of information about persons: some determinants and implications'. *Journal of Personality and Social Psychology* 37: 1660–1672.

Steffens, M **(2004)** 'Is the implicit association test immune to faking?' *Experimental Psychology* 51(3): 165–179.

Steinpreis, RE, Anders, K, Ritzke, D **(1999)** 'The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: a national empirical study'. *Sex Roles* 41(7/8): 509–528.

Stewart, BD & Payne, BK **(2008)** 'Bringing automatic stereotyping under control: implementation intentions as efficient means of thought control'. *Personality and Social Psychology Bulletin* 34(10): 1332–1345.

Stewart, TL, Latu, IM, Kawakami, K & Myers, AC **(2010)** 'Consider the situation: reducing automatic stereotyping through situational attribution training.' *Journal of Experimental Social Psychology* 46(1): 221–225.

Tajfel, H **(1970)** 'Experiments in intergroup discrimination'. *Scientific American* 223(5): 96–102.

Tajfel, H & Forgas, JP **(1981)** 'Social categorization: cognitions, values and groups'. In Forgas, JP (ed.) *Social cognition* (pp113–140). Academic Press, New York.

Tajfel, H & Turner, JC **(1979)** 'An integrative theory of intergroup conflict'. In Austin, W & Worchel, S (eds) *The social psychology of intergroup relations* (pp33–48). Brooks/Cole, Pacific Grove, CA.

Tidball, ME **(1985)** 'Baccalaureate origins of entrants into American medical schools'. *Journal of Higher Education* 56: 385–402.

Tidball, ME & Kistiakowsky, V **(1976)** 'Baccalaureate origins of American scientists and scholars'. *Science* 193: 646–652.

Trawalter, S, Richeson, JA & Shelton, JN **(2009)** 'Predicting behavior during interracial interactions: a stress and coping approach'. *Personality and Social Psychology Review* 13(4): 243–268.

Turner, RN & Crisp, RJ **(2010)** 'Imagining intergroup contact reduces implicit prejudice'. *British Journal of Social Psychology* 49(1): 129–142.

Uhlmann, EL & Cohen, GL **(2005)** 'Constructed criteria redefining merit to justify discrimination'. *American Psychological Society* 16(6): 474–480.

Vorauer, JD **(2012)** 'Completing the implicit association test reduces positive intergroup interaction behavior'. *Psychological Science* 23(10): 1168–1175.

Vorauer, JD & Turpie, CA **(2004)** 'Disruptive effects of vigilance on dominant group members' treatment of out-group members: choking versus shining under pressure'. *Journal of Personality and Social Psychology* 87(3): 384–399.

Wegener, DT & Petty, RE **(1997)** 'The flexible correction model: the role of native theories of bias in bias correction'. In Zanna, MP (ed.) *Advances in experimental social psychology 29* (pp141–208). Academic Press, New York.

Wilson, TD & Brekke, N **(1994)** 'Mental contamination and mental correction: unwanted influences on judgments and evaluations'. *Psychological Bulletin* 116: 117–142.

Wolsko, C, Park, B, Judd, CM & Wittenbrink, B **(2000)** 'Framing interethnic ideology: effects of multicultural and color-blind perspectives on judgments of groups and individuals'. *Journal of Personality and Social Psychology* 78(4): 635–654.

Wood, M, Hales, J, Purdon, S, Sejersen, T & Hayllar, O **(2009)** *A test for racial discrimination in recruitment practice in British cities: research report no 607*. Department for Work and Pensions, London.

Ziegert, JC & Hanges, PJ **(2009)** 'Strong rebuttal for weak criticisms'. *Journal of Applied Psychology* 94(3): 590–597.

Ziegert, J & Hanges, P **(2005)** 'Employment discrimination: the role of implicit attitudes, motivation, and a climate for racial bias'. *Journal of Applied Psychology* 90: 553–562.

Zogmaister, C, Arcuri, L, Castelli, L & Smith, ER **(2008)** 'The impact of loyalty and equality on implicit in-group favoritism'. *Group Processes and Intergroup Relations* 11(4): 493–512.

Equality Challenge Unit

ECU works to further and support equality and diversity for staff and students in higher education across all four nations of the UK, and in colleges in Scotland.

ECU works closely with colleges and universities to seek to ensure that staff and students are not unfairly excluded, marginalised or disadvantaged because of age, disability, gender identity, marital or civil partnership status, pregnancy or maternity status, race, religion or belief, sex, sexual orientation or through any combination of these characteristics or other unfair treatment.

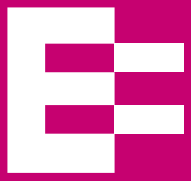
Providing a central source of expertise, research, advice and leadership, we support institutions in building a culture that provides equality of both opportunity and outcome, promotes good relations, values the benefits of diversity and provides a model of equality for the wider UK society.

Did you find this publication useful?

Your feedback will help us to improve and develop our publications and resources, and help us to ensure that we produce materials that support your work.

Please take a few minutes to complete our publications feedback survey: **www.surveymonkey.com/s/ecu-publications-feedback**

You can also email us with your feedback: **pubs@ecu.ac.uk**



Equality Challenge Unit

7th floor, Queen's House
55/56 Lincoln's Inn Fields
London, WC2A 3LJ
T 020 7438 1010
F 020 7438 1011
E info@ecu.ac.uk
www.ecu.ac.uk

© Equality Challenge Unit
ECU's publications are produced free of charge to the UK HE sector and also for colleges in Scotland. Information can be reproduced as long as it is accurate, the source is identified and it will not be used for profit. Alternative formats are available: E **pubs@ecu.ac.uk**

Company limited by guarantee. Registered in England and Wales, No. 05689975.
Charity no. 1114417 (England and Wales) and SC043601 (Scotland).